Schools face new test as pandemic resurges p. 1023

When machines collude, can our laws stop them? p. 1040

Real-time tracking of roaming dynamics p. 1072

Science

$15
27 NOVEMBER 2020
sciencemag.org

AAAS

VARIATION MATTERS

Heterozygosity improves translocation success in endangered desert tortoises p. 1086

CONTENTS

27 NOVEMBER 2020
VOLUME 370 • ISSUE 6520

1023

PHOTO: ANDY BUCHANAN/DIGITAL/AFP VIA GETTY IMAGES

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

PERSPECTIVES

POLICY FORUM

BOOKS *ET AL.*

Published by AAAS

PHOTO: WILLI ROLFES/MINDEN PICTURES

RESEARCH

IN BRIEF

REVIEW

RESEARCH ARTICLES

REPORTS

DEPARTMENTS

ON THE COVER

For many imperiled species, including the Mojave desert tortoise (*Gopherus agassizii*), translocation of individuals to augment populations is an important conservation strategy. Researchers now demonstrate that individual heterozygosity is a key predictor of translocation success in these tortoises—a finding that may have broad implications for improving survival outcomes for many endangered species. See page 1086. *Photo: Cameron Rognan*

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. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. 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 available from backissues.sciencemag.org; 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 can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Published by AAAS

Science AAAS

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

REPRINTS AND PERMISSIONS
sciencemag.org/help/reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services

science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.sciencemag.org/science-careers

advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org

support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

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

Published by AAAS

EDITORIAL

Modeling, post COVID-19

Much of the public first learned about epidemiological modeling during the early months of the coronavirus disease 2019 (COVID-19) pandemic. The first models resulted in more confusion than clarity. Even though coronavirus cases were rising exponentially in the United States and Europe, some models predicted a rapid peak followed by a rapid decline, whereas other models predicted cycles of infection continuing over several years. Much has been learned since those early months. In retrospect, it is clear that modeling requires both reliable data and an accurate understanding of how disease spreads, and that the field of epidemiological modeling requires a diversity of approaches. Support for this field must increase and be coordinated, with a designation of responsibilities among funding agencies.

Early models relied on sparse, sometimes unreliable, data, and modelers did not anticipate the emergence of important new facts—frequent transmission by asymptomatic carriers, the disproportionate impact of superspreaders, and the role of aerosol-mediated transmission. As data accumulated and new information about transmission was incorporated, the modeling improved.

Basic research in forecasting epidemics ranges from traditional models that can be run on a laptop to those that simulate the daily activities of a hundred thousand or more people and require a supercomputer. However, research on modeling is funded haphazardly in the United States, with no single federal agency having ownership. If the nation expects to do better in the inevitable next pandemic, basic research needs more support, and it needs a lead agency to strengthen coordination across research groups from a wide variety of disciplines including public health, medicine, statistics, computer science, and the behavioral sciences.

For use in an emergency, models developed through basic research need to be "operationalized"—that is, made robust for evaluating specific policy interventions. "Nowcasting" requires models that integrate incomplete, real-time data and emerging medical knowledge to provide situational awareness. Such specialized models can address questions such as, "If we do this, how many fewer deaths are likely?" Models must also incorporate behavioral responses to policy interventions that may change the course of an epidemic. By contrast, forecasts of a hurricane's path and intensity need not take into account behavioral responses: The forecaster may tell people to take shelter; whether they do so does not change the path of the hurricane.

The comparison to weather research and prediction is helpful when considering the changes needed to bolster the effectiveness of epidemiological modeling. For the weather, two federal agencies have clear responsibilities: The National Science Foundation (NSF) supports basic research in weather and climate modeling, whereas the National Oceanic and Atmospheric Administration's National Weather Service distributes results of its operational models to weather-forecast offices. The United States needs an analogous structure for epidemiological modeling and forecasting—one or more agencies with clearly defined responsibilities for supporting basic research in modeling, coordinated with one or more agencies with defined operational responsibilities.

> "To conquer the next pandemic, it will be essential to have a diverse array of improved models..."

We and other former members of the Obama administration's President's Council of Advisors on Science and Technology (http://opcast.org), after consultation with modelers and policy analysts, recently concluded that NSF, because of its unmatched outreach into computational and data sciences, is well-positioned to act as the lead research agency and should work closely with the National Institutes of Health to ensure that epidemiological models are consistent with what is known about human disease. The U.S. Centers for Disease Control and Prevention should be the lead operational agency, with a new office for epidemic forecasting and analytics. The national supercomputing infrastructure of the Department of Energy could help support the computing requirements of the combined programs.

To conquer the next pandemic, it will be essential to have a diverse array of improved models that quickly incorporate new data and emerging medical knowledge, and that simulate the effects of various public policy interventions. If the United States takes action now by increasing funding through agencies with well-defined roles and responsibilities, it will be better prepared to save lives in the future.

–William H. Press and Richard C. Levin

William H. Press
is a former member of the President's Council of Advisors on Science and Technology (PCAST) in the Obama administration. He is a professor at The University of Texas at Austin, TX, USA, and a past president of the American Association for the Advancement of Science (AAAS, the publisher of *Science*), Washington, DC, USA.
wpress@utexas.edu

Richard C. Levin
is a former member of PCAST in the Obama administration and president emeritus of Yale University, New Haven, CT, USA.
richard.levin@yale.edu

10.1126/science.abf7914

NEWS

$1,000,000

Gift from entertainer **Dolly Parton** in April to support development of Moderna's coronavirus vaccine, which the company last week said showed an efficacy of 94.5%. "I felt so proud to have been part of that little seed money," Parton told BBC.

IN BRIEF

Edited by **Jeffrey Brainard**

A woman in Équateur province in the DRC is examined for symptoms of Ebola. She tested negative.

INFECTIOUS DISEASES

Congo declares end to Ebola outbreak

The 11th Ebola outbreak in the Democratic Republic of the Congo (DRC) is officially over, giving the country respite from the disease for the first time in more than 2 years. On 18 November, the World Health Organization (WHO) announced that no new cases had been identified for 42 days, twice the incubation period for the deadly virus. The outbreak, in the western Équateur province, started in late May, just as a bigger one in the eastern DRC was coming to an end. (That outbreak had killed 2200 people.) The Équateur outbreak sickened 130 and killed 55; a campaign that vaccinated more than 40,000 people is credited with helping end it. Special portable coolers that keep the vaccine at −80°C for up to 1 week allowed health workers to administer the shots in communities deep in the rainforest, accessible only by boat or helicopter. The same technology will be useful in efforts to distribute COVID-19 vaccines in Africa, says Matshidiso Moeti, WHO's regional director. The coronavirus pandemic complicated the fight against Ebola, WHO says, but the expertise gained by local health workers in earlier outbreaks in the region was a major advantage. They will remain on the lookout for potential flare-ups.

FDA OKs at-home COVID-19 test

DIAGNOSTICS | The U.S. Food and Drug Administration (FDA) issued its first emergency use authorization last week for an at-home diagnostic test that can detect the pandemic coronavirus in just minutes. However, the test might not be widely available until spring 2021. Produced by Lucira Health, a biotech company, it is expected to cost less than $50 and require a doctor's prescription. The company says it will soon distribute tests in parts of California and Florida; it says it needs time to scale up manufacturing for national distribution. Lucira's test amplifies viral genetic material, making it nearly as accurate as laboratory tests that use the polymerase chain reaction, the current gold standard. FDA previously approved at-home tests that must be mailed to a laboratory for analysis. Several other companies are working on rapid antigen tests, which detect viral particles, for home use. But concerns remain about antigen tests' reliability. Still, some public health specialists consider widely available, low-cost, at-home testing vital for controlling the pandemic.

NIH boosts young scientists

FUNDING | A new U.S. National Institutes of Health (NIH) award will allow early-career investigators who want to shift research directions when applying for their first independent award to submit a proposal without first generating preliminary data to support their idea. Reviewers will instead assess the soundness of the project's approach. The Katz award is named for Stephen Katz, a longtime champion of young researchers who was director of the National Institute of Arthritis and Musculoskeletal and Skin Diseases when he died in 2018. The grant will build on an NIH policy that prioritizes proposals from early-stage investigators—those no more than 10 years from completing their training who are applying for their first research grant. The policy has been credited with raising their numbers from fewer than 600 supported in 2013 to more than 1300 last year. Applications for the first Katz awards are due on 26 January 2021.

PHOTO: CAROLINE THIRION/MSF

Published by AAAS

NIST appointee draws protest

LEADERSHIP | Democrats in Congress say a political appointee given a senior post at the U.S. National Institute of Standards and Technology (NIST) is unfit for the job because he lacks technical skills and holds pseudoscientific views about racial differences on IQ tests. On 9 November, Jason Richwine, an independent public policy analyst, took up the new position of deputy undersecretary of commerce for standards and technology, and Commerce Secretary Wilbur Ross subsequently issued an order that would put Richwine in charge of the $1 billion research agency if NIST Director Walter Copan leaves or is fired. On 17 November, Representative Eddie Bernice Johnson (D–TX), who leads the science committee in the U.S. House of Representatives, asked Ross to justify the moves. Richwine has advocated for more restrictive immigration policies, and his 2009 doctoral thesis argued that lower IQ scores by Mexican and Hispanic immigrants suggest a genetic component to intelligence that is "likely to persist over several generations."

Mentoring paper draws scrutiny

DIVERSITY | The editors of *Nature Communications* say they are reviewing a paper that drew scalding criticism after it suggested that encouraging female junior scientists to work with female mentors could "hinder the careers of women." The 17 November study, led by data scientist Bedoor AlShebli of New York University, Abu Dhabi, examined 3 million mentor-protégé pairs and how gender influenced the impact of papers later published by the protégés. Female protégés, it concluded, did better if they worked with male mentors. Critics pounced, noting the authors ignored reviewer complaints about the study's methods and arguing the journal was promoting a harmful and unfounded message. The article's authors said they welcome the review.

Europe warns of avian flu spread

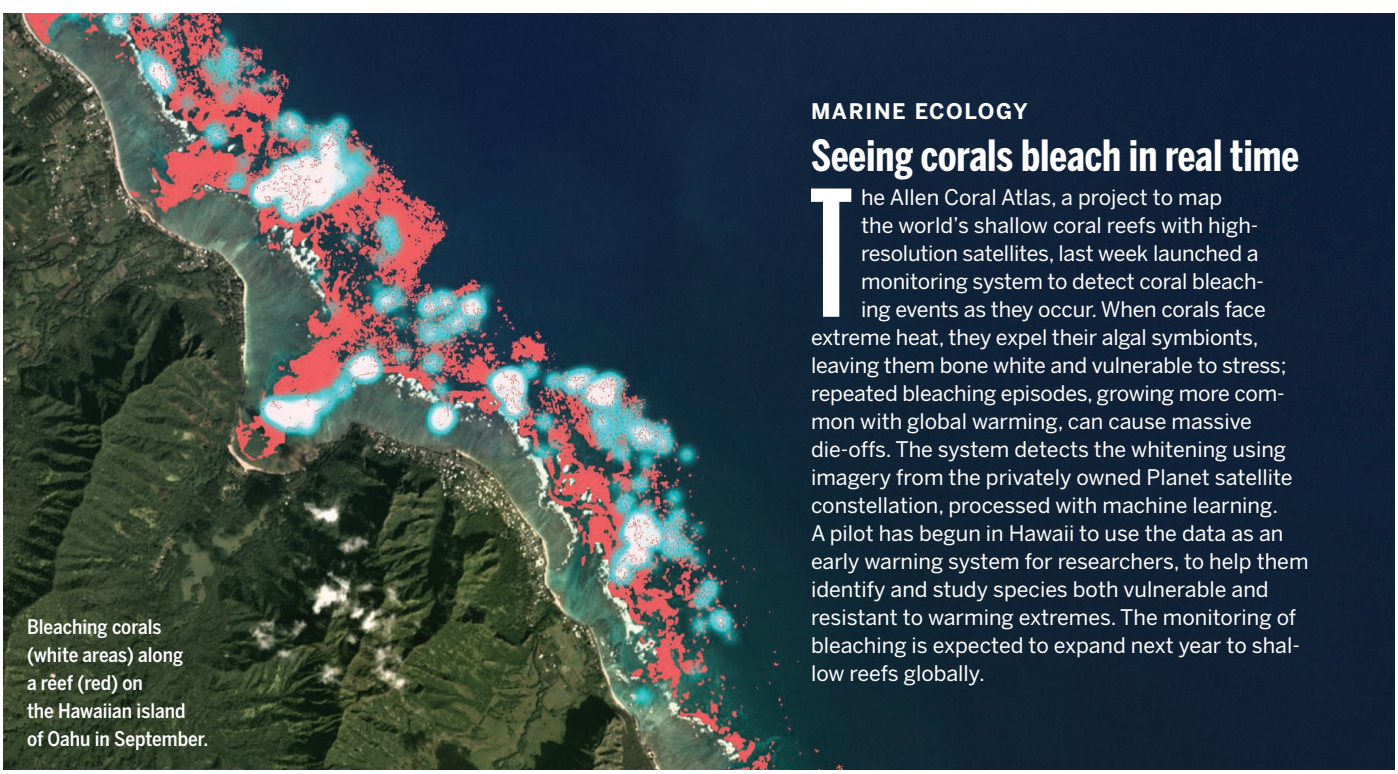
ANIMAL DISEASES | European authorities reported on 19 November they have detected highly pathogenic avian influenza in 302 birds in eight countries. Only 18 cases were in poultry; most of the rest were in wild birds, the European Food Safety Authority and its partners said. The number of infected birds is expected to rise with winter migrations. Several flu strains were identified, but no people were reported to be infected, and the risk of that occurring is considered low; researchers studying the viruses found no genetic markers indicating they had adapted to infect mammals. But the threat to poultry is high, and the report's authors recommended bird producers increase precautions against infections.

IN OTHER NEWS

VACCINE APPLICATION Days after making public the final analysis of their 40,000-person COVID-19 vaccine trial, which found 95% efficacy, Pfizer and its German partner BioNTech filed for emergency authorization of the messenger RNA vaccine from the U.S. Food and Drug Administration—the first such request for a vaccine during the pandemic. They plan to seek additional approvals in other countries soon. Pfizer hopes to supply up to 50 million doses this year.

REMDESIVIR PANNED A World Health Organization panel recommended against using the antiviral drug remdesivir to treat most hospitalized COVID-19 patients. Its review of four studies of 7000 people found that the drug, which the U.S. Food and Drug Administration approved last month for hospitalized patients, did not reduce mortality or speed recovery. But the panel encouraged further study of it.

AMMO BAN Denmark has become the first nation to ban all lead-based hunting ammunition, including bullets and shotgun pellets, to protect wildlife. Hunters annually release about 2 tons of lead into Denmark's environment; waterbirds and other species eat the toxic material and die. European regulators are considering a ban like Denmark's.



Bleaching corals (white areas) along a reef (red) on the Hawaiian island of Oahu in September.

MAP: ALLEN CORAL ATLAS

MARINE ECOLOGY

Seeing corals bleach in real time

The Allen Coral Atlas, a project to map the world's shallow coral reefs with high-resolution satellites, last week launched a monitoring system to detect coral bleaching events as they occur. When corals face extreme heat, they expel their algal symbionts, leaving them bone white and vulnerable to stress; repeated bleaching episodes, growing more common with global warming, can cause massive die-offs. The system detects the whitening using imagery from the privately owned Planet satellite constellation, processed with machine learning. A pilot has begun in Hawaii to use the data as an early warning system for researchers, to help them identify and study species both vulnerable and resistant to warming extremes. The monitoring of bleaching is expected to expand next year to shallow reefs globally.

Broken suspension cables have torn through the Arecibo Observatory's dish panels and threaten to send the 900-ton instrument platform crashing down.

PHOTO: ARECIBO OBSERVATORY/UNIVERSITY OF CENTRAL FLORIDA

ASTRONOMY

Arecibo radio telescope to be decommissioned

Famed observatory set for demolition after cable failures put it on the brink of collapse

By **Daniel Clery**

The giant radio dish, nestled in a karst sinkhole amid lush hills near Arecibo, Puerto Rico, saw plenty of fame in its time. It was the first to detect indirect evidence for gravitational waves, the first to map Venus's surface, and the first to detect exoplanets. Now, following cable breaks that have brought the 57-year-old facility to the brink of collapse, the Arecibo Observatory will be demolished, officials at the National Science Foundation (NSF) announced on 19 November.

Researchers are mourning a facility that won acclaim in many fields: atmospheric science, planetary science, and astrophysics. "It's like the death of a family member," says former observatory Director Robert Kerr. Jill Tarter of the SETI Institute, who used the 305-meter-wide dish to scan nearby stars for possible alien communications, says Arecibo is unlike any other observatory. Its striking appearance earned it supporting roles in movies such as *Contact* and *GoldenEye*. "It's a really special place, with a somewhat exotic cultural dimension."

In an effort to save money for newer telescopes, NSF in 2018 handed Arecibo's management to a consortium led by the University of Central Florida, which shouldered more of its finances. But Maura McLaughlin, co-director of the Nanohertz Observatory for Gravitational Waves (NANOGrav) project, says Arecibo was still doing frontier research. The loss will set back NANOGrav's hunt by as much as 10 years, she says. "It's a real blow."

NSF officials say Arecibo is now so fragile that attempting repairs would put staff and workers in danger. The agency wants to preserve other instruments at the site, as well as the popular visitor's center. But they could be damaged if the 900-ton instrument platform, suspended 137 meters above the dish, comes crashing down—or if any of the three support towers fell. "There is a serious risk of an unexpected and uncontrolled collapse," Ralph Gaume, director of NSF's astronomy division, said at a news briefing last week. Over the next few weeks, engineers will plan for a controlled dismantling beginning in the spring, which will likely involve an explosive release of the platform from its cables.

Hurricane Maria battered Arecibo in 2017. Since December 2019, a string of earthquakes, including 11 of magnitude 5 or greater, have shaken the site. It's unknown whether those stresses contributed to the failures of the two suspension cables. The first came on 10 August, when an auxiliary cable, installed in the 1990s as 300 tons of instruments were added to the platform, broke away at one end, damaging some instruments and leaving a 30-meter gash in the dish below.

Engineers investigating the failure noticed a dozen broken wires in a main suspension cable—one of 12—that was connected to the same tower as the failed auxiliary cable. Because these 9-centimeter-thick cables are made up of 160 wires, they thought that one could shoulder the extra load. But on 7 November, it also broke.

The three engineering firms already onsite set to analyzing the structure's safety. NSF sent another firm and the Army Corps of Engineers. Three of the five said controlled demolition was the only way forward. "Now all the cables are suspect," Ashley Zauderer, NSF's program director for the observatory, said at the briefing. If any of three remaining main cables on that one tower failed, the engineers found, the platform would collapse.

A forensic examination of the cable failures is too dangerous at the moment, but they may have degraded in the warm, moist environment, which Kerr says is "difficult for steel." Keeping the cables in condition "was always a maintenance challenge," he says. Tarter speculates that corners may have been cut. "When funding has diminished, maintenance gets deferred or minimized," she says. Gaume said at the briefing that Arecibo managers "have done everything they were required to do and inspected the cables as necessary."

From 1960 to 1963, the Pentagon's Ad-

vanced Research Projects Agency built Arecibo to study Earth's upper atmosphere for ballistic missile defense. Astronomers soon realized the potential of the huge dish to gather radio signals from outer space, and in the late 1960s, NSF took it over. The NASA-funded search for extraterrestrial intelligence (SETI) used it to listen to nearby stars and, after funding was cut in 1992, the private SETI Institute continued to fund the Arecibo search. Because Arecibo can generate as well as receive radio signals, SETI pioneer Frank Drake in 1974 turned to it to send a message out to the star cluster M13.

Later, researchers used its radar beams to discover icy craters on Mercury and to peer through the thick sulfurous clouds of Venus, revealing hidden mountain ranges. From its early days, Arecibo studied pulsars—rapidly spinning neutron stars that emit metronomic radio pulses. In 1974, with Arecibo's help, Joseph Taylor and Russell Hulse discovered a pulsar-neutron star pair whose mutual orbit was slowly contracting—a sign they were shedding energy as gravitational waves. The researchers won the 1993 Nobel Prize in Physics for the discovery.

Pulsars that didn't keep perfect time led to other discoveries. In 1992, researchers detected variations in the beats from one, which indicated something was regularly tugging on it. That turned out to be two planets—the first confirmed exoplanet detections. At Arecibo, Kerr says, "Every night was exciting."

Many Arecibo researchers can turn to other facilities. But some need its extreme sensitivity: The single-dish collecting area was surpassed only by China's Five-hundred-meter Aperture Spherical radio Telescope, which opened in 2016. McLaughlin's NANOGrav project has for 15 years recorded faint pulsar beats in the hope that tiny variations could reveal the passing of gravitational waves. Two-thirds of its observations were done at Arecibo.

NASA researchers will also miss Arecibo's ability to monitor nearby asteroids that could threaten Earth. Former Arecibo Director Michael Nolan used the telescope for more than 20 years to bounce radio pulses off asteroid Bennu. The resulting maps helped guide NASA's OSIRIS-REx mission, which last month grabbed a sample from Bennu and is now heading home. Its absence "will definitely affect future mission planning" to other asteroids, Nolan says.

The loss of the telescope will also affect staff and their families. "The economic impact on the local community really is a sad note," Kerr says. "We are thinking of all the staff at the observatory, who've dealt with so many setbacks over the past few years," McLaughlin says. They "truly have done anything they can to keep Arecibo going." ■

FUSION ENERGY

Laser fusion reactor approaches 'burning plasma' milestone

After a decade, National Ignition Facility nears a self-heated, sustained reaction, though net energy gain is still elusive

By **Daniel Clery**

In October 2010, in a building the size of three U.S. football fields, researchers at the Lawrence Livermore National Laboratory powered up 192 laser beams, focused their energy into a pulse with the punch of a speeding truck, and fired it at a pellet of nuclear fuel the size of a peppercorn. So began a campaign by the National Ignition Facility (NIF) to achieve the goal it is named for: igniting a fusion reaction that produces more energy than the laser puts in.

A decade and nearly 3000 shots later, NIF is still generating more fizz than bang, hampered by the complex, poorly understood behavior of the laser targets when they vaporize and implode. But with new target designs and laser pulse shapes, along with better tools to monitor the miniature explosions, NIF researchers believe they are close to an important intermediate milestone known as "burning plasma": a fusion burn sustained by the heat of the reaction itself rather than the input of laser energy.

Self-heating is key to burning up all the fuel and getting runaway energy gain. Once NIF reaches the threshold, simulations suggest it will have an easier path to ignition, says Mark Herrmann, who oversees Livermore's fusion program. "We're pushing as hard as we can," he says. "You can feel the acceleration in our understanding." Outsiders are impressed, too. "You kind of feel there's steady progress and less guesswork," says Steven Rose, co-director of the Centre for Inertial Fusion Studies at Imperial College London. "They're moving away from designs traditionally held and trying new things."

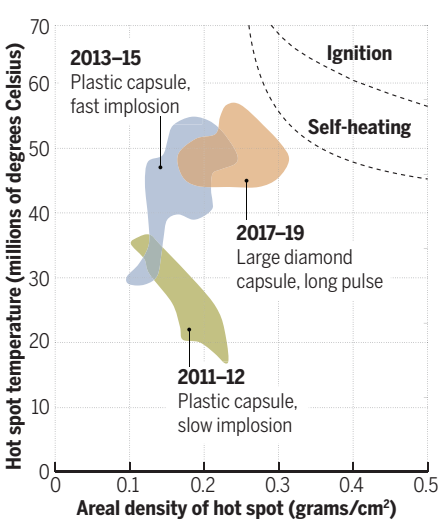
NIF may not have the luxury of time, however. The proportion of NIF shots devoted to the ignition effort has been cut from a high of nearly 60% in 2012 to less than 30% today to reserve more shots for stockpile stewardship—experiments that simulate nuclear detonations to help verify the reliability of warheads. Presidential budget requests in recent years have repeatedly sought to slash research into inertial confinement fusion at NIF and elsewhere, only to have Congress preserve it. NIF's funder, the National Nuclear Security Administration (NNSA), is reviewing the machine's progress for the first time in 5 years. Under pressure to modernize the nuclear arsenal, the agency could decide on a further shift toward stockpile stewardship. "Will the ignition program be squeezed out?" asks Mike Dunne, who directed Livermore's fusion energy efforts from 2010 to 2014. "The jury's out."

Fusion has long been held up as a carbon-free source of energy, fueled by readily available isotopes of hydrogen and producing no long-lived radioactive waste. But it remains a distant dream, even for the slow-burning, doughnut-shaped magnetic furnaces like the ITER project in France, which aims to achieve energy gain sometime after 2035.

NIF and other inertial fusion devices would be less like a furnace and more like an internal combustion engine, producing energy through rapid-fire explosions of the

Fire by trial

The National Ignition Facility has closed in on fusion ignition—getting more energy out than goes in—by altering its laser pulses and targets. It is even closer to the temperatures and pressures needed for an intermediate goal: a self-heating "burning plasma."



GRAPHIC: PRAV PATEL/LLNL. ADAPTED BY N. DESAI/*SCIENCE*

diminutive fuel pellets. Whereas some fusion lasers aim their beams straight at the pellets, NIF's shots are indirect: The beams heat a gold can the size of a pencil eraser called a hohlraum, which emits a pulse of x-rays meant to ignite fusion by heating the fuel capsule at its center to tens of millions of degrees and compressing it to billions of atmospheres.

But shots in the first 3 years of the ignition campaign only yielded about 1 kilojoule (kJ) of energy each, short of the 21 kJ pumped into the capsule by the x-ray pulse and far short of the 1.8 megajoules (MJ) in the original laser pulse. Siegfried Glenzer, who led the initial campaign, says the team was "overly ambitious" about reaching ignition. "We were overly reliant on simulations," says Glenzer, now at the SLAC National Accelerator Laboratory.

After the failed ignition campaign, NIF researchers beefed up their diagnostic instruments. They added more neutron detectors to give them a 3D view of where the fusion reactions were happening. They also adapted four of their laser beams to produce high-power, ultrashort pulses moments after the implosion in order to vaporize thin wires close to the target. The wires act as an x-ray flashbulb, able to probe the fuel as it compresses. "It's like a CAT scan," says planetary scientist Raymond Jeanloz of the University of California, Berkeley, who uses NIF to replicate the pressures at the core of giant planets such as Jupiter. (About 10% of NIF shots are devoted to basic science.)

With their sharper vision, researchers have tracked down energy leaks from the imploding fuel pellet. One came at the point where a tiny tube injected fuel into the capsule before the shot. To plug the leak, the team made the tube even thinner. Other leaks were traced back to the capsule's plastic shell, so researchers revamped manufacturing to smooth out imperfections of just a millionth of a meter. The improved diagnostics "really helps the scientists to understand what improvements are required," says Mingsheng Wei of the University of Rochester's Laboratory for Laser Energetics.

The team has also played with the shape of the 20-nanosecond laser pulses. Early shots ramped up in power slowly, to avoid heating the fuel too quickly and making it harder to compress. Later pulses ramped up more aggressively so that the plastic capsule had less time to mix with the fuel during compression, a tactic that boosted yields somewhat.

In the current campaign, begun in 2017, researchers are boosting temperatures by enlarging the hohlraum and the capsule by up to 20%, increasing the x-ray energy the capsule can absorb. To up the pressure, they're extending the duration of the pulse and switching from plastic capsules to denser diamond ones to compress the fuel more efficiently.

NIF has repeatedly achieved yields approaching 60 kJ. But Herrmann says a recent shot, discussed at the American Physical Society's Division of Plasma Physics meeting earlier this month, has exceeded that. Repeat shots are planned to gauge how close they got to a burning plasma, which is predicted to occur around 100 kJ. "It's pretty exciting," he says.

Even at maximum compression, the NIF researchers believe only the very center of the fuel is hot enough to fuse. But in an encouraging finding, they see evidence that the hot spot is getting a heating boost from frenetically moving helium nuclei, or alpha particles, created by the fusion reactions. If NIF can pump in just a bit more energy, it should spark a wave that will race out from the hot spot, burning fuel as it goes.

> **"How far can we get? I'm an optimist. We'll push NIF as far as we possibly can."**
> **Mark Herrmann**, Lawrence Livermore National Laboratory

Herrmann says the team still has a few more tricks to try out—each of which could drive temperatures and pressures to levels high enough to sustain burning plasma and ignition. They are testing different hohlraum shapes to better focus energy onto the capsule. They're experimenting with double-walled capsules that could trap and transfer x-ray energy more efficiently. And by soaking the fuel into a foam within the capsule, rather than freezing it as ice to the capsule walls, they hope to form a better central hot spot.

Will that be enough to reach ignition? If these steps don't suffice, boosting the laser energy would be the next option. NIF researchers have tested upgrades on four of the beamlines and managed to get an energy boost that, if the upgrades were applied to all the beams, would bring the full facility close to 3 MJ.

Those upgrades would, of course, take time and money NIF may not end up getting. Fusion scientists at NIF and elsewhere are anxiously awaiting the conclusions of the NNSA review. "How far can we get?" Herrmann asks. "I'm an optimist. We'll push NIF as far as we possibly can." ■

CANCER RESEARCH

Study homes in on 'exceptional responders' to cancer drugs

Project reveals clues to why some patients respond much better than others

By **Jocelyn Kaiser**

Although even the best cancer drugs don't buy much time for most people with advanced cancer, there are rare exceptions: patients whose tumors melt away and who remain healthy years later. Researchers have long dismissed these "exceptional responders" as unexplainable outliers. Now, an effort to systematically study these unusual patients is yielding data that could help improve cancer treatments.

The project, led by the U.S. National Cancer Institute (NCI), examined tumor DNA and the immune cells found around or within the cancers in 111 exceptional responders. In 26 of the patients, scientists found clues that may explain why a drug that didn't work for most people shrank the tumors for months or years. Some cases suggest new drug combinations could yield better outcomes for other patients. The study, published last week in *Cancer Cell*, "opens new avenues for treating comparable cancers in the wider population," says Dale Garsed of the Peter MacCallum Cancer Centre in Australia.

The former NCI director who launched the project in 2014 is also excited. "It is gratifying to see so much novel information from this initial survey of patients who have done unexpectedly well with existing therapies," says cancer biologist Harold Varmus of Weill Cornell Medicine. The results are "complex," he says, but they "promote unique hypotheses" and highlight the value of conducting genomic tests of patients' tumors in order to customize treatments.

Varmus was inspired in part by a bladder cancer patient who responded to a generally lackluster drug—an outcome traced to mutations in her tumor. He funded a team led by NCI's Louis Staudt and Percy Ivy, who

with Baylor College of Medicine and other academic collaborators combed through some 500 cases, largely cancers that had spread, for those that fit criteria for an exceptional response—such as having tumors shrink or disappear in response to a drug that worked for less than 10% of patients overall. After removing cases that lacked adequate medical data or tumor samples, they ended up with 111 patients. Their tumors went through a battery of genomic analyses and tests for immune cells.

In 26 cases, the data appeared to explain the patient's exceptional response. For example, a patient with brain cancer who was still alive after more than 10 years had received a chemotherapy drug called temozolomide that kills tumor cells by damaging their DNA. The patient's tumor had genomic changes that crippled two DNA repair pathways, which cells use to counter the drug's assault. Another patient, in remission from colon cancer for nearly 4 years after temozolomide treatment, also had two disabled DNA repair pathways, and had received a second drug that blocked a third. "Every backup system that would have reversed the damage was inactivated" in this person, Staudt says. These results suggest a cocktail of drugs, blocking different DNA repair pathways, could help some patients, Ivy says.

In other cases, tumors shrank after the patients received a drug that blocks a protein driving cell growth. The tumors had DNA changes that spurred activity of the protein's gene, which made the tumor cells highly dependent on the growth signal; as a result, the drug worked unusually well.

Still other exceptional responders had tumors that were infiltrated with high levels of certain immune cells. This suggests their immune systems were primed to swoop in and destroy tumors once a cancer drug started to kill some cells, Staudt says.

The findings suggest more genomic testing of tumors could improve treatments. But the DNA results may still be hard to interpret. Many cancers had confusing combinations of DNA and immune cell changes, and the evidence wasn't strong enough to draw firm conclusions in 85 of the cases. NCI is putting data for all 111 patients in an online database so that other researchers can study it and look for similar cases. "Maybe we missed something," Staudt says.

Researchers in North America, Europe, and Australia have launched similar exceptional responder projects, and NCI researchers hope some of these efforts can pool their data. Staudt would like to see a study of at least 1000 patients. "These are puzzles to be solved," he says. "I do think they teach us something." ■

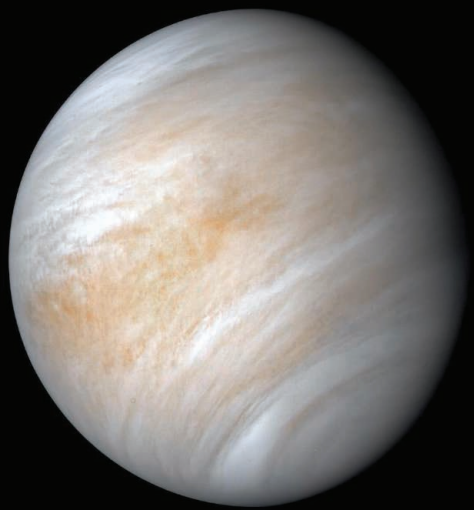
PHOTO: NASA/JPL-CALTECH

PLANETARY SCIENCE

Potential signs of life on Venus are fading fast

Amid mounting criticism, claims for cloud-borne phosphine are downgraded

By **Paul Voosen**



After reanalyzing data from Venus, researchers say they see much less phosphine, a possible sign of life.

The announcement in September took the world by storm: In radio emissions from Venus's atmosphere, researchers found signs of phosphine, a toxic compound that on Earth is made in significant amounts only by microbes and chemists. The unexpected detection could point to a microbial biosphere floating in the venusian clouds, the researchers suggested in *Nature Astronomy*. But almost immediately, other astronomers began to point out questionable methods or said they couldn't reproduce results.

Now, after reanalyzing their data, the original proponents are downgrading their claims. Phosphine levels are at least seven times lower than first claimed, the authors reported in a preprint posted on 17 November to arXiv. But the team still believes the gas is there, Jane Greaves, an astronomer at Cardiff University who led the work, said in a talk last week to a NASA Venus science group. "We have again a phosphine line."

The original observations were made in 2017 at the James Clerk Maxwell Telescope (JCMT) in Hawaii, and in 2019 at the Atacama Large Millimeter/submillimeter Array (ALMA) in Chile. In Venus's radio spectrum, Greaves and her colleagues detected an absorption line they attributed to phosphine. The researchers went to great lengths to remove confounding effects such as absorption by Earth's own atmosphere. But critics said such aggressive fixes made the discovery of a false positive more likely.

ALMA scientists have since found a new noise source: telescope calibration errors. After reanalyzing the ALMA data, Greaves said her team now finds phosphine at just 1 part per billion (ppb). That's still above levels that can be explained by natural processes such as volcanic eruptions or lightning strikes, Greaves said.

A study published last month in *Astronomy & Astrophysics*, led by Therese Encrenaz, an astronomer at the Paris Observatory, ruled out higher phosphine levels. Her team analyzed observations made in 2015 by NASA's Infrared Telescope Facility in Hawaii. Phosphine should have popped out if it had existed at levels above 5 ppb. "It's easy to see there's no phosphine line," Encrenaz says.

If the line does exist, it might not be due to phosphine, according to a critique submitted to *Nature Astronomy*. It argues that the dip in the JCMT spectrum can be explained by an overlapping absorption line from sulfur dioxide (SO_2), the gas that makes up most venusian clouds. The Greaves team concedes the point in its reanalysis. "We emphasize that there could be a contribution from SO_2," they write. But the width of the absorption line in the ALMA data suggests the feature isn't "solely SO_2," they write.

Just where any signal is coming from is also in dispute. ALMA is only sensitive to absorption from substances at altitudes above 70 kilometers (km), Encrenaz says. But the *Nature Astronomy* paper suggested the signal originated some 55 km up, in warmer, more hospitable cloud layers. "This is very difficult to conceive," Encrenaz says. Greaves and her co-authors argue in their reanalysis that ALMA is unable to capture the full width—and therefore depth—of the signal. "There is no empirical evidence that [phosphine] lies only above 70 km."

Colin Wilson, a co-author of the *Nature Astronomy* critique, says it's too early to say where the "phosphine roller coaster will end up." More observations at ALMA might settle the issue, he says. "Whether or not we find phosphine, we're likely to find something new." ■

COVID-19

Public needs to prep for vaccine side effects

"Take Tylenol and suck it up," says one researcher. Fever, aches show vaccine works

By **Meredith Wadman**

This summer, computational biologist Luke Hutchison volunteered for a trial of Moderna's COVID-19 vaccine. But after the second injection, his arm swelled up to the size of a "goose egg," Hutchison says. He can't be sure he got the vaccine and not a placebo, but within a few hours, Hutchison, who was healthy and 43, was beset by bone and muscle aches and a 38.9°C fever. "I started shaking. I had cold and hot rushes," he says. "I was sitting by the phone all night long thinking: 'Should I call 911?'"

Hutchison's symptoms resolved after 12 hours. But, he says, "Nobody prepared me for the severity of this."

He says the public should be better prepared than he was, because a subset of people may face intense, if transient, side effects, called reactogenicity, from Moderna's vaccine. Some health experts agree.

"Somebody needs to address the elephant: What about vaccine reactogenicity? While it's … not going to cause any long-term issues … how is that perception going to go with the public once they start receiving it?" asks Deborah Fuller, a vaccinologist at the University of Washington, Seattle, whose lab is developing second-generation RNA vaccines against COVID-19. She worries the side effects could feed vaccine hesitancy. "I feel like it's being glossed over."

Those concerns arise after a week of good news about coronavirus vaccines: Both Moderna and Pfizer, with BioNTech, announced their messenger RNA (mRNA) vaccines reached 95% efficacy in clinical trials of tens of thousands of people. The firms added that the trials showed no serious safety concerns.

Both vaccines consist of a snippet of genetic code directing production of the corona-virus' spike protein, delivered in a tiny fat bubble called a lipid nanoparticle. Some researchers suspect the immune system's response to that delivery vehicle is causing the short-term side effects.

Those transient reactions should not dissuade people from getting vaccinated in the face of a pandemic virus that kills at least one in 200 of those it infects, says Florian Krammer, a vaccinologist at the Icahn School of Medicine at Mount Sinai, who participated in Pfizer's trial. Sore arms, fevers, and fatigue are "unpleasant but not dangerous," he says. I'm not concerned about [reactogenicity] at all."

Most people will escape "severe" side effects, defined as those that prevent daily activity. Fewer than 2% of recipients of the Pfizer and Moderna vaccines developed severe fevers of 39°C to 40°C. But if the companies win regulatory approvals, they're aiming to supply vaccine to 35 million people worldwide by the end of December. If 2% experienced severe fever, that would be 700,000 people.

Other transient side effects would likely affect even more people. The independent board that conducted the interim analysis of Moderna's huge trial found that severe side effects included fatigue in 9.7% of participants, muscle pain in 8.9%, joint pain in 5.2%, and headache in 4.5%. In the Pfizer/BioNTech vaccine trial, the numbers were lower: Severe side effects included fatigue (3.8%) and headache (2%).

But that's a higher rate of severe reactions than people may be accustomed to. "This is higher reactogenicity than is ordinarily seen with most flu vaccines, even the high-dose ones," says Arnold Monto, an epidemiologist at the University of Michigan School of Public Health.

So front-line public health workers will need "to have a story that gets out in front of [stories like Hutchison's]—that responds to the way that people are going to try to make that a story about vaccine injury," says Bernice Hausman, an expert on vaccine controversy at the Pennsylvania State University College of Medicine.

Transparency is key, Hausman emphasizes. Rather than minimizing the chance of fever, for instance, vaccine administrators could alert people that they may experience a fever that can feel severe but is temporary. "That would go a significant way toward people feeling like they are being told the truth." Adds Drew Weissman, an immunologist at the University of Pennsylvania whose research contributed to both vaccines: "The companies just have to warn people: 'This is what you need to expect. Take Tylenol and suck it up for a day.'"

Hausman also sees a need to support people who have serious reactions. For example, people may need "a hotline with a nurse triaging … figuring out if you need to go to the hospital or not. Will your medical expenses be covered if you do? These are important questions."

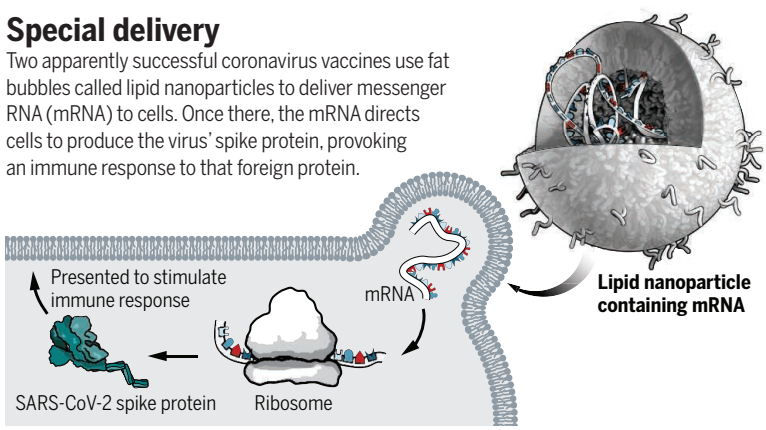
Both Moderna's and Pfizer/BioNTech's vaccines require two doses separated by several weeks. Reactogenicity is typically higher after a second dose, Weissman says. The side effects "mean the vaccine is working well. … [It] means you had such a good immune response to the first dose and now you are seeing the effects of that," he says.

"We suspect the lipid nanoparticle causes the reactogenicity, because lipid nanoparticles without mRNA in them do the same thing in animals," Weissman says. "We see production, in the muscle, of inflammatory mediators that cause pain, [redness], swelling, fever, flulike symptoms, etc."

Hutchison hopes better vaccines are on the way. Still, he says, "Given that COVID can kill or incapacitate people, everybody should bite the bullet and expect a rough night. … Get lots of naproxen." ■

Special delivery

Two apparently successful coronavirus vaccines use fat bubbles called lipid nanoparticles to deliver messenger RNA (mRNA) to cells. Once there, the mRNA directs cells to produce the virus' spike protein, provoking an immune response to that foreign protein.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Science's COVID-19 reporting is supported by the Pulitzer Center and the Heising-Simons Foundation.

In California, a child confronts the new reality of school in the time of COVID-19.

PHOTO: © PRENTICE C. JAMES/CSM VIA ZUMA WIRE

GRADE: INCOMPLETE

As COVID-19 soars in many communities, schools attempt to find ways through the crisis

By **Gretchen Vogel** *and* **Jennifer Couzin-Frankel**

Schools around the world are again the site of a large, and largely uncontrolled, experiment.

When schools from New Zealand to Norway to Japan reopened in April and May as the first wave of COVID-19 cases subsided, the virus stayed mostly at bay. Health and education officials cheered, having bet that the huge benefits of in-person schooling outweighed the risk of viral spread among children and teachers—and from schools to wider communities.

As a result, many places that had moved cautiously at first threw open classroom doors in August and September. Schools in the United Kingdom, Denmark, and the Netherlands shifted from cycling in small groups of students to full-size classes. Cities like Montreal that had kept schools shuttered welcomed students back inside. In Manaus, Brazil, a city with a COVID-19 death toll among the world's highest, more than 100,000 students returned to class-

Science's COVID-19 reporting is supported by the Pulitzer Center and the Heising-Simons Foundation.

rooms. Teenagers thronged hallways in Georgia, Iowa, and Texas. But the backdrop is very different now: In many areas, COVID-19 has surged to even higher levels than early in the year.

In July, *Science* examined the mostly encouraging lessons from the first reopenings, among schools in areas with minimal COVID-19 percolating. Now, scrutiny of school openings in countries where the virus is resurgent paints a more complex picture of the risks and how they might be managed.

The virus has exposed disparities between and within countries, and among the most unsettling are in schools. In many countries, such as India, Mexico, and Indonesia, most schools remain shut. In the United States, students enrolled in urban public schools from Los Angeles to Chicago, which in normal times may struggle to provide enough soap and toilet paper, continue to learn from home, whereas wealthy private schools have installed tents for outdoor learning and hired more teachers to shrink already-small classes. "The inequities from school to school are inexcusable and heart-wrenching," says Tom Kelly, head of the Horace Mann School, a private school in New York City that drew on many resources to open.

Early evidence, often gathered by researchers with children in school or a teacher spouse, suggests schools can stay open even in the face of significant community spread, given strong safety measures and political will. Many countries are closing restaurants, bars, and gyms, and begging residents to steer clear of social gatherings in bids to contain spread and keep schools open. Sometimes, that hasn't been enough: The Czech Republic, Russia, and Austria closed schools in the face of skyrocketing case numbers in October and early November.

"I think schools should close last," says Michael Wagner, a microbial ecologist at the University of Vienna who is part of a consortium of four universities studying the prevalence of the virus in Austria's schools. But he cautions that it's wishful thinking to suggest open schools can't fuel spread of the virus. Closing them can be "one of the most powerful measures we have, but also one of the most costly" to children.

In Austria, schools hung on until 17 November. However, other countries, such as South Korea and Australia, have closed many schools at the first sign of rising cases as authorities worked to quash even modest community transmission. "The conversation is fairly polarized right now as to whether schools should be open or should be closed," says pediatric infectious disease doctor Nisha Thampi of the Children's Hospital of Eastern Ontario. "People interpret the data one way or another to justify one end or another."

PHOTO: ANDY BUCHANAN/DIGITAL/AFP VIA GETTY IMAGES

How common are school outbreaks?

A gnawing anxiety for teachers and parents is the silent spread of virus through hallways and classrooms. Most schools have layers of protection such as mask requirements and physical distancing to impede an outbreak if a student or staff member brings COVID-19 into the building. But with virus cases surging in many communities, those guardrails are facing a stress test. "You're on pins and needles all the time," says Bradford Gioia, head of Montgomery Bell Academy in Nashville, Tennessee, an 800-person boys' school that runs from seventh to 12th grade.

So far, scientists say, school outbreaks appear less common than initially feared, although data are sparse. At Duke University, Danny Benjamin and Kanecia Zimmerman, both pediatricians and epidemiologists, are collaborating with more than 50 North Carolina school districts as well as local health departments to study COVID-19 in schools. The effort includes gathering data on clusters and single cases from a subset of six school districts—50,000 students and staff—in the first 9 weeks of in-person school. Community spread has been high, and the team has recorded 197 COVID-19 cases acquired outside of school and just eight confirmed to be "secondary transmission," or spread from one person to another inside a school. These numbers are almost certainly missing some asymptomatic infections. But Benjamin thinks the data support the approach North Carolina schools are using, with small classes and masks for everyone. He estimates that for every 10,000 people in school, there will be between one and five cases of secondary transmission about every 2 months.

Understanding why outbreaks happen can help schools shore up protections. In Benjamin's data, one outbreak traced to a group of teachers who carpooled unmasked to lunch. Inconsistent masking in a pre-K class in Tennessee was linked to a small outbreak there. Kelly, the Horace Mann head, was alarmed when three teachers tested positive in quick succession. He shut the middle and upper school for 2 weeks. But contact tracing suggested the cases were unrelated, and no one else tested positive during the closure.

Multiple cases among students rightly trigger worries about in-school spread. But young people's lives are intertwined, and the virus has plenty of chances to infect young people outside of school. "Kids have dance class, soccer class, a school bus," says Gail Carter-Hamilton, a nurse at the Philadelphia Department of Public Health

Students crowd the hallway of a school in Glasgow, U.K.

PHOTO: MATTHEW PUTNEY MEDIA

who provides support to local schools.

Sequencing the virus could help show whether multiple cases in a school are related. But it's almost never done, says Trevor Bedford, an expert in viral genome sequencing at the Fred Hutchinson Cancer Research Center. "It's really frustrating," he says.

Many experts lament that although health officials often tout low school case numbers, record keeping is inconsistent, as is transparency, particularly on outbreak investigations. "Show us the data," says Amy Greer, an infectious disease epidemiologist at the University of Guelph. She acknowledges that protecting individual privacy is a must—but deidentified data can still be studied and shared. "We have to be able to understand what the data we have tell us about school transmission," she says.

Montgomery Bell Academy offers testing on-site, and on many days, a few students are at home after testing positive, with a larger number quarantined because of close contact with someone infected. Most cases have been traced to outside activities, although three boys in a six-person study hall contracted the virus. Transparency about viral reach in a school can be difficult, but more and more Gioia comes down on the side of openness. "Most people appreciate honesty," he says.

Do open schools change risk perception?

Sports. Dating. Birthday parties. Orchestra practice. When schools open, students' other activities may be more likely to resume. And that has researchers worried.

"Families look to schools to communicate what's OK," says Jennifer Lerner, who studies the psychology of judgment and decision-making at Harvard University. Even when schools are doing all they can to mitigate COVID-19 spread inside their buildings, the mere act of opening can send an unintended message that mixing together is benign—and provide more opportunities to do so.

In considering how people assess risk, Lerner points to a seminal paper published in 1987 in *Science*, in which psychologist Paul Slovic at the University of Oregon wrote that the more uncertain and uncontrollable something seems, the riskier people deem it. Activities associated with school feel familiar and controllable and so may seem less risky, says Greer, of the University of Guelph. She found in a national survey that 40% of families have their children in at least one after-school activity, and some "have kids participating in extracurriculars 5 days a week."

To many people, it's especially hard to imagine school without sports. But the potential for the virus to spread there looms. In the United States, numerous outbreaks in August were traced to football practices. Early this month, the Iowa High School Girls Athletic Union hosted a state volleyball championship that brought together 20,000 fans and players in an indoor arena, as cases in the host city, Cedar Rapids, hit record highs and hospitals filled to capacity. In Canada, outbreaks have been linked to youth hockey; whether COVID-19 spread during play itself or gatherings with family and friends afterward is not known. "Schools are going to have a really tough time with sports," says Benjamin, of Duke University. "It's hard to make them safe."

Parties, too, have been an issue around the world. In Cape Town, South Africa, a gathering of high schoolers at a bar sparked an outbreak that ultimately infected more than 80 people. At Lerner's daughter's school, all families signed a pledge that, among other points, stressed compliance with state restrictions on social gatherings. When several girls held a party, they were required to quarantine at home for 2 weeks.

For health officials, a tricky balancing act may be in order. The message to school communities, Lerner says, needs to be, "There's such an enormous benefit to having schools open." And to sustain that, "We have to make sure that we are reducing risk in all other possible places."

High school athletes celebrated at a volleyball championship in Cedar Rapids, Iowa, where cases are surging.

Published by AAAS

In Bonn, Germany, children open windows to enhance the flow of outdoor air—and hopefully minimize any virus-laden aerosols in the classroom.

PHOTO: REUTERS/WOLFGANG RATTAY

How much fresh air is enough?

As temperatures drop in the Northern Hemisphere, many schools aren't radiating the comforting warmth they used to. The coronavirus pandemic has created a new routine: Open windows, no matter the weather.

In Germany, students wear coats and winter hats in class. In the United Kingdom, they're permitted to don extra clothes over uniforms. It's part of an effort to disperse any exhaled viral particles before someone can breathe them in.

"The air flow patterns you have inside make a lot of difference to your potential exposure," says Paul Linden, who studies fluid mechanics at the University of Cambridge and in September published a paper on how ventilation can help prevent viral spread. But with variability in weather, ventilation systems, and window size and placement, guidance is elusive. "It's very hard to be prescriptive," he says.

Rather than dive into calculations for every indoor space, scientists like Linden are embracing a simple alternative: high-quality carbon dioxide (CO_2) monitors, which cost as little as $100. Because CO_2 is exhaled as people breathe, it can serve as a proxy for how much exhaled air, and possible virus, has accumulated. Outdoors, the CO_2 concentration is about 400 parts per million (ppm). "What we've been recommending for schools is that CO_2 be below 700 ppm," even if everyone is wearing a mask, says Jose-Luis Jimenez, an aerosol scientist at the University of Colorado, Boulder, who models transmission risk. In a church in Washington state where a March outbreak among choir members originated, Jimenez's modeling suggests the CO_2 levels were about 2500 ppm.

Preliminary evidence from CO_2 monitors in schools suggests there's still work to be done. Linden found CO_2 levels in classrooms before the pandemic were about twice as high in winter as in summer. In Madrid, Javier Ballester, a fluid dynamics expert at the University of Zaragoza, found that, when windows are closed, a standard classroom with 15 students passes 1000 ppm in just 15 to 20 minutes.

Part of the challenge is practical. If children are "freezing cold, that's not going to help their learning experience," says Henry Burridge, a fluid mechanics specialist at Imperial College London. But Ballester's calculations suggest opening multiple windows by 15 centimeters each is likely sufficient. Germany is trying a different compromise: Classrooms can leave windows closed for 20 minutes and then open them wide for 5 minutes. (Berlin schools had to install tens of thousands of new handles on windows that had been secured shut.)

Some schools are adding professional-grade air filters to try to remove virus, and scientists are developing other creative solutions. Frank Helleis, a physicist at the Max Planck Institute for Chemistry, has developed a system of fume hoods over students' and teachers' desks. He and his colleagues are testing the setups at a school in Mainz, Germany, where his wife is a teacher. Cone-shaped hoods hang from the ceiling, connected to tubes that lead to a window, where a fan blows air outside. Warm air around a person rises, carrying exhaled aerosols to the hood, which collects and removes about 90% of aerosols before they can circulate, Helleis says. "It happens fast—within 10 seconds." Built from supplies available at home-improvement stores, the design is freely available.

Ballester, whose wife is also a teacher, has tried attaching a standard air filter to a fan. Initial tests show it's almost as effective as professional-grade machines. Most schools can't spend $500 per classroom, Ballester says, "but if it's $50 or $60, they might." Fans with filters that clean indoor air "work very well," Jimenez says, and are already used in regions with forest fires or air pollution. Such solutions may be especially valuable for classrooms with few or no windows—a common setup in U.S. schools that has fueled worries about reopening.

Does testing make a difference?

A student receives a coronavirus test on the outskirts of São Paulo, Brazil.

Since May, teenagers at the Gymnasium Carolinum, a school in Neustrelitz, Germany, have swabbed their own throats twice a week. Along with students, staff, and family members at six other schools and one day care, the teens send the samples to Centogene, a biotechnology company. The company's website trumpets, "School in spite of coronavirus, but safe!" It has run nearly 40,000 tests so far, which Volkmar Weckesser, Centogene's chief information officer, says have identified "multiple" cases and no outbreaks. "We can't say what would have happened if we hadn't been there," he acknowledges, but isolating the cases removed their chance to spark more infections.

Coronavirus testing in schools is scattershot, reflecting key uncertainties including how much children spread the virus and varying accuracy of tests. Some programs use tests for surveillance, as in New York City, with monthly tests on 10% to 20% of staff and students at many public schools. Wagner, of the University of Vienna, and colleagues are testing students and teachers in Austrian schools, and this fall found that roughly one in 250 people were infected without symptoms.

Blanket testing has its uses. But it takes up potentially scarce resources and can give a false picture, Duke's Benjamin warns. Even the most accurate tests can miss early stage infections. "Your public health interventions should assume that everybody's infected," he says.

Several studies are turning to tests to probe a big unknown: whether people with no symptoms spread SARS-CoV-2 at school. At the Charité University Hospital in Berlin, a team is working with 48 schools and day cares to regularly test staff, students, and their family members for both virus and antibodies. In Nashville, a team from Vanderbilt University descends every 2 weeks on a school, where children as young as age 4 spit into a cup and hand over their sample. "Does it even matter if a small percentage have detectable virus in saliva but they're not symptomatic and they're masked?" wonders Ritu Banerjee, one of three pediatric infectious disease specialists running the study, along with Sophie Katz and Kathryn Edwards. So far, she and her colleagues have collected four batches of more than 180 samples each and received test results on three of them. One positive case turned up in each batch, and none seemed to have infected anyone at school.

In Montreal, health surveillance expert David Buckeridge and pediatric infectious disease specialist Caroline Quach-Thanh at McGill University are planning an experiment in two schools. They want to know whether it's safe to shave a 14-day quarantine for close contacts to 7 days with a test before returning to school. The Montreal and Nashville researchers have something in common: Their own children attend the schools that welcomed the researchers in. That connection, Buckeridge says, was vital to making the work possible.

Should schools stay open as cases surge?

Early school reopenings spurred optimism. But many experts caution that that experience has limited relevance to high-transmission regions today. "The areas that reopened schools in the spring … had very, very little circulating virus in the community," says Matthew Oughton, an infectious disease doctor at McGill. But Denmark, for example, has had more than five times as many cases per week as in spring and France more than 10 times. Officials are facing difficult decisions about whether and when schools should close.

Scientific uncertainties aren't helping. Initial studies suggested children under age 10 were less likely than older ones and adults to catch and transmit SARS-CoV-2. But newer data have muddied the picture. In September, a study of families of U.K. health care workers found no difference in susceptibility by age. Antibody surveys in Brazil and southern Germany reported similar results. In a day care in Poland, five toddlers, none with symptoms, apparently infected nine family members. "I think asymptomatic infections have allowed children to fly under the radar," says Zoë Hyde, an epidemiologist at the University of Western Australia, Perth.

Yet some countries are finding they can suppress the virus while schools stay open. In mid-October, Ireland shut down many businesses and restricted people to within 5 kilometers of their home, but in-person instruction continued. Around the same time, the Netherlands closed restaurants, bars, and museums but also kept schools open. In both countries, new cases have fallen significantly.

Without clarity on in-school transmission, schools are hunting for signposts on when to throw in the towel and shift to remote learning. Officials in Iowa won't consider local school closures until a county's test positivity rate exceeds 15%, whereas New York City closed schools on 19 November, after reaching 3%. Other areas analyze virus levels in neighborhoods from which a school draws. Berlin authorities focus on what's happening inside a school, assessing weekly the numbers of new cases and people in quarantine.

Hyde and David Rubin, head of the PolicyLab at the Children's Hospital of Philadelphia, both think schools should probably shut if there are so many cases that contact tracing in the community is no longer feasible. In the Philadelphia area, "contact tracing is falling apart," Rubin says. In mid-November, with transmission rates escalating, he recommended area schools consider closing, especially for older children, until January 2021.

Numbers can act as a guide, but many say decisions around opening and closing schools are as much moral and political as they are scientific. "I don't think the right question is, at what point do we close schools," says Greer, of the University of Guelph. Instead, it's what do we need to do to keep schools open?

Like Ireland and the Netherlands, France, Spain, and Germany have kept students in classrooms while shutting other parts of public life. But U.S. cities including Boston and San Francisco have held off on or reversed public school openings, without major clampdowns on businesses. "A lot of school districts are not getting the support they need," including funding for safety measures, says Meagan Fitzpatrick, an infectious disease modeler at the University of Maryland School of Medicine.

Ultimately, she says, researchers can only offer so much. Until the pandemic subsides, likely with the help of a vaccine, officials, parents, and teachers face questions that lie outside science: "What do you mean by safe? And what levels of risk are you willing to accept for opening your school?" ■

With reporting by Linda Nordling and Emiliano Rodríguez Mega.

PHOTO: REUTERS/AMANDA PEROBELLI

GUIDING LIGHTS

Precise maps of millions of bright quasars show our place in the cosmos as never before

By **Joshua Sokol**

On a Friday at the end of 2018, the top brass of a NASA deep space mission convened for a tense meeting.

Hour by hour the New Horizons probe was hurtling toward a New Year's Day rendezvous with Arrokoth, an ancient, icy rock at the edge of the Solar System. The team had one last chance to send instructions for pointing the probe's cameras. Success would ensure in-frame pictures of Arrokoth, and the clues it held for how the planets formed. Failure would mean expensive pictures of an empty void.

The mission managers who gathered at New Horizons headquarters in suburban Maryland realized they had a "massive problem," says Marc Buie, a team member and planetary scientist at the Southwest Research Institute in Boulder, Colorado. Something was off in images already beamed back. Either the flying spacecraft or the orbiting rock was a teensy bit lost in a universe where nothing is nailed down.

The team debated what to fix. Some thought the probe's position, calculated from Earth-based measurements, was correct, in which case Arrokoth was in an unexpected place. But Buie believed the rock was right where it should be, which suggested thrusters had nudged the spacecraft itself a hair off course.

Buie was confident because he was tracking Arrokoth's position relative to an ultraprecise map of far-off beacons called quasars: cosmic lighthouses generated by black holes in distant galactic centers. But the map was largely untested, having just been released by a European Space Agency star-mapping satellite called Gaia. It was the basis of a brand-new celestial reference frame, a fixed, imaginary grid against which everything else moves, akin to lines of latitude and longitude on Earth. And Buie was gambling the Arrokoth flyby on that new grid.

For the past few decades, astronomers have based their celestial grid on radio observations of several thousand quasars. These radio beacons not only guide the pointing of telescopes, but they are also the bedrock of the reference frame for the spinning, bucking Earth. Without them, GPS devices would lose their accuracy and many ultraprecise studies of processes such as plate tectonics and climate change would be impossible. But observations of these beacons are costly and rely on radio telescopes.

By 2018, when New Horizons was approaching Arrokoth, Gaia had produced its own version of a reference frame, based on half a million quasars seen in the visible wavelengths most astronomers use, not radio. Buie persuaded the New Horizons team to trust the new framework. A correction based on the Gaia positions went up to the probe.

The team got it right: When the closest flyby images came back, Arrokoth was framed perfectly. "None of that would have happened if we hadn't had the Gaia catalog," Buie says. "It's a fundamental rewriting of how we do positional astronomy."

The rewriting has continued. Next week, on 3 December, Gaia will release, along with the latest data about billions of Milky Way stars, its newest reference frame, built from 1.6 million quasars scattered across the sky. "It is improved, larger, better, more beautiful," says François Mignard, an astronomer at the Côte d'Azur Observatory in France who leads Gaia's reference frame team.

THE GAIA REFERENCE FRAME is only the latest solution to a very, very old problem. From planets to comets to asteroids, much of the sky drifts from night to night. Studying these

PHOTO: J. PAULSON/ALAMY STOCK PHOTO

Quasars, generated by black holes in distant galaxies, are far more stable beacons than nearby stars.

objects would be hopeless without comparing them with points that stay still.

At first, the stars looked like trustworthy reference points. In the second century C.E., the Alexandrian astronomer Ptolemy revisited constellations his predecessor Hipparcos had observed some 3 centuries earlier. With his naked eyes, Ptolemy couldn't find any movement among the stars, which he assumed were fixed points on a sphere rotating around Earth.

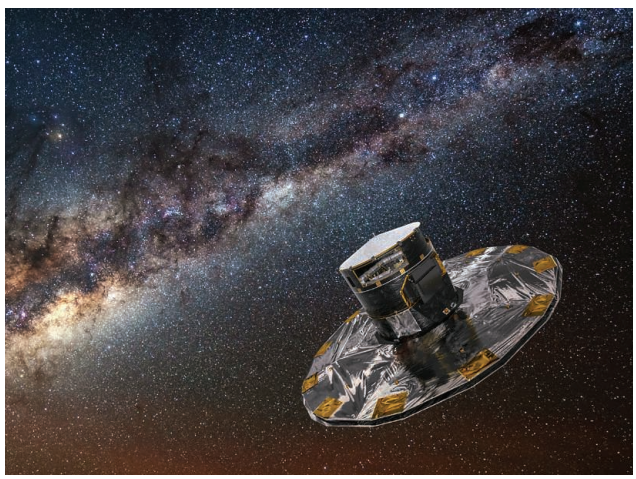
But by the 1700s, careful observations with telescopes proved the stars' apparent positions in the sky do shift over the years, as they move through the cosmos. In response, astronomers spent whole lifetimes building catalogs of stars far enough away or slow enough to mostly stay still.

The game changed again when astronomers started to observe quasars in the 1970s, using radio dishes on different continents to make hyperprecise measurements of their positions. Like stars, quasars appear as points of light. But they are billions of light-years away, so they barely budge within human lifetimes. Finally, the distant sky, not Earth, was the ultimate arbiter of where things are.

Today, the radio measurements feed into a global bureaucracy that maintains reference frames, imposing order on space in the same way astronomical observatories used to keep time. Many of the quasar measurements are accurate to about a hundred-millionth of a degree—smaller than the apparent size of a basketball on the Moon. They not only hold the sky in place, but also reveal jerks in Earth's rotation speed and wobbles in its axis that arise from earthquakes and hurricanes. The calculated tweaks are used in turn to correct GPS devices, which would otherwise lose track of Earth's spinning surface.

The growing number of rock-solid quasars, now in the thousands, also transformed interplanetary navigation. For decades, NASA tracked its spacecraft mainly by measuring their velocities as they flew away, which made it possible to calculate their distances from Earth. Their positions in other dimensions were only coarsely estimated by the sensors on the spacecraft. But after a pair of high-profile Mars failures in 1999, the agency added another method: It looks for quasars that are near the craft's current location in the sky—anchoring the probe to the reference frame. The approach has enabled subsequent bull's-eye landings on Mars and elsewhere, says Barry Geldzahler, who recently retired as NASA's navigation lead. "We make the hard things routine, and kind of boring."

For the many space scientists who work outside of radio wavelengths, however, a radio-based reference frame isn't so useful. Astronomers had spent decades trying to build up rival reference points in visible light. But quasars are faint specks at those wavelengths, and optical telescopes peering through Earth's blurry atmosphere struggled to match the precision of radio arrays.



Next week, Europe's star-mapping Gaia mission will release a new celestial reference frame, built from the positions of 1.6 million quasars.

Then the 2018 Gaia data set dropped, after the probe scanned the whole sky with sensitive space-based detectors. "Ninety-eight percent of that work was obliterated after the Gaia release," says Leonid Petrov, an astronomer at NASA's Goddard Space Flight Center who conducts radio quasar observations to build reference frames. "In 1 day, they became history."

THE CONSENSUS GRID for outer space—the third iteration of the official International Celestial Reference Frame, maintained by the International Astronomical Union (IAU)—still relies on radio quasars. But at the next IAU general assembly in 2021, Mignard says he plans to propose a multiwavelength system, with the optical quasar positions listed alongside the radio ones.

Tiny offsets between the two systems can already be seen, but they are not errors. They reflect astrophysical reality—and a tantalizing research opportunity. Quasars are powered by gas swirling around supermassive black holes at the centers of galaxies. As the gas circles the drain, it kicks out bright jets of plasma at nearly the speed of light. The radio telescopes are trained on the black hole itself, whereas Gaia picks up an average position between the black hole and the jets. No single telescope can distinguish between these locations. But the discrepancies between the radio and optical positions point to these fine details and offer a new way to investigate the physics at galaxy centers.

"If you're a fan of active galactic nuclei, this is a great time to be alive," says Bryan Dorland, an astronomer at the U.S. Naval Observatory. "The last time positional astronomy was exciting was around, like, Ptolemy. Right?"

Eventually, the Gaia data might even feed back into terrestrial position-finding systems, but not before lengthy studies and negotiations, says Manuela Seitz, of the German Geodetic Research Institute. "It's a long way between showing, OK, you can have an improvement if you use it to, OK, you now have products which are really consistent," she says.

To stay useful, the Gaia system will require tending. Right now, it has a firm handle on not just its quasars, but also more than 1 billion closer, drifting stars. These stars, anchored to the quasar-based grid, are useful guides for spacecraft with simple star trackers, or when no quasar is visible in a particular part of the sky. But the Gaia mission is set to end its vigil in 2025. After that, the stars will meander relative to background quasars unless astronomers dispatch a follow-up mission to remap the sky.

Meanwhile, the quasars themselves will drift glacially. Ultimately, Mignard says, high-precision reference frames of the future might require anchors even more stable than quasars: perhaps points on the cosmic microwave background, the afterglow of the big bang, which lies at the farthest observable distance in the cosmos.

Buie, for his part, plans to use the Gaia reference frame for as long as he can in his work pinpointing and studying tiny outer Solar System rocks. The Gaia data make it easier for him to calculate when and where to go on Earth to watch a star wink out as a remote object crosses in front of it—a so-called occultation event that relies on the momentary backlight to reveal details about the object.

He's also playing the same game he did with New Horizons and Arrokoth for Lucy, an upcoming NASA mission that plans to buzz past five small asteroids near the orbit of Jupiter. The extra precision of the Gaia system will help mission controllers home in precisely on their targets. "They don't think they need it, but they do," Buie says. ■

IMAGE: ESA/ATG MEDIALAB; ESO/S. BRUNIER

ECOLOGY

Surplus and stress control autumn timing

Climate change might cause an early shedding of leaves if trees have stored enough carbon

By **Christine R. Rollinson**

The presence of leaves on deciduous trees not only marks the changing of the seasons, but also defines the period of time in which trees store carbon from the air in leaves, wood, and roots. Warming winters causing earlier spring-leaf emergence is a widespread pattern of climate-change impacts across temperate tree species and their locations (*1*). Much more idiosyncratic is the timing of leaf senescence (deterioration), which offers no clear indication of whether future warming will cause an extended autumn growing season that leads to greater carbon-storage potential (*2*). On page 1066 of this issue, Zani *et al.* (*3*) demonstrate that there might be limits to how much carbon a tree can use or store in a single year. Further, if all carbon needs are met, leaves might senesce earlier rather than later in the autumn.

Leaf shedding at the end of the growing season in cold-temperate regions is an evolutionary adaptation to acute stressors (such as freezing temperatures and a lack of moisture available in frozen soils), which often limit species' survival and geographic distributions (*4*, *5*). The mechanistic impacts of mortality-inducing events on trees have likely contributed to senescence being commonly portrayed through a stress-based, supply-side framework. From the photosynthetic carbon supply perspective, leaves become a liability rather than an asset to a tree when the leaves do not have the available resources to be net carbon producers. This explains early leaf loss and mortality in response to stressors such as drought or heat (*6*, *7*). The well-understood role of leaves as carbon suppliers to the tree forms the basis for the common assumption that alleviating plant stress will allow leaves to persist and trees to continue storing carbon for a longer time period in the autumn.

The new work of Zani *et al.* demonstrates that focusing only on carbon supply is a flawed line of investigation because it entails the assumption that photosynthesis can be sustained indefinitely so long as resources

Center for Tree Science, The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL 60532, USA.
Email: crollinson@mortonarb.org

PHOTO: WILLI ROLFES/MINDEN PICTURES

Trees of the Teutoburg Forest in Lower Saxony, Germany, display colors of autumn leaf senescence.

are available, and thus, that senescence is triggered solely by photosynthetic limiters (such as short days or cold temperatures). Zani *et al.* revealed this shortcoming by demonstrating with observational, experimental, and modeling evidence that increased growing-season productivity can lead to earlier, rather than later, leaf senescence in temperate trees. These findings build on a growing body of literature focused on the demand side of the carbon cycle, in which growth and productivity are limited by the ability of different tissues, such as roots or stems, to use and store carbon (carbon "sinks").

In particular, sink limitation has been highlighted as a mechanism to explain the hotly debated lack of a clear carbon fertilization effect on tree growth. Carbon fertilization hypotheses are typically supply focused: Increasing atmospheric carbon from anthropogenic emissions increases the carbon available for plant photosynthesis, making carbon movement into leaves more efficient. Although carbon-enrichment studies and analyses of carbon isotopes stored in tree rings provide evidence for an increase in the efficiency of plant photosynthesis with increased atmospheric carbon, there is limited evidence that this translates into increased long-term tree growth and forest carbon storage (*8*). If more carbon is unnecessary because the other requirements for biological activity in sink tissues are met, trees will fail to grow more rapidly in response to additional carbon. From this demand-side perspective, leaves become unneeded and consequently senesce when the carbon they would fix in photosynthesis has nowhere to go because the available sinks are no longer taking more carbon.

Sink-limitation frameworks often lack mechanistic processes that cause various tissues or functions to no longer use or store carbon. Sink limitation can be an outcome of low productivity and allocation to tissues such as roots or wood that store carbon (*9*). However, in other cases, a highly productive early season can result in the production of an unsustainably large leaf area that leads to late-season stress and subsequent early senescence (*10*). Sink-based limitation from lack of carbon demand also might simply be the result of asynchronous tissue phenology arising from independent environmental controls on biological activity in different parts of a plant. For example, root growth in many, if not most, species occurs asynchronously with leaf production (*11*). When root and wood activity cease to use or store carbon, those tissues become unavailable as carbon sinks. Thus, leaves become costly to maintain, not because of their decreased carbon productivity potential but because there is no room for the carbon they make. Unavailability of potential sinks gives the tree a carbon economic incentive to senesce early despite suitable environmental conditions for photosynthesis.

In the end, the timing of leaf senescence and the autumn season is unlikely to be universally controlled by carbon source or sink dynamics alone (*12*). In portions of the temperate biome where warm summers increase photosynthesis (as is the case in Zani *et al.*) and climate change is projected to benefit plant growth, more sink-based limitation of forest productivity seems likely. In regions where stress from heat or low moisture currently causes marked declines in productivity, source-based models might adequately explain autumn leaf senescence. Given the geographic and temporal variability in limitations on tree productivity (*13*), it is reasonable to expect that the controls on autumn senescence across the temperate biome are likely to include a range from sink-based to source-based limitation. The sink-based models shown to perform well in Zani *et al.* improve the prediction of autumn senescence in some regions, but the universality of this pattern remains unknown.

Regardless of the drivers of autumn senescence, the potential productivity of temperate deciduous forests is not infinite. Forest carbon sequestration alone cannot be expected to keep pace with anthropogenic carbon emissions or climate change (*14*). The study of Zani *et al.* shows that the forest carbon sink is limited in ways that are not yet fully understood. Thus, whereas trees and forests remain one solution for mitigating the impacts of climate change, they cannot be the sole means of climate change response. A diverse portfolio of climate actions that include emissions reductions and tree conservation and planting is essential to mitigate anthropogenic carbon emissions and climate change. ■

REFERENCES AND NOTES

1. S. Piao *et al.*, *Glob. Change Biol.* **25**, 1922 (2019).
2. Z. A. Panchen *et al.*, *Ann. Bot.* **116**, 865 (2015).
3. D. Zani, T. W. Crowther, L. Mo, S. S. Renner, C. M. Zohner, *Science* **370**, 1066 (2020).
4. N. Delpierre *et al.*, *Ann. For. Sci.* **73**, 5 (2016).
5. I. Chuine, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 3149 (2010).
6. L. Chen *et al.*, *Nat. Clim. Chang.* **10**, 777 (2020).
7. Y. Xie, X. Wang, A. M. Wilson, J. A. Silander Jr., *Agric. For. Meteorol.* **250–251**, 127 (2018).
8. A. P. Walker *et al.*, *New Phytol.* (2020). 10.1111/nph.16866
9. J. Dolezal *et al.*, *Funct. Ecol.* **33**, 553 (2019).
10. A. S. Jump *et al.*, *Glob. Change Biol.* **23**, 3742 (2017).
11. K. Makoto, S. D. Wilson, T. Sato, G. Blume-Werry, J. H. C. Cornelissen, *Oikos* **129**, 643 (2020).
12. M. Mund *et al.*, *New Phytol.* **226**, 111 (2020).
13. F. Babst *et al.*, *Sci. Adv.* **5**, eaat4313 (2019).
14. B. W. Griscom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645 (2017).

10.1126/science.abf4481

ECOLOGY

Witnessing ice habitat collapse in the Arctic

Abrupt ice loss signals major changes ahead in a north polar conservation zone

By **Warwick F. Vincent**[1] *and* **Derek Mueller**[2]

Over a 2-day period at the end of July 2020, the Milne Ice Shelf underwent fracturing and collapse, losing 43% of its vast expanse as ice islands calved into the Arctic Ocean (see figure) (*1*). This sudden attrition of a thick Arctic ice shelf harboring diverse animal and microbial communities (*2*, *3*) is one of many recent events along Canada's far northern coastline that underscore the vulnerability of polar ice habitats to ongoing climate change. The observed impacts are of special concern given the conservation value of ecosystems in this High Arctic region.

The northern edge of Greenland and the Canadian Arctic Archipelago is the world's most northerly coastal zone. It contains the oldest, thickest sea ice in the Arctic Ocean and is considered the ultimate refuge for ice-dependent species in the rapidly warming North. In recognition of this global value, the Canadian government in partnership with the Qikiqtani Inuit Association designated a large portion of this "Last Ice Area" for interim protection as a Marine Protected Area (MPA) named Tuvaijuittuq, meaning "lasting ice" in Inuktitut (*4*). There is a strong reciprocal interaction between this marine system and the land margins of the High Arctic. These margins include Canada's northernmost land conservation region, Quttinirpaaq National Park, where extreme cold is critical to maintaining prominent ice features, including numerous glaciers, ice shelf remnants, ice-capped lakes, permafrost, and perennial snowbanks that feed water tracks across the polar desert landscape. Molecular analysis has shown that diverse microorganisms [including cold-water viruses (*5*)] thrive in all of these cryoenvironments.

[1]Centre for Northern Studies (CEN), Takuvik Joint International Laboratory, and Department of Biology, Université Laval, Quebec City, QC, Canada. [2]Water and Ice Research Laboratory, Department of Geography and Environmental Studies, Carleton University, Ottawa, ON, Canada. Email: warwick.vincent@bio.ulaval.ca

The Ward Hunt Ice Shelf, the largest platform of thick landfast ice in Arctic Canada, experienced several fracturing events in recent decades, and an automated camera (*6*) on Ward Hunt Island captured its collapse into ice islands over the summers of 2011 and 2012, leaving only two remnants on either side of its embayment. The initial fracturing in 2002 was accompanied by draining and complete loss of a freshwater lake overlying seawater (*2*). The lake dammed by the Milne Ice Shelf, with its distinctive microbial community (*7*), is the last of this ecosystem type in Canada and may be similarly close to disappearance.

Along the coast adjacent to the MPA, lakes that were once covered by thick, multiyear ice are now intermittently ice-free, with abrupt changes in their underlying ecosystems. The first observations of Ward Hunt Lake in the 1950s showed a summer ice cover of more than 4 m (*8*). Over subsequent decades, this perennial ice thinned, and three late-summer ice-out events have occurred in the past 10 years, including 31 days of ice-free conditions in 2016. That year also marked anomalously warm air temperatures at Ward Hunt Island and at Alert, Nunavut, 170 km east of the island, where melting degree-days (summation of daily mean temperatures above 0°C) were more than twice the 1951–2017 average (*8*). Weather extremes are becoming increasingly common throughout the Arctic and have the potential to drive lake ice, sea ice, and other cryoenvironments across thresholds, making them increasingly sensitive to year-to-year variations in climate.

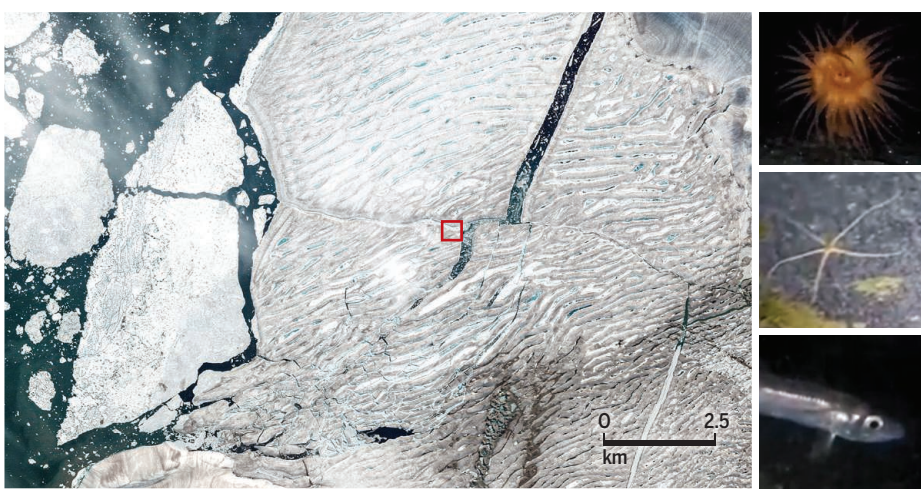
The surface area of glaciers along the northern margin of Ellesmere Island retracted by up to 5% between 1999 and 2015 (*9*), accompanied by reductions in supraglacial features such as cryoconite holes that provide microbial habitats. New cold-loving yeast species were recently discovered on Walker Glacier, 28 km east of Ward Hunt Island, but this cryohabitat is now undergoing accelerated melting and attrition (*10*). Summer meltwater on the surface of ice shelves and floating ice tongues collects in ponds and cryoconite holes that contain diverse microbial consortia. From 1998 to 2015, 51% of this habitat was lost as a result of episodic calving (*2*); little time may be left to fully characterize this component of Arctic biodiversity.

Thick multiyear landfast sea ice, some more than 50 years old, was a prominent feature of this coastline until recently, when all but the smallest remnants broke away (*2*). Mobile pack ice, including thick multiyear sea ice floes, are tightly concentrated in the coastal zone owing to the action of the Transpolar Drift and the clockwise Beaufort Gyre. The average thickness of this ice has decreased by 1.5 m since the 1970s, and this region is losing sea ice mass at double the rate of the Arctic Ocean (*11*). Ice supply to this region is highly variable and depends on currents that are readily modified by atmospheric circulation patterns, which may shift as the climate continues to warm. Increased ice advection and a marked drop in ice concentration in recent years underscore the vulnerability of this environment (*11*), which provides a cryohabitat for sea ice algae and associated biota (*12*). The implications of increased meltwater runoff from coastal lands to the marine ecosystem are unknown but could increase the occurrence of bloom-forming algae that may harm marine mammals (*13*)—this requires close monitoring.

The thick sea ice zone and its associated coastal land environments harbor diverse ecosystems that are only now being discovered. Conservation areas in this region, such as Tuvaijuittuq and Quttinirpaaq, are more important than ever for protection of these polar habitats and provide refuges that are free of additional stressors such as shipping and resource extraction. The Canadian MPA restrictions will expire in 2024 (*4*). This protection needs to be made permanent and extended across northern Canada and Greenland to include the entire thick ice zone and its associated coasts, despite the jurisdictional hurdles. Efforts to define, culture, and cryopreserve environmental microbiomes from this region should also be enhanced. Climate models underestimate the current acceleration of Arctic warming, and meeting the Paris Agreement carbon targets is essential to avoid abrupt future change (*14*). The ice attrition recorded over the past two decades shows that while this area of biological richness and ice-dependent ecosystems has been resistant to warming, it is not immune, and it can only be fully protected by the greenhouse gas mitigation that is urgently needed at a global scale. ■

Ice ecosystem collapse

An aerial view shows the breakup of the Milne Ice Shelf in late July 2020 (left) and animals found living in its ice cavity (right) at the site boxed in red (*3*). The ice shelf is in the so-called Last Ice Area, which observations and modeling (*15*) show as the band of thickest sea ice in the Arctic Ocean that occurs along the northern coast of Canada and Greenland.



PHOTOS: (LEFT) JOSEPH MASCARO/PLANET LABS INC.; (RIGHT) WATER AND ICE LABORATORY/CARLETON UNIVERSITY

REFERENCES AND NOTES

1. ECCC Canadian Ice Service (@ECCC-CIS), Twitter, 2 August 2020; https://twitter.com/ECCC_CIS/status/1290045576692658176.
2. L. Copland, D. Mueller, Eds., *Arctic Ice Shelves and Ice Islands* (Springer, 2017).
3. Water and Ice Research Laboratory, Carleton University, "Life Inside the Milne Ice Shelf" (2020); https://wirl.carleton.ca/research/ice/ice-shelves/calving-2020/#life-inside-milne-ice-shelf.
4. Fisheries and Oceans Canada, "Tuvaijuittuq Marine Protected Area (MPA)" (2019); www.dfo-mpo.gc.ca/oceans/mpa-zpm/tuvaijuittuq/index-eng.html.
5. M. Labbé *et al.*, *MSphere* **5**, e00334-20 (2020).
6. Northern Ellesmere Island in the Global Environment, *Nordicana D* 10.5885/45654CE-270458B4B50F45AA (2020).
7. M. Thaler *et al.*, *Front. Mar. Sci.* **3**, 275 (2017).
8. P. N. Bégin *et al.*, *Limnol. Oceanogr.* 10.1002/lno.11546 (2020).
9. A. White, L. Copland, *J. Glaciol.* **64**, 609 (2018).
10. M. Tsuji *et al.*, *Int. J. Syst. Evol. Microbiol.* **69**, 696 (2019).
11. G. W. K. Moore *et al.*, *Geophys. Res. Lett.* **46**, 11237 (2019).
12. B. A. Lange *et al.*, *Geophys. Res. Lett.* **46**, 10834 (2019).
13. N. Joli *et al.*, *Sci. Rep.* **8**, 9405 (2018).
14. E. Jansen *et al.*, *Nat. Clim. Chang.* **10**, 714 (2020).
15. K. S. Madsen *et al.*, *Polarforschung* **85**, 101 (2016).

ACKNOWLEDGMENTS

This contribution to Terrestrial Multidisciplinary distributed Observatories for the Study of Arctic Connections (T-MOSAiC) is supported by Sentinel North (CFREF), ArcticNet (NCE), the Polar Continental Shelf Program, and the Natural Sciences and Engineering Research Council. We thank C. Lovejoy and A. Culley for suggestions on the manuscript. W.F.V. and D.M. are contributors to the Canadian Science Advisory Secretariat review of Tuvaijuittuq, and W.F.V. is a member of the Quttinirpaaq planning team. The opinions in this Perspective are their individual views.

10.1126/science.abe4491

PHYSICS

Crisis and catharsis in atomic physics

Precise measurement of an atomic hydrogen transition resolves the proton size puzzle

By **Wim Ubachs**

The spectrum of the simplest atomic species, the hydrogen (H) atom with only a single electron, was initially described with Bohr's theory of 1913 and was refined with newer theories, from Schrödinger's quantum mechanics to Dirac's relativistic formalism and ultimately Feynman-Schwinger-Tomonaga's quantum electrodynamics (QED). The latter includes the effects of virtual particles that emerge from the vacuum and deals with the problem that the energy of a charged point particle is infinite. The comparison between theory and precise measurements of the H atom ran into a crisis in 2010, when measurements on muonic hydrogen (where muons replace electrons) (*1*) led to two different values of the size of the proton, r_p (see the figure). A decade-long period of the "proton size puzzle" spurred renewed experimental activity and many far-reaching hypothetical theories. On page 1061, Grinin *et al.* (*2*) report the precision measurement of the 1S-3S transition to help finally resolve this crisis.

The quantum-level structure of the hydrogen atom can be described for each state as the energy E that depends on quantum numbers n, l, and j and is the sum of three terms:

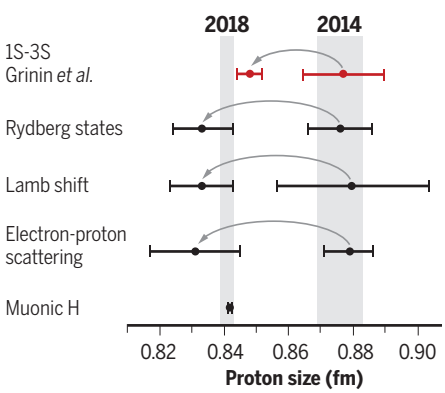
$$E(n,l,j) = E_{\mathrm{Bohr}} f(m_p, m_e) + E_{\mathrm{NS}}(r_p, n, l) + E_{\mathrm{QED}}(n,l,j) \quad (1)$$

where E_{Bohr} represents the Bohr structure, $f(m_p, m_e)$ is a small correction involving the proton mass m_p and the electron mass m_e, E_{NS} describes nuclear size effects, and E_{QED} represents the QED corrections. E_{Bohr} is proportional to the Rydberg constant R_∞. E_{NS} accounts for an electron penetrating the inner region of the extended proton, where it experiences an attractive force that deviates from the usual Coulomb force.

The entire level structure of the hydrogen atom can be cast by the two unknowns: R_∞, representing the energy scale of all atomic physics and of chemistry, and r_p. These two unknowns can be determined from several sets of two precise measurements. The measurement of the 1S-2S interval (*3*) stands out as the most precise (reaching 15-digit accuracy) because the upper 2S level exists for ~1 s and is not affected by the Heisenberg uncertainty principle. This result may then be combined with a second result. Various groups (*4*–*6*) have measured transitions from the long-lived 2S level to levels with nS, nP, or nD states (called Rydberg states), and these measurements were considered statistically independent and averaged to obtain a set of values for R_∞ and r_p. Another approach is in the measurement of the Lamb shift (the 2S-2P splitting) initially performed by Lamb and Retherford in 1948 and improved thereafter (*7*).

A historical correction

The right side shows the consistent picture of the proton size of 2010 before the muonic hydrogen measurement. The left side displays the current consistent picture, and arrows indicate the historical correction. Gray zones are the recommended values from the Committee on Data for Science and Technology (CODATA) for 2014 and 2018.



Electron-proton scattering measurements yield a fully independent determination of the proton size (*8*), although this nuclear physics approach requires a complicated extrapolation to a zero-momentum scattering vector. By 2010, these approaches led to a consistent set of values for R_∞ and r_p. In the figure, the values for r_p are plotted, and r_p was determined at 0.88 fm.

Because the Standard Model of physics assumes that the muon exhibits the same physics as the electron (lepton universality), except for being 207 times more massive and prone to radioactive decay, the level structure of muonic hydrogen (μH) should be calculable with Eq. 1. The heavier mass makes the muon overlap more with the nucleus, a reason why comparison between experiment and theory led to a very accurate value for r_p at 0.84 fm (*1*).

Thus, the atomic physics community was shaken to have the "consistent value" of r_p disagree with this μH measurement well beyond their error limits. This situation spurred enormous experimental activity on the precision spectroscopy of the H atom, as well as tests of the far-reaching hypothesis of the breakdown of lepton universality, and in turn the Standard Model of physics. Painstaking experiments remeasured the Lamb shift (*9*), excitation to a 4P Rydberg state (*10*), and electron-proton scattering (*11*), and all produced r_p of 0.84 fm.

However, in 2018, a continuous-wave laser experiment on the 1S-3S transition (*12*) still obtained the larger value. The situation is now finally resolved with the more accurate direct frequency-comb laser experiment performed by Grinin *et al.* on the same 1S-3S transition. The smaller r_p value supports a QED framework of electronic and muonic hydrogen described by the two constants, R_∞ and $r_p = 0.84$ fm. This struggle for consistency in atomic physics should provide an intriguing topic for historians and sociologists of science.

The H atom, and the interesting alternative of the He^+ ion, has only a single long-lived excited state, which hampers precision experiments on a set of quantum levels. Molecules such as the H_2^+ and HD^+ ions support many ro-vibrationally excited states that can all be subjected to precision measurements. Studies of these molecules have already led to a determination of the proton-electron mass ratio (*13*, *14*). The neutral H_2 molecule, which supports >300 ro-vibrational states with lifetimes of 1 week, could also be used in precision experiments for testing fundamental physics, possibly at 20-digit accuracy (*15*). ■

Vrije Universiteit, Amsterdam, Netherlands.
Email: w.m.g.ubachs@vu.nl

GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. R. Pohl *et al.*, *Nature* **466**, 213 (2010).
2. A. Grinin *et al.*, *Science* **370**, 1061 (2020).
3. A. Matveev *et al.*, *Phys. Rev. Lett.* **110**, 230801 (2013).
4. D. J. Berkeland *et al.*, *Phys. Rev. Lett.* **75**, 2470 (1995).
5. M. Weitz *et al.*, *Phys. Rev. A* **52**, 2664 (1995).
6. B. de Beauvoir *et al.*, *Phys. Rev. Lett.* **78**, 440 (1997).
7. E. W. Hagley, F. M. Pipkin, *Phys. Rev. Lett.* **72**, 1172 (1994).
8. J. C. Bernauer *et al.*, *Phys. Rev. Lett.* **105**, 242001 (2010).
9. N. Bezginov *et al.*, *Science* **365**, 1007 (2019).
10. A. Beyer *et al.*, *Science* **358**, 79 (2017).
11. W. Xiong *et al.*, *Nature* **575**, 147 (2019).
12. H. Fleurbaey *et al.*, *Phys. Rev. Lett.* **120**, 183001 (2018).
13. S. Alighanbari *et al.*, *Nature* **581**, 152 (2020).
14. S. Patra *et al.*, *Science* **369**, 1238 (2020).
15. W. Ubachs, J. C. J. Koelemeij, K. S. E. Eikema, E. J. Salumbides, *J. Mol. Spectrosc.* **320**, 1 (2016).

10.1126/science.abf0589

IMMUNOTHERAPY

Engineering cytokines and cytokine circuits

Learning to speak the secret language of immune cells could improve immunotherapies

By **Aileen W. Li**[1,2] *and* **Wendell A. Lim**[1,2,3]

Cytokines have far-reaching effects on the behavior of immune cells. Given their powerful roles, there has been a long history of trying to harness cytokines as therapeutic drugs for cancer and other diseases. However, there are several problems that severely limit the therapeutic use of cytokines, including their pleiotropic actions and systemic toxicity. Overcoming these issues to create the next generation of cytokine-based therapies will require sophisticated control over their spatial-temporal function. New approaches in protein and cell engineering are emerging that allow distinct and multiple levels at which to program cytokine regulation—from engineering individual cytokines, to cytokine-receptor pairs, and ultimately, more complex cytokine-sensing, -secreting, and -consuming cell circuits. These technologies may confer the ability to precisely sculpt the local cytokine environment, and by doing so, improve the potency of cytokine drugs and deepen our understanding of the language of cytokine communication.

The biological function of cytokines is broad, encompassing immune cell proliferation, death, activation, and inhibition. The effects of these secreted signaling molecules depends on their local concentration, which is driven by the rates of cytokine production, diffusion, and consumption. Cytokine-mediated cell-cell communication can be autocrine, paracrine, or endocrine. Together, these core features of cytokine communication are thought to shape the ecosystem of specific tissues or tumors. Perhaps most notable is how this set of secreted factors can achieve such diverse yet highly spatially coordinated physiological outcomes within the complex environment of the body.

Interleukins and interferons are cytokines that have clinical relevance in cancer. Direct infusion of cytokines into a tissue can have potent therapeutic effects—killing transformed cells in a tumor or stimulating the expansion and cytotoxic activities of host or adoptively transferred immune cells. So far, two cytokine drugs [interferon-α (IFN-α) and interleukin-2 (IL-2)] have been approved by the U.S. Food and Drug Administration (FDA) for the treatment of hairy cell leukemia, melanoma, and other cancers.

Nonetheless, there are fundamental problems that severely limit the therapeutic use of natural cytokines: short circulation half-life, off-target effects, and inherent pleiotropic functions. Clinically, repeated systemic administration of IL-2 at high doses is typically needed to achieve therapeutic response as a result of its short circulation half-life (the serum half-life of IL-2 is ~90 min). Most seriously, cytokines act as a double-edged sword—they target many cell types. Thus, for example, high dosing regimens of IL-2 elicit severe systemic toxicity because the cytokine accumulates not only in the disease tissue, but also in healthy bystander organs, where IL-2 induces severe adverse effects including vascular leak syndrome and pulmonary edema (*1*). IL-2 causes many changes in immune cells, some that may be desired and some that are therapeutically detrimental. IL-2 acts on multiple immune cells—it drives proliferation of effector T cells, but also stimulates T regulatory cells (T_{reg}) that cause suppressive outcomes. T_{reg} stimulation can promote tumor growth by serving as an IL-2 cytokine sink to deplete the growth factor necessary for effector T cell–mediated antitumor activity, and by directly disarming effector T cells.

Much of the existing efforts to engineer improved cytokines have focused on IL-2 because of its long history as a cancer therapeutic target. A more-conventional chemical strategy is to attach IL-2 to moieties such as polyethylene glycol (PEG) to extend its serum half-life. PEGylating IL-2 creates an IL-2 prodrug that mitigates rapid systemic activation upon administration by hindering receptor binding. Once the PEG is slowly released from the prodrug, the active free IL-2 becomes bioavailable over time (*2*). This modified IL-2 showed significantly longer serum half-life and was well tolerated in recent phase 1 trials in patients with advanced solid tumors (NCT02983045). Similarly, a PEGylated form of IFN-α showed longer half-life, and was approved by the FDA for the treatment of melanoma. Nonetheless, current evidence suggests that these approaches do not sufficiently address the major challenges of systemic toxicity and pleiotropic action.

Creating the next generation of cytokine-based therapies that address pleiotropic toxicity will require far greater control over cytokine function. Advances in protein and cell engineering are emerging that provide multiple new levels at which to program the time and space of cytokine-driven immune responses (see the figure). Protein engineering and screening have allowed investigators to more rationally engineer synthetic cytokines with selective bias toward a desired function. Pioneering studies using phage display screens created a human growth hormone (hGH) mutant that bound ~400 fold more tightly to its receptor than the wild-type form (*3*). Following this example, most cytokine engineering strategies use a combination of directed mutagenesis and library-based screens. For instance, an IL-2 mutant (BAY 50-4798) with reduced affinity for IL-2 receptor-β (IL-2Rβ) showed preferential activation for T cells over natural killer (NK) cells (which can cause toxicity) 3000-fold higher than the wild-type IL-2 (*4*). Even though this mutant was shown to be less toxic when tested in preclinical models, phase 1 trials in patients with metastatic melanoma or renal cancer failed to show significant benefit or reduction in side effects over IL-2 (*5*), likely because multiple IL-2–responsive populations can contribute to toxicity.

In a different approach, a superagonist form of IL-2, called "Super2," was engineered to have increased binding affinity for IL-2Rβ, rationalizing that it would preferentially trigger naïve T cells that are otherwise insensitive to IL-2 owing to their low expression of IL-2Rα (which stabilizes IL-2 interaction with IL-2Rβ). Indeed, Super2 showed superior expansion of cytotoxic T cells relative to regulatory T cells than did IL-2 and also reduced pulmonary toxicity in preclinical tumor models (*6*). Building on this work, an entirely new cytokine termed "neo-2/15" was designed in silico that signals through the shared chains of IL-2 and IL-15 receptors (the heterodimer of IL-2Rβ and IL-2Rγc) but has no binding sites for their respective private chains (IL-2Rα and IL-15Rα). Bypassing the private receptors allows neo-2/15 to preferentially signal to antitumor lymphocytes. In preclinical tumor models, neo-2/15 shows superior therapeutic activity to IL-2 and reduced toxicity (*7*). Recent efforts in cytokine engineering have also resulted in a "decoy-

[1]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, USA. [2]Cell Design Institute, University of California, San Francisco, CA, USA. [3]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA, USA.
Email: wendell.lim@ucsf.edu

resistant" IL-18 (DR-18), which maintains native IL-18 signaling but is impervious to inhibition by IL-18 binding peptide (IL-18BP), an endogenous secreted antagonist for wild-type IL-18 (*8*). Unlike IL-18, DR-18 showed effective antitumor effects in mice resistant to immune-checkpoint therapies. Clinical examples of designer cytokines include Pitrakinra, an engineered IL-4 variant that acts as an antagonist. In completed phase 2 trials, Pitrakinra showed some benefits for treating IL–4-associated asthma, with fewer adverse events (*9*).

A more radical emerging approach to limiting detrimental cytokine action is to engineer orthogonal cytokine-receptor pairs. This approach entails changing both the cytokine molecule and the way a target cell recognizes the engineered cytokine—an approach that fits well with engineered immune cell therapies [such as adoptive transfer of chimeric antigen receptor (CAR) T cells], which already involves a commitment to engineering a target effector immune cell. For example, to precisely target IL-2 functions to specific target T cells, an orthogonal IL-2/IL-2R pair (ortho2 and ortho2R, respectively) was developed (*10*). Ortho2 is a mutant IL-2 that can no longer bind to the native IL-2R; similarly, ortho2R is a mutant IL-2R that does not recognize the native IL-2. The ortho2/2R pair are engineered to only interact with each other. Thus, ortho2 stimulates only the complementary T cells that have been engineered to express ortho2R. Although engineering perfect orthogonal pairs with wild-type like potency remains a challenge, this pioneering work shows the power of the approach. In mouse models, ortho2 cytokine-receptor pairs show a high degree of specificity and orthogonality in vivo, suggesting that ortho2 may be a powerful tool to precisely control the proliferation of engineered cells while remaining inert to the endogenous immune system. This concept can be broadly applied to other cytokines and could be used to control CAR T cells or any other engineered therapeutic cell.

Moving beyond cytokines that already exist in nature, non-natural cytokines, or "synthekines," have also been described (*11*). These synthekines do not bind to natural cytokine receptor pairings, but instead assemble non-natural receptor heterodimers that lead to previously undescribed responses. Together, these important advances demonstrate the possibility of going beyond the proteins that our genomes naturally encode and open exciting therapeutic opportunities.

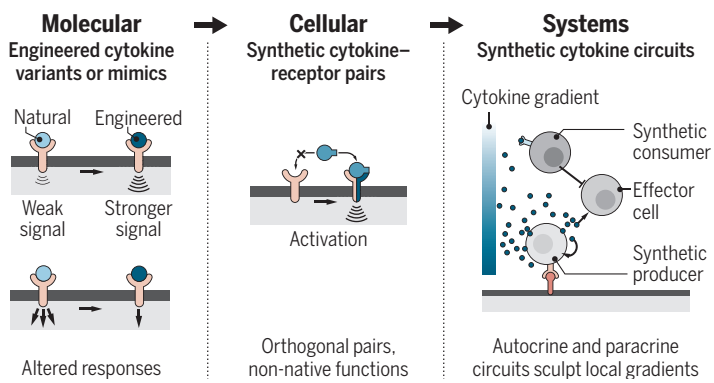
An even higher level of emerging engineering involves the creation of new multicellular cytokine systems and circuits. The highly localized action of cytokines originates from the ability of specific cells to read local signals that control both the production and consumption of cytokines—in essence, the immune system sculpts spatial gradients and niches using source and sink cells (in addition to effector cells that read the gradients) (*12*). With our mechanistic understanding of cellular biology and cell-cell communication, it may now be possible to rationally sculpt cytokine gradients, using cells that are synthetically engineered to act as sources and sinks. Engineering such gradients will likely require dynamic and discrete combinations of agonists and antagonists in the forms of cytokines, inhibitors, and cytokine receptors.

An early approach to engineering "source cells" has been to design CAR T cells to express proinflammatory cytokines (e.g., IL-12), either constitutively or under a CAR-controlled promoter (*13*). Engineering of cytokine consuming "sink" cells can also be a complementary powerful tool for sculpting cytokine milieus. A recent example of this nascent concept is engineered T cells constitutively expressing a nonsignaling membrane-bound IL-6R to effectively deplete IL-6 and thus reduce IL-6–mediated toxicity in mice (*14*). More controlled approaches are emerging in which modular sensing receptors, such as synNotch receptors (*15*), can be used to induce cytokine secretion or consumption in response to local disease or tissue antigen signals, yielding the potential of highly localized and programmable sink or source cells. Such engineered cellular delivery systems may offer one of the best ways to autonomously target and modulate local disease environments (including metastases) to drive antitumor responses and to remodel immunosuppressive responses, especially when combined with engineered autocrine or paracrine signaling that can locally amplify activity through positive feedback. Conversely, similar approaches could be used to create locally suppressed microenvironments in the case of autoimmunity.

These concepts are still at an early stage, and much experimental and theoretical validation are needed before they can reach the clinic. As a therapy, it is also important to critically evaluate the timing of intervention during disease progression. Ultimately, these multicellular cytokine control circuits may allow modulation of the expansion and death of engineered and host cells, and tuning the amplitude and duration of cytokines in a precisely targeted local environment. The future for engineered cytokines and cellular circuits is promising given that they could have many advantages compared to current cytokine therapies, including higher specificity, local and tissue-specific actions, and reduced off-target effects. It is expected that these strategies will be broadly impactful in treating other diseases involving inflammatory imbalances, such as autoimmunity, fibrosis, and tissue or wound regeneration. As more attempts are made to sculpt local cytokine microenvironments, deeper understanding of the language and grammar of cytokine-based communication will be gained. ■

Engineering cytokine communication

Emerging protein- and cell-engineering technologies may provide multiple levels at which to program cytokine-driven immune responses. These tools may lead to powerful therapeutics and improve understanding of cytokine-based communication.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. J. A. Klapper *et al.*, *Cancer* **113**, 293 (2008).
2. D. H. Charych *et al.*, *Clin. Cancer Res.* **22**, 680 (2016).
3. H. B. Lowman, J. A. Wells, *J. Mol. Biol.* **234**, 564 (1993).
4. A. B. Shanafelt *et al.*, *Nat. Biotechnol.* **18**, 1197 (2000).
5. K. Margolin *et al.*, *Clin. Cancer Res.* **13**, 3312 (2007).
6. A. M. Levin *et al.*, *Nature* **484**, 529 (2012).
7. D.-A. DSilva *et al*, *Nature* **565**, 186 (2019).
8. T. Zhou *et al.*, *Nature* **583**, 609 (2020).
9. S. Wenzel, D. Wilbraham, R. Fuller, E. B. Getz, M. Longphre, *Lancet* **370**, 1422 (2007).
10. J. T. Sockolosky *et al.*, *Science* **359**, 1037 (2018).
11. I. Moraga *et al.*, *eLife* **6**, e22882 (2017).
12. A. Oyler-Yaniv *et al.*, *Immunity* **46**, 609 (2017).
13. O. O. Yeku, T. J. Purdon, M. Koneru, D. Spriggs, R. J. Brentjens, *Sci. Rep.* **7**, 10541 (2017).
14. A. H. J. Tan, N. Vinanica, D. Campana, *Blood Adv.* **4**, 1419 (2020).
15. K. T. Roybal *et al.*, *Cell* **167**, 419 (2016).

ACKNOWLEDGMENTS

The authors are supported by the Howard Hughes Medical Institute (W.A.L.), the NIH (R01CA196277, P50GM081879, UC4DK116264, U54CA244438), and the Cancer Research Institute (A.W.L.). Thanks to members of the Lim lab and H. El-Samad. W.A.L is adviser to Allogene, a shareholder of Gilead, and has applied for patents on cytokine delivery circuits. A.W.L. is an employee of Lyell.

10.1126/science.abb5607

DEVELOPMENT

Staying connected under tension

A tyrosine kinase connects tension with strength of junctions between three adjacent cells

By **Srikala Raghavan**[1] *and* **Valeri Vasioukhin**[2]

The primary function of cells in all epithelial tissues is to form a physical barrier separating different compartments. Cell-cell adhesion structures seal neighboring membranes to form a barrier. Depending on the number of cells that are clumped together, these are called bicellular or tricellular junctions (*1*). During development and normal tissue homeostasis, cells in epithelia move and change their respective positions without complete loss of membrane connections and disruption of junctions. The mechanisms responsible for this junctional remodeling are not well understood. On page 1060 of this issue, Yu and Zallen (*2*) identify a signaling pathway that connects changes in mechanical forces at tricellular junctions with biochemical signals that either reinforce or weaken these adhesion structures. These results provide a molecular basis for understanding how cells sense and react to the mechanical changes in their environment.

Yu and Zallen studied early embryonic development in vivo using the fruit fly *Drosophila melanogaster*. During a process called convergent extension, cells in the surface epithelial sheet migrate and change their respective positions in a highly organized manner and without disrupting the tissue architecture (*3*). Epithelial cells are held together by adherens junctions that are connected through transmembrane cadherins, proteins that also link to the actomyosin cytoskeleton. Adherens junctions are used by cells to sense mechanical signals from their neighbors (*4*, *5*).

Cells use various mechanisms for mechanotransduction, to convert mechanical stimuli into biochemical signals. The mechanisms underpinning mechanotransduction at bicellular junctions in cultured cells have been identified. However, less is known about how these signals govern collective cell rearrangements in vivo and whether distinct mechanisms are involved at tricellular junctions (*5*). Tricellular adherens junctions are special: They need to withstand tremendous forces, and besides cadherins, they also contain the cell-cell adhesion protein Sidekick (*6*). In addition, actin-binding proteins Canoe and Polychaetoid are enriched at tricellular junctions. They interact with each other and act in parallel in junctional remodeling and maintenance of epithelial integrity during morphogenesis (*7*). Sidekick physically interacts with Polychaetoid and is necessary to organize Polychaetoid and Canoe at tricellular junctions and link these structures to the actin cytoskeleton (*8*). Canoe membrane localization is also regulated by the small guanosine triphosphatase (GTPase) Rap1 and its effector protein Dizzy (*9*).

Tricellular junctions nucleate actomyosin fibers, which help to limit elongation of larger cells and regulate the Hippo signaling pathway to favor the divisions of larger cells to control the final number and size of cells within a tissue (*10*). The mechanisms connecting mechanical forces with biochemical pathways that regulate tricellular junction assembly and function were not known. Yu and Zallen found that an increase in mechanical tension at the tricellular junction results in the Abelson tyrosine kinase (Abl)–mediated phosphorylation of Canoe, which directs it to tricellular junctions to promote connection to the actin cytoskeleton, thus increasing its mechanical strength during convergent extension movement (*3*).

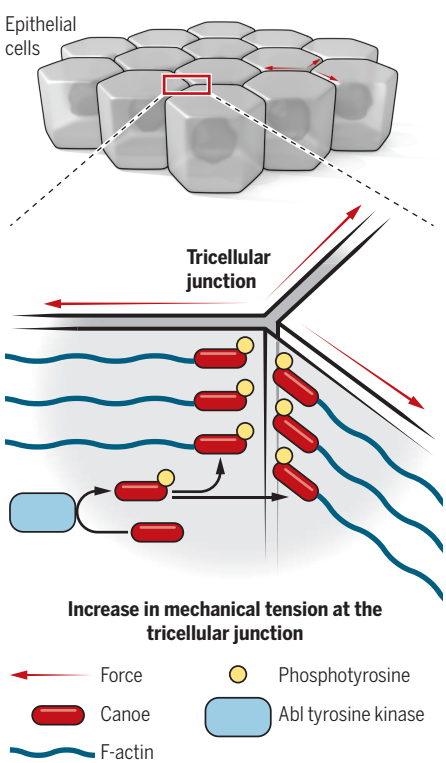
Abl was first discovered as a primary driver of leukemia and later was implicated in regulation of diverse cellular functions involving cytoskeletal changes, cellular motility, polarity, and adhesion (*11*). Imatinib, a prototypical example of targeted cancer therapy, inhibits abnormally activated ABL in the BRC-ABL fusion protein that causes chronic myeloid leukemia. Additionally, Abl was previously identified as a kinase that phosphorylates the cell adhesion protein β-catenin and directs bicellular junction remodeling during *D. melanogaster* convergent extension (*12*).

If Abl-mediated phosphorylation of Canoe is necessary for its function, what regulates this phosphorylation? This is the most exciting discovery of the study. Yu and Zallen found that Canoe phosphorylation and its localization at tricellular junctions are mechanosensitive. The dynamic and tension-regulated localization of Canoe to tricellular junctions was necessary for their remodeling and proper cellular movements during convergent extension.

Yu and Zallen observed that Canoe recruitment was strongly correlated with bursts of myosin accumulation at tricellular junctions and that its localization was uncoupled from myosin in the absence of Abl activity. This finding prompted the authors to ask if the localization of Canoe was regulated by force and if this was part of a force-regulated mechanism that strengthens adhesion at tricellular junctions under tension. They found that Canoe localization at tricellular junctions requires mechanical forces generated by actomyosin contractility (see the figure). Moreover, constitutive and mechanically uncoupled localization of Canoe to tricellular junctions resulted in their overstabilization and delayed or reduced cell rearrangements. Therefore, mechanosensitive coupling of

Regulating junction strength

The strengths of tricellular junctions are dynamically regulated in response to changes in mechanical tension during *Drosophila melanogaster* morphogenesis. Increased tension results in Abelson tyrosine kinase (Abl)–mediated phosphorylation of Canoe, which then accumulates at tricellular junctions and reinforces them by promoting connection with the actin cytoskeleton.



GRAPHIC: C. BICKEL/*SCIENCE*

[1]Institute for Stem Cell Science and Regenerative Medicine, GKVK Campus, Bellary Road, Bangalore 560065, India.
[2]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.
Email: srikala@instem.res.in; vvasiouk@fredhutch.org

Canoe localization to tricellular junctions was necessary for proper junction remodeling during morphogenesis. The loss of this localization caused disruption of tricellular junctions, whereas constitutive localization resulted in tricellular junction overstabilization and prevented normal morphogenesis.

Substantial knowledge has been accumulated concerning the mechanotransduction mechanisms at bicellular junctions (*5*). α-Catenin and vinculin, two actin-binding proteins that link junctional cadherins with the cytoskeleton, play a central role in this process. Increased actomyosin tension stretches α-catenin, which in turn increases the strength of its F-actin binding domain and, in addition, opens up the binding site for vinculin, which then accumulates at the junctions and reinforces their connection to the actin cytoskeleton (*13*, *14*). Vinculin can also be phosphorylated by Abl at Tyr^{822}, and this stabilizes its active open conformation, promotes association with F-actin, and strengthens cell-cell junctions (*15*). The connections between cytoskeletal tension, Abl-mediated Canoe phosphorylation, and its targeting to tricellular junctions found by Yu and Zallen provide important insights about the distinct mechanisms governing mechanotransduction.

As with all exciting discoveries, the study of Yu and Zallen raises new questions. What is the primary tension-dependent mechanosensor at tricellular junctions? How exactly is phosphorylated Canoe targeted to tricellular junctions? It is also unknown if additional scaffolding or signaling molecules are recruited to tricellular junctions in a tension-dependent manner. The availability of powerful in vivo model systems and tools will help to answer these questions and uncover additional mechanisms connecting mechanical forces to changes in cellular functions enabling adaptive control of morphogenesis. ■

REFERENCES AND NOTES

1. F. Bosveld *et al.*, *Curr. Opin. Cell Biol.* **54**, 80 (2018).
2. H. H. Yu, J. A. Zallen, *Science* **370**, eaba5528 (2020).
3. K. Z. Perez-Vale, M. Peifer, *Development* **147**, dev191049 (2020).
4. O. Klezovitch, V. Vasioukhin, *F1000 Res.* **4**, 550 (2015).
5. A. Angulo-Urarte, T. van der Wal, S. Huveneers, *Biochim. Biophys. Acta Biomembr.* **1862**, 183316 (2020).
6. T. Higashi, H. Chiba, *Biochim. Biophys. Acta Biomembr.* **1862**, 183143 (2020).
7. L. A. Manning *et al.*, *Mol. Biol. Cell* **30**, 1938 (2019).
8. A. Letizia *et al.*, *Dev. Cell* **50**, 313 (2019).
9. T. T. Bonello *et al.*, *Development* **145**, dev157941 (2018).
10. J. M. López-Gay *et al.*, *Science* **370**, eabb2169 (2020).
11. E. K. Greuber *et al.*, *Nat. Rev. Cancer* **13**, 559 (2013).
12. M. Tamada *et al.*, *Dev. Cell* **22**, 309 (2012).
13. S. Yonemura *et al.*, *Nat. Cell Biol.* **12**, 533 (2010).
14. C. D. Buckley *et al.*, *Science* **346**, 1254211 (2014).
15. J. L. Bays *et al.*, *J. Cell Biol.* **205**, 251 (2014).

ACKNOWLEDGMENTS

S.R. is supported by inStem core funds and Department of Biotechnology (DBT) grant BT/PR31418/BRB/10/1758/2019. V.V. is funded by U.S. National Cancer Institute grant R01CA234050.

10.1126/science.abf2782

A cross section cut reveals climate-sensitive tree rings in a Scots pine (*Pinus sylvestris* L.), one of the species used for soil-moisture reconstruction in inner East Asia.

PHOTO: PENG ZHANG

CLIMATE CHANGE

Tree rings circle an abrupt shift in climate

A recent drier-hotter climate stands out in the context of past climate variability

By **Qi-Bin Zhang**[1,2] *and* **Ouya Fang**[2]

An abrupt and substantial change in a climate system from one steady state to another—called a climate regime shift—has a considerable impact on ecosystems and society because of the abbreviated time period for adaptation (*1*, *2*). Recognizing the occurrence of regime shifts in climate is crucial for the timely development of effective policy responses to cope with the change. On page 1095 of this issue, Zhang *et al.* (*3*) reveal a recent regime shift to a drier and hotter climate throughout inner East Asia. The shift was detected with tree ring–based reconstructions of heat waves and soil moisture over the past 260 years.

Scientists have found it difficult to accurately identify a threshold over which the climate will irreversibly step into a new regime. The threshold can only be determined by understanding the natural range of climate variability over a time scale that is much longer than the new regime. Yet, long climate records are rarely directly available for a specific region of concern. In addition, onset of the regime shift is usually caused by various complex factors that interact in poorly understood ways. This lessens a scientist's ability to judge whether a change in climate is indeed a regime shift or merely an extreme climate event that will return to a normal state afterward.

[1]Institute of Tibet Plateau Ecology, Tibet Agricultural and Animal Husbandry University, Linzhi, Xizang, China. [2]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China. Email: qbzhang@ibcas.ac.cn

Tree rings have long been recognized as a useful proxy for past climate variations because of their special characteristics, such as precise dating, annual resolution, long time series, and climate sensitivity (*4*). The International Tree-Ring Data Bank (ITRDB) archives a wealth of tree-ring records contributed by scientists around the world. However, the use of these data for reconstruction of a specific climate variable over a large geographical region is not a straightforward procedure. Because a variety of environmental characteristics (such as macroclimate, local habitats, environmental disturbances, and tree physiology) influence tree growth, the signals from the macroclimate must be extracted while removing the nonclimate noise embedded in the rings.

The nonclimate noises from local habitats and disturbances are usually random in nature; thus, these noises can be canceled out by averaging tree-ring series from a large number of trees. Sustained growth reduction in trees at a sampling site should be interpreted carefully, because it might be the result of either prolonged adverse climate or damage of tree health after site disturbances or climate extremes (*5*). Also, trees do not merely respond passively to adverse climate but develop ecophysiological resilience to resist and recover from the influence; this makes the macroclimate signals more difficult to extract (*6*). In addition, sites where tree growth is not strongly sensitive to macroclimate should be excluded from regional climate reconstruction.

Zhang *et al*. compiled ITRDB tree-ring width data from 76 sites throughout inner East Asia and screened them for signals of two climate variables: summer heat-wave frequency and soil moisture content. They evaluated the linear relationships between series of tree-ring widths and observed climate records at interannual (2 to 4 years) and above-interannual (more than 4 years) time scales for each site. The authors selected two groups of tree-ring width data, each independently composed of 10 climate-sensitive sites, for reconstructing the history of the two climate variables. In the multicentury context of the reconstruction, the authors observed an unusual "drier-hotter" climate that emerged in the 1990s and continues to the present day. This finding opened a window through which one could tackle the physical mechanisms that trigger the onset of a climate regime shift.

Most studies on climate regime shifts have concentrated primarily on one climate variable at a time, to detect when and how it persistently exceeds the threshold (also called "tipping point") of its natural variability. Real-world climate systems involve many variables whose complex interactions could either cause negative feedbacks that reduce the probability that any single variable crosses its tipping point or cause positive feedbacks that increase the probability that multiple variables cross their tipping points (*7*). A challenge for scientists is to discover what circumstances cause the interaction of climate variables to generate negative or positive feedbacks. By investigating the characteristics of land-atmosphere interactions over inner East Asia, Zhang *et al*. demonstrated that the coupling of summer heat waves and droughts has been intensified by positive feedback loops since the late 20th century. This positive feedback was fueled by an enhanced soil-moisture deficit that might have nudged the climate over a tipping point.

Much remains to be discovered before scientists can fully understand the occurrence and mechanisms of climate regime shifts. Despite the concurrence of extreme droughts and heat waves reported in the literature (*8*–*10*), it remains unknown whether these phenomena resulted from a regime shift associated with global warming. In addition, the nature of the land-atmosphere interactions differs among geographical regions. For example, in addition to the shift to a drier-hotter climate in inner East Asia described by Zhang *et al*., other researchers have reported a hotter-wetter climate in the Tibetan Plateau (*11*).

Given that climate systems involve feedback among multiple variables, it is crucial to be able to predict when each variable will exceed its tipping point and when changes in these variables will lead to a domino effect, which has more severe deleterious consequences for ecosystems and society than do individual variable changes (*12*, *13*). Development of trustworthy long-term climate records is a critical prerequisite for accurate detection of temporal changes in the interactions of climate variables, for discovering past climate regime shifts, and for predicting thresholds of potential future shifts. ■

REFERENCES AND NOTES

1. T. M. Lenton *et al*., *Nature* **575**, 592 (2019).
2. S. Trumbore, P. Brando, H. Hartmann, *Science* **349**, 814 (2015).
3. P. Zhang *et al*., *Science* **370**, 1095 (2020).
4. J. H. Speer, *Fundamentals of Tree-ring Research* (Univ. Arizona Press, 2010).
5. Y.-M. Mu, Q.-B. Zhang, O. Fang, L. Lyu, P. Cherubini, *For. Ecol. Manage.* **480**, 118642 (2021).
6. O. Fang, Q.-B. Zhang, *Glob. Chang. Biol.* **25**, 245 (2019).
7. C. Raymond *et al*., *Nat. Clim. Chang.* **10**, 611 (2020).
8. V. Mishra, K. Thirumalai, D. Singh, S. Aadhar, *NPJ Clim. Atmos. Sci* **3**, 10 (2020).
9. Q. Kong, S. B. Guerreiro, S. Blenkinsop, X.-F. Li, H. J. Fowler, *Weather Clim. Extrem.* **28**, 100242 (2020).
10. D. G. Miralles, P. Gentine, S. I. Seneviratne, A. J. Teuling, *Ann. N. Y. Acad. Sci.* **1436**, 19 (2019).
11. D. L. Chen *et al*., *Chinese Sci. Bull.* **60**, 3025 (2015).
12. D. Lemoine, C. P. Traeger, *Nat. Clim. Chang.* **6**, 514 (2016).
13. P. C. Reid *et al*., *Glob. Chang. Biol.* **22**, 682 (2016).

10.1126/science.abf1700

GENOMICS

Sequencing perturbed cortex development

Single-cell genomics show promise for understanding neurodevelopmental disorders

By **Barbara Treutlein**[1] *and* **J. Gray Camp**[2]

Tissues are complex and contain a diverse compendium of cell types and states. Genetic disruptions can affect any one or all of these cells and, in doing so, sometimes give rise to disease. There are thousands of genes associated with human disease, and studying the effects of genetic disruption across all cell types within mammalian tissues remains difficult. Perturbations with single-cell sequencing readouts have promised extraordinary insight through increased throughput and enhanced resolution. On page 1057 of this issue, Jin *et al*. (*1*) combine in vivo genetic disruptions of genes associated with the risk of developing autism spectrum disorder (ASD) or neurodevelopmental delay (ND) with single-cell transcriptome sequencing to explore effects across diverse cells in the developing mouse cortex.

Over the past decades, cleverly designed in vivo genetic screens have provided insight into the mechanisms that underpin organism development (*2*). Enormous effort has been applied to increase throughput and enhance readouts of genetic screen technologies. Still, most screen designs do not provide insight into phenotype and mechanism in the same experiment, and certainly not across all cell types in a tissue. The combination of single-cell RNA-sequencing and CRISPR-Cas technologies (*3*–*5*) has provided researchers with tools to generate information-rich descriptions of cells (such as the transcriptome) and at the same time introduce and assess the

[1]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. [2]Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland.
Email: barbara.treutlein@bsse.ethz.ch; grayson.camp@iob.ch

effects of a genetic perturbation. Because each cell is sequenced independently, a mosaic tissue with diverse genetic manipulations can be created and the effect of many genes assessed in one experiment across all cell types and states. This is the advantage of the experimental design in Jin *et al.*

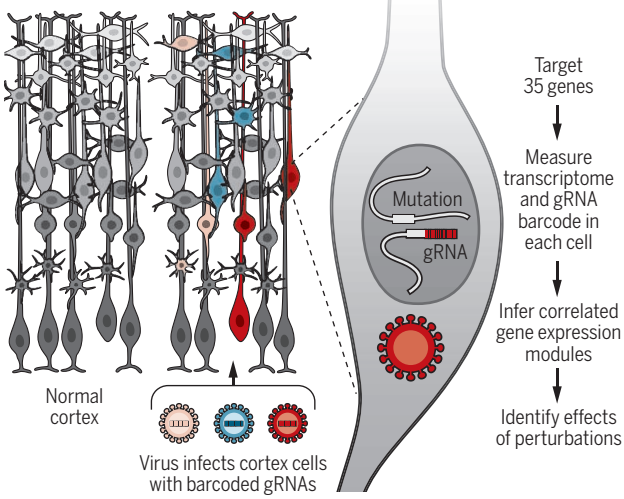
There are dozens of genes associated with ASD, and some of these genes are expressed in specific cell types, whereas others are broadly expressed in glial and neuronal cells. Through sparse infection of the developing cortex of mice constitutively expressing Cas9 (*6*) with lentiviral particles encoding guide RNAs (gRNAs) targeting these genes, Jin *et al.* used the Perturb-Seq method (*4*) to, in principle, rapidly assess all ASD-associated genes in one experiment using a single-cell RNA-sequencing readout (see the figure). Jin *et al.* find that recurrent gene regulatory modules may be perturbed by different gene mutations and also implicate additional cell types not previously linked to ASD or ND pathology.

However, the reality of single-cell perturbation screens is that the data are sparse, the effects are correlative, and interpretations can be hazy. With current technologies, the mutations introduced in the analyzed cells are not read out directly but are linked to the detection of the gRNAs or, in the case of Perturb-Seq, the detection of a barcode linked to the gRNA. The efficiency of CRISPR-mediated "gene deletions" can be estimated on the basis of quantifying the proportion of frameshifts on a per-guide basis; however, this does not currently scale well. For example, Jin *et al.* evaluated perturbations of 35 genes and provide frameshift efficiencies in the mouse cortex for two genes. Therefore, even though gene ablation efficiency is likely high, it is not measured for all gRNAs. It is possible that a gene deletion with no effect is not effectively mutated, and hence, important genes may not be identified in the screen.

The authors show that out of the 35 ASD- and ND-associated genes, 11 have a detectable effect. The data highlight the complexity of ASD and ND pathologies because both glial and neuronal cells showed perturbed expression networks. For example, perturbation of chromodomain helicase DNA-binding protein 8 (*Chd8*) affected expression modules in glial cells, including oligodendrocytes and astrocytes. *Chd8* is highly expressed in excitatory neurons in the mouse cortex, and previous work suggested that haploinsufficiency affects neuronal network development in the cortex (*7*, *8*) and striatum (*9*). However, there was no significant effect of *Chd8* perturbations on excitatory neurons. Thus, complex neuronal and glial interactions could lead to the diverse and variable pathologies associated with ASD and ND, and there could be specific effects within each brain region.

Understanding genetic disease risk

Perturb-Seq involves sparse infection of lentiviurses carrying guide RNAs (gRNAs) to ablate target genes followed by single-cell RNA-sequencing to identify the effects. Jin *et al*. targeted 35 genes associated with the risk of autism spectrum disorder and neurodevelopmental delay in the developing mouse cortex and uncovered common gene regulatory effects.



This study was a success because more is now known about the cell types affected by certain ASD risk–associated genes. However, it is still unclear whether at least 24 of the genes were deleted and have no effect. With Perturb-Seq, negatives are difficult to clearly interpret. Additionally, in contrast to classical in vivo genetic screens, the phenotypic effect of gene deletion is not obvious. Single-cell genomic perturbation measurements require layers of filtering, computational analyses, and statistical tests to assess the phenotypic effect. A simple differential expression analysis between gene-deleted and normal cells could not be performed because of the small number of cells for any given perturbation per cell type. The perturbation effect size was instead calculated on correlated expression modules across cell types compared with cells receiving control vectors. Data sparseness, technical variation between batches, uncertainty of genetic ablation status, and low cell numbers for certain cell types make it difficult to properly assess perturbation-associated phenotypes by using single-cell RNA-sequencing data. The authors validated certain key associations by using follow-up experiments in mice with single gene deletions, as well as with human brain tissues from patients with ASD. The data point to a small set of genes that are consistently dysregulated in humans and mice with ASD or ND.

Emerging approaches that use massively parallel cell barcoding (called split and pool barcoding) (*10*), compressed sensing through measuring portions of the transcriptome (*11*), or image-based in situ sequencing (*12*) could enhance screen scalability with single-cell resolution. However, it is not yet clear whether these emerging technologies will yield practical solutions for complex tissues while providing high-information content phenotypes. Furthermore, many disease-associated mutations are not loss of function, but the approach used by Jin *et al.* aims to introduce mutations that cause gene ablation. Precise CRISPR editors could be an approach for more nuanced dissection of disease-associated mutations.

The need to understand and to develop therapies for ASD and all other human diseases is urgent. Scientists are equipped with unprecedented tools. However, each scientist must decide how to spend valuable time and resources. Perhaps more pooled single-cell genomic perturbation screens in mice and human organoids should be the focus, or maybe disease-associated mutations should be introduced into models one by one for deep explorations of the phenotypic effect—or possibly both. Or perhaps that decision should be postponed until new technologies are developed. Progress often poses as many decisions as it provides answers. ■

REFERENCES AND NOTES

1. X. Jin *et al.*, *Science* **370**, eaaz6063 (2020).
2. E. Wieschaus, C. Nüsslein-Volhard, *Annu. Rev. Cell Dev. Biol.* **32**, 1 (2016).
3. P. Datlinger *et al.*, *Nat. Methods* **14**, 297 (2017).
4. A. Dixit *et al.*, *Cell* **167**, 1853 (2016).
5. D. A. Jaitin *et al.*, *Cell* **167**, 1883 (2016).
6. R. J. Platt *et al.*, *Cell* **159**, 440 (2014).
7. O. Durak *et al.*, *Nat. Neurosci.* **19**, 1477 (2016).
8. Y. Katayama *et al.*, *Nature* **537**, 675 (2016).
9. R. J. Platt *et al.*, *Cell Rep.* **19**, 335 (2017).
10. J. Cao *et al.*, *Science* **357**, 661 (2017).
11. D. Schraivogel *et al.*, *Nat. Methods* **17**, 629 (2020).
12. D. Feldman *et al.*, *Cell* **179**, 787 (2019).

ACKNOWLEDGMENTS

The authors thank R. Platt, J. Fleck, and Z. He for thoughts and comments.

10.1126/science.abf3661

GRAPHIC: A. KITTERMAN/*SCIENCE*

POLICY FORUM

TECHNOLOGY AND LAW

Protecting consumers from collusive prices due to AI

Price-setting algorithms can lead to noncompetitive prices, but the law is ill equipped to stop it

By **Emilio Calvano[1,2], Giacomo Calzolari[2,3,4], Vincenzo Denicolò[1,2], Joseph E. Harrington Jr.[5], Sergio Pastorello[1]**

The efficacy of a market system is rooted in competition. In striving to attract customers, firms are led to charge lower prices and deliver better products and services. Nothing more fundamentally undermines this process than collusion, when firms agree not to compete with one another and consequently consumers are harmed by higher prices. Collusion is generally condemned by economists and policy-makers and is unlawful in almost all countries. But the increasing delegation of price-setting to algorithms (*1*) has the potential for opening a back door through which firms could collude lawfully (*2*). Such algorithmic collusion can occur when artificial intelligence (AI) algorithms learn to adopt collusive pricing rules without human intervention, oversight, or even knowledge. This possibility poses a challenge for policy. To meet this challenge, we propose a direction for policy change and call for computer scientists, economists, and legal scholars to act in concert to operationalize the proposed change.

HUMAN COLLUSION

Collusion among humans typically involves three stages (see the table). First, firms' employees with price-setting authority communicate with the intent of agreeing on a collusive rule of conduct. This rule encompasses a higher price and an arrangement to incentivize firms to comply with that higher price rather than undercut it in order to pick up more market share. For example, in 1995 the CEOs of Christie's and Sotheby's hatched their plans in a limo at Kennedy International Airport, and in 1994 the U.S. Federal Bureau of Investigation secretly taped the lysine cartel as they conspired in a Maui hotel room. At those meetings, they spoke about charging higher prices and how to enforce them. Second, successful communication results in the mutual adoption of a collusive rule of conduct, which commonly takes the form of a collusive pricing rule. A crucial component of this pricing rule is retaliatory pricing: Each firm raises its price and maintains that higher price under the threat of a "punishment," such as a temporary price war, should it cheat and deviate from the higher price (*3*). It is this threat that sustains higher prices than would arise under competition. Third, firms set the higher prices that are the consequence of having adopted those collusive pricing rules.

The process that produces higher prices

	COMMUNICATIONS	COLLUSIVE PRICING RULES	HIGHER PRICES
Humans	Present, discoverable	Latent, not discoverable	Observable, difficult to evaluate
Algorithms	Not present	Latent, discoverable	Observable, difficult to evaluate

To determine whether firms are colluding, one could look for evidence at any of the three stages. However, evidence related to the last two stages—pricing rules and higher prices—is generally regarded as insufficient to achieve the requisite level of confidence in the judicial realm. Economists know how to calculate competitive prices given demand, costs, and other relevant market conditions. But many of these factors are difficult to observe and, when observable, are challenging to measure with precision. Consequently, courts do not use the competitive price level as a benchmark to identify collusion. Likewise, it is difficult to assess whether the firms' rules of conduct are collusive because such rules are latent, residing in employees' heads. In practice, we may never observe the retaliatory lower prices from a firm that cheated, even though that response is there in the minds of the employees and it is the anticipation of such a response that sustains higher prices. In other words, we might lack the events that produce the data that could identify the collusive pricing rules. Furthermore, even if one could observe what looks like a price war, it would be difficult to rule out innocent explanations (such as a decrease in the firms' costs or a fall in demand).

Given the latency of collusive pricing rules and the difficulty of determining whether prices are collusive or competitive, antitrust law and its enforcement have focused on the first stage: communications. Firms are found to be in violation of the law when communications (perhaps supplemented by other evidence) are sufficient to establish that firms have a "meeting of minds," a "concurrence of wills," or a "conscious commitment" that they will not compete (*4*). In the United States, more specifically, there must be evidence that one firm invited a competitor to collude and that the competitor accepted that invitation. The risk of false positives (i.e., wrongly finding firms guilty of collusion) has led courts to avoid basing their judgments on evidence of collusive pricing rules or collusive prices and instead to rely on evidence of communications.

ALGORITHMIC COLLUSION

Although the use of pricing algorithms has a long history—airline companies, for instance, have been using revenue management software for decades—concerns regarding algorithmic collusion have only recently arisen for two reasons. First, pricing algorithms had once been based on pricing rules set by programmers but now often rely on AI systems that learn autonomously through active experimentation. After the programmer has set a goal, such as profit maximization, algorithms are capable of autonomously learning rules of conduct that achieve the goal, possibly with no human intervention. The enhanced sophistication of learning algorithms makes it more likely that AI systems will discover profit-enhancing collusive pricing rules, just as they have succeeded in discovering winning strategies in complex board games such as chess and Go (*5*).

Second, a feature of online markets is that competitors' prices are available to a firm in real time. Such information is es-

[1]Department of Economics, University of Bologna, Bologna, Italy. [2]Center for Economic Policy Research, London, UK. [3]Department of Economics, European University Institute, Florence, Italy. [4]Department of Statistics, University of Bologna, Bologna, Italy. [5]Department of Business Economics and Public Policy, Wharton School, University of Pennsylvania, Philadelphia, PA, USA. Email: vincenzo.denicolo@unibo.it

sential to the operation of collusive pricing rules. In order for firms to settle on some common higher price, firms' prices must be observed frequently enough because sustaining those higher prices requires the prospect of punishing a firm that deviates from the collusive agreement. The more quickly the punishment is meted out, the less temptation to cheat. Thus, the emergence and persistence of higher prices through collusion is facilitated by rapid detection of competitors' prices, which is now often possible in online markets. For example, the prices of products listed on Amazon may change several times per day but can be monitored with practically no delay.

In light of these developments, concerns regarding the possibility of algorithmic collusion have been raised by government authorities, including the U.S. Federal Trade Commission (FTC) (*6*) and the European Commission (*7*). These concerns are justified, as enough evidence has accumulated that autonomous algorithmic collusion is a real risk.

The evidence is both experimental and empirical. On the experimental side, recent research has found the spontaneous emergence of collusion in computer-simulated markets. In these studies, commonly used reinforcement-learning algorithms learned to initiate and sustain collusion in the context of well-accepted economic models of an industry (*8*, *9*) (see the figure). Collusion arose with no human intervention other than instructing the AI-enabled learning algorithm to maximize profit (i.e., algorithms were not programmed to collude). Although the extent to which prices were higher in such virtual markets varied, prices were almost always substantially above the competitive level.

On the empirical side, a recent study (*10*) has provided possible evidence of algorithmic collusion in Germany's retail gasoline markets. The delegation of pricing to algorithms was found to be associated with a substantial 20 to 30% increase in the markup of stations' prices over cost. Although the evidence is indirect—because the authors of the study could not directly observe the timing of adoption of the pricing algorithms and thus had to infer it from other data—their findings are consistent with the results of computer-simulated market experiments.

A NEW POLICY APPROACH

Algorithmic collusion is as bad as human collusion. Consumers are harmed by the higher prices, irrespective of how firms arrive at charging these prices. However, should algorithmic collusion emerge in a market and be discovered, society lacks an effective defense to stop it. This is because algorithmic collusion does not involve the communications that have been the route to proving unlawful collusion (as distinguished from instances in which firms' employees might communicate and then collude with the assistance of algorithms, as in a recent case involving poster sellers on Amazon Marketplace). And even if alternative evidentiary approaches were to arise, there is no liability unless courts are prepared to conclude that AI has a "mind" or a "will" or is "conscious," for otherwise there can be no "meeting of minds" with algorithmic collusion.

> **"...pricing algorithms had once been based on pricing rules set by programmers but now often rely on AI systems that learn autonomously through active experimentation."**

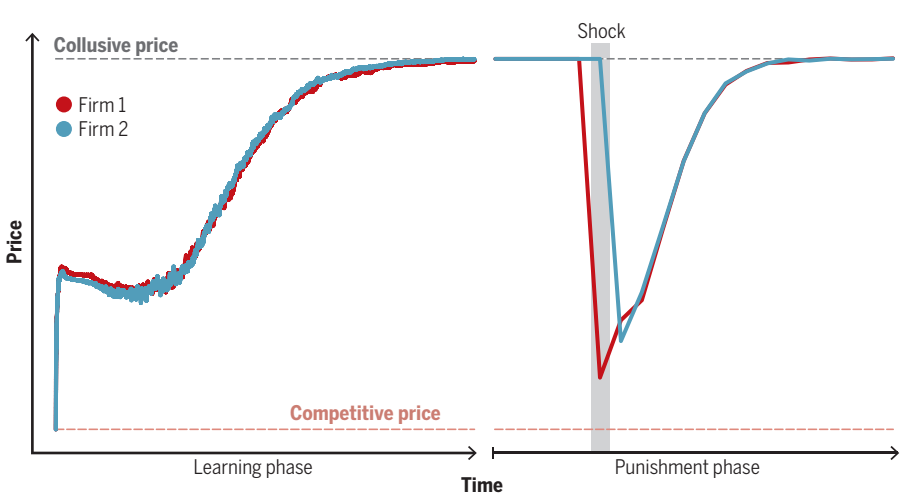
As a result, if algorithmic collusion occurs and is discovered by the authorities, currently it cannot be considered a violation of antitrust or competition law. Society would then have no recourse and consumers would be forced to continue to suffer the harm from algorithmic collusion's higher prices.

There is an alternative path, which is to target the collusive pricing rules learned by the algorithms that result in higher prices (*11*). These latent rules of conduct may be uncovered when they have been adopted by algorithms. Whereas a court cannot get inside the head of an employee to determine why prices are what they are, firms' pricing algorithms can be audited and tested in controlled environments. One can then simulate all sorts of possible deviations from existing prices and observe the algorithms' reaction in the absence of any confounding factor. In principle, the latent pricing rules can thus be identified precisely.

This approach was successfully used by researchers in (*8*) to verify that the pricing algorithms have indeed learned the collusive property of reward (keeping prices high unless a price cut occurs) and punishment (through retaliatory price wars should a price cut occur). To show this, the researchers momentarily overrode the pricing algorithm of one firm, forcing it to set a lower price. As soon as the algorithms regained control of the pricing, they engaged in a temporary price war, where lower prices were charged but then gradually returned to the collusive level. Having learned that undercutting the other firm's price brings forth a price war

Collusive pricing rules uncovered

After the two algorithms have found their way to collusive prices ("learning phase," left side), an attempt to cheat so as to gain market share is simulated by exogenously forcing Firm 1's algorithm to cut its price ("punishment phase," right side). From the "shock" period onward, the algorithm regains control of the pricing. Firm 1's deviation is punished by the other algorithm, so firms enter into a price war that lasts for several periods and then gradually ends as the algorithms return to pricing at a collusive level. For better graphical representation, the time scales on the right and left sides of the figure are different.



GRAPHIC: N. CARY/*SCIENCE* FROM CALVANO *ET AL.* (*8*)

(with the associated lower profits), the algorithms evolved to maintain high prices (see the figure).

It may seem paradoxical that collusion can be identified by the low retaliatory prices, which could be close to the competitive level, rather than by the high prices that are the ultimate concern for policy. But there are two important differences between retaliatory price wars and healthy competition. First, in the absence of the low-price perturbation, the price war remains hypothetical in that it is a threat that is not executed. Second, the price war shown in the figure is only temporary: Instead of permanently reverting to the competitive price level, the algorithms gradually return to the pre-shock prices. This is evidence that the price war is there to support high prices, not to produce low prices.

Focusing on the collusive pricing rules is the key to identifying, preventing, and prosecuting algorithmic collusion (see the table). Policy cannot target the higher prices directly, nor can it target communications as they may not be present (unlike with human collusion). But the retaliatory pricing rules may now be observable, as firms' pricing algorithms can be audited and tested. We therefore propose that antitrust policy shift its focus from communications (with humans) to rules of conduct (with algorithms).

Making the proposed change operational involves a broad research program that requires the combined efforts of economists, computer scientists, and legal scholars. One strand of this program is a three-step experimental procedure. The first step creates collusion in the lab for descriptively realistic models of markets. As the competitive price would be known by the experimenter, collusion is identified by high prices. Having identified an episode of collusion, the second step is to perform a post hoc auditing exercise to uncover the properties of the collusive pricing rules that produced those high prices.

Some progress has been made on the identification of collusive rules of conduct adopted by algorithms, but much more work needs to be done. Economics provides several properties to watch out for. Of course, there is the retaliatory price war discussed above, which is what existing research has focused on (*8*, *9*). Another property is price matching, whereby firms' prices move in sync: one firm changing its price and the other firm subsequently matching that change. Price matching has been documented for human collusion in various markets, but we do not yet know whether algorithms are capable of learning it. A third property is the asymmetry of price responses. When firms collude, they typically respond to a competitor's price cut more strongly—as part of a punishment—than to a price increase. No such asymmetry is to be expected when firms compete.

"Another path is to make firms legally responsible for the pricing rules that their learning algorithms adopt. Firms may then be incentivized to prevent collusion by routinely monitoring the output of their learning algorithms."

The aforementioned properties are based on economic theory and studies of human collusion. Learning algorithms may devise rules of conduct that neither economists nor managers have imagined (just as learning algorithms have done, for instance, in chess). To investigate this possibility, computer scientists might develop algorithms that explain their own behavior, thereby making the collusive properties more apparent. One way of doing so is to add a second module to the reinforcement-learning module that maximizes profits; this second module maps the state representation of the first one onto a verbal explanation of its strategy (*12*).

Having uncovered collusive pricing rules, the third step is to experiment with constraining the learning algorithm to prevent it from evolving to collusion. Computer scientists are particularly valuable here, given that they are involved in similar tasks such as trying to constrain algorithms so that, for instance, they do not exhibit racial and gender bias (*13*).

Once the capacities to audit pricing algorithms for collusive properties and to constrain learning algorithms so that they do not adopt collusive pricing rules have been developed, legal scholars are called upon to use that knowledge for purposes of prosecution and prevention. One route is to make certain pricing algorithms unlawful, perhaps under Section 5 of the FTC Act, which prohibits unfair methods of competition. In the area of securities law, the 2017 case *U.S.* v. *Michael Coscia* made illegal the use of certain programmed trading rules and thus provides a legal precedent for prohibiting algorithms. Another path is to make firms legally responsible for the pricing rules that their learning algorithms adopt (*14*). Firms may then be incentivized to prevent collusion by routinely monitoring the output of their learning algorithms.

These are some of the avenues that can be pursued for preventing and shutting down algorithmic collusion. There are several obstacles down the road, including the difficulty of making a collusive property test operational, the lack of transparency and interpretability of algorithms, and courts' willingness and ability to incorporate technical material of this nature. In addition, there is the challenge of addressing algorithmic collusion without giving up the efficiency gains from pricing algorithms such as the quicker response to changing market conditions. As authorities prepare to take action (*15*), it is vital that computer scientists, economists, and legal scholars work together to protect consumers from the potential harm of higher prices. ■

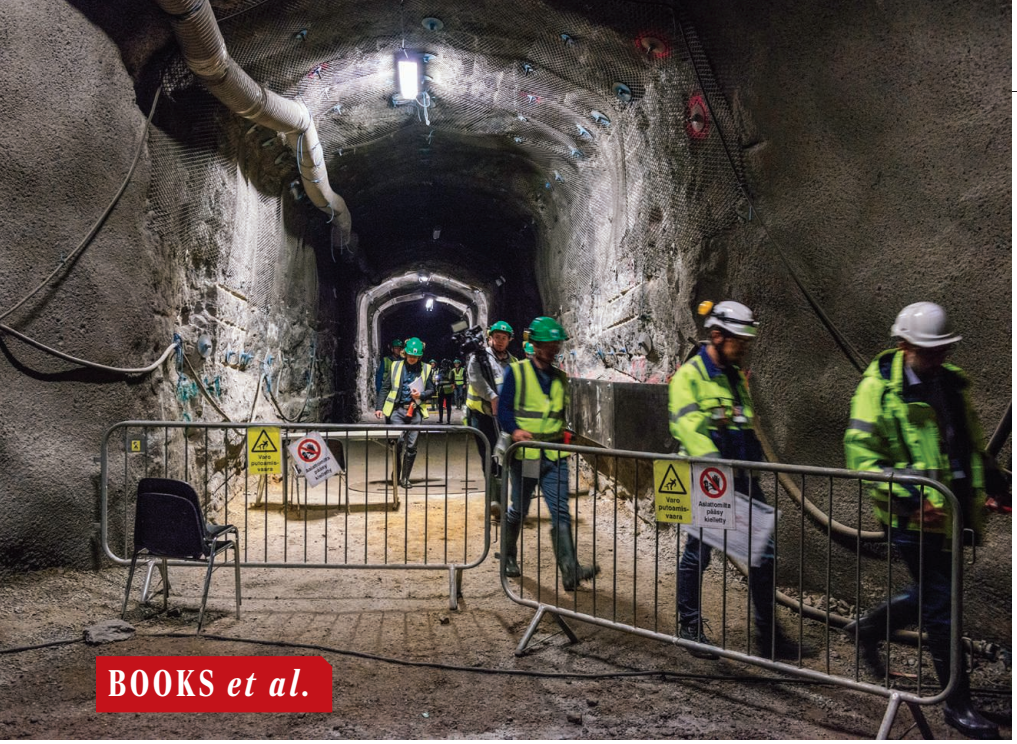
REFERENCES AND NOTES

1. A. Ezrachi, M. Stucke, *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy* (Harvard Univ. Press, 2016).
2. S. Mehra, *Minn. Law Rev.* **100**, 1323 (2016).
3. J. Harrington, *The Theory of Collusion and Competition Policy* (MIT Press, 2017).
4. L. Kaplow, *Competition Policy and Price Fixing* (Princeton Univ. Press, 2013).
5. D. Silver *et al.*, *Science* **362**, 1140 (2018).
6. "The Competition and Consumer Protection Issues of Algorithms, Artificial Intelligence, and Predictive Analytics," Hearing on Competition and Consumer Protection in the 21st Century, U.S. Federal Trade Commission, 13–14 November 2018; www.ftc.gov/news-events/events-calendar/ftc-hearing-7-competition-consumer-protection-21st-century.
7. "Algorithms and Collusion—Note from the European Union," OECD Roundtable, June 2017; www.oecd.org/competition/algorithms-and-collusion.htm.
8. E. Calvano, G. Calzolari, V. Denicolo, S. Pastorello, *Am. Econ. Rev.* **110**, 3267 (2020).
9. T. Klein, "Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing," Amsterdam Law School Research Paper 2018-15 (2019).
10. S. Assad, R. Clark, D. Ershov, L. Xu, "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market," CESifo Working Paper No. 8521 (2020).
11. J. Harrington, *J. Compet. Law Econ.* **14**, 331 (2018).
12. Z. C. Lipton, *ACM Queue* **16**, 30 (2018).
13. P. S. Thomas *et al.*, *Science* **366**, 999 (2019).
14. S. Chopra, L. White, *A Legal Theory for Autonomous Artificial Agents* (Univ. of Michigan Press, 2011).
15. European Commission, document Ares(2020)2877634.

ACKNOWLEDGMENTS

The paper benefited from detailed and insightful comments by three anonymous reviewers. All authors contributed equally. The authors declare no competing interests.

10.1126/science.abe3796



A tour proceeds in a Posiva demonstration tunnel in Olkiluoto, Eurajoki, Finland, in 2017.

PHOTO: ANTTI YRJONEN/NURPHOTO/GETTY IMAGES

BOOKS *et al.*

ANTHROPOCENE

Learning to think long-term

Lessons from nuclear waste management could help us move beyond shortsighted climate thinking

By **Jacob Darwin Hamblin**

Before we were worried about fossil fuels and plastic pollution, nuclear waste stretched our minds to think about our place in geological time. When asked about the most challenging aspect of her work, one expert charged with finding a final resting place for this waste observed, "We have to build canisters that are supposed to last hundreds of thousands of years. Nothing has lasted so long, nothing."

In *Deep Time Reckoning*, cultural anthropologist Vincent Ialenti would have us look to such experts for inspiration. Rather than allowing ourselves to be overwhelmed by the magnitude of humanity's effects on the planet, he proposes that we see nuclear waste storage as he does, as a case study for how to extend our intellects and acquire habits associated with extremely long-term thinking. "I am a techno-optimist," he writes, "but just barely."

Deep Time Reckoning: How Future Thinking Can Help Earth Now
Vincent Ialenti
MIT Press, 2020. 208 pp.

Because of the myriad challenges of the Anthropocene, Ialenti tells us, deep time reckoning is no longer the province solely of geologists, evolutionary biologists, or astrophysicists. It is now our collective responsibility, and we need the tools to do it.

Ialenti spent time in Finland between 2012 and 2014 observing employees of Posiva, a nuclear waste management firm. To obtain permission to operate a permanent storage facility in the Olkiluoto area, Posiva needed to demonstrate that the nuclear waste would pose no threat to future generations.

The project was depicted in a 2010 Danish documentary, *Into Eternity*, which Ialenti thinks was too focused on existential dread. "My informants never quoted philosophical works on the sublime or the uncanny," he notes. Instead, they were mostly focused on engineering, regulation, financing, and systems modeling. Ialenti emerged from his fieldwork motivated to develop "long-termist intellectual calisthenics" for a scientifically informed public.

Posiva's "safety case" experts—the individuals charged with assessing the project's risks—often worked by analogy, he observes, analyzing the remains of a 2100-year-old human cadaver in China, for example, or ancient Roman nails found in Scotland, or a bronze ship's cannon embedded in sea ice in order to predict the fate of radioactive waste canisters. A comparable activity in which we might engage would be to read books such as Alan Weisman's *The World Without Us* and then to go into outdoor places and imagine their distant pasts and futures, suggests Ialenti. Inspired by scientists' models of future biospheres, he introduces readers to "deep time heuristics," which might include doing research on climatological forecasts and comparing them to predictable patterns in everyday routines.

One of Ialenti's informants draws an analogy to Peru's Nazca lines, whose creators could only imagine the impressive aerial view they were crafting. Ialenti wants us to become similarly adept at zooming in and out of different scales of time. "Multitimescale awareness training" might be offered at fossil fuel corporations, he suggests, challenging employees to reflect on the long-term consequences of extraction. He also envisions the establishment of a global nonprofit think tank to work toward scientifically informed portrayals of future worlds.

Those familiar with the politics of the nuclear industry may find it difficult to accept nuclear experts as exemplars of future thinking. Many have argued that the safety case studies exist to justify Finland's continued public investment in nuclear energy, and at least one of the scientists Ialenti interviewed confided that the archeological analog studies they cited were cherry-picked to suit the nuclear industry's predetermined conclusions. There are limits, after all, to what a bronze cannon submerged at sea for centuries can tell us about copper nuclear waste canisters buried in granite for millennia. Ialenti argues, however, that his book should not be read as an endorsement of the nuclear experts' conclusions, even if it does celebrate their process.

He worries about what he calls the "deflation of expertise," as manifested in the rising antiscience attitudes that undermine the application of expertise to major policy issues. In its place, he wants us to be guardedly enthusiastic about scientific work, which is superior in value to "freewheeling podcast rants, talking head television pundits, and impulsive Twitter posts." Fair enough.

While he does not idealize Finland, Ialenti seems to envy the degree of public trust in technocratic expertise there. We need tools for restoring such trust, he contends, because our survival in the Anthropocene depends on it. ■

10.1126/science.abf2843

The reviewer is at the School of History, Philosophy, and Religion, Oregon State University, Corvallis, OR 97331, USA. Email: jacob.hamblin@oregonstate.edu

GENE EDITING

Purse strings, patients, and personal glory

An anthropologist investigates the factors shaping modern genetic medicine

By **Stephan Güttinger**

When the first gene-edited babies, Lulu and Nana, were born in China in October 2018, He Jiankui was at the hospital. The scientist behind the secretive trial had flown in from Shenzhen with some of his team members as soon as he had heard that the babies' mother was in labor. Apart from wanting to collect samples for genetic testing, he was also hoping to be able to hold the twins for a promotional photo opportunity. The right to do so—and the right to use the photos on billboards, cars, and calendars—was written into the consent form that the parents were required to sign before they could join the trial. He and his team were able to collect most of the samples they were after, but he was not allowed to hold the babies. Contrary to He's later claim that the twins were born "as healthy as any other babies," they were in neonatal intensive care after their premature birth.

The Mutant Project
Eben Kirksey
St. Martin's Press, 2020. 304 pp.

This scene captures but a few of the many intimate details about the first clinical use of heritable human genome editing (HHGE) that are described in *The Mutant Project*, the third monograph by cultural anthropologist Eben Kirksey. Apart from offering new insights into the events surrounding the birth of the twins Lulu and Nana, Kirksey also provides firsthand accounts from parents who took part in He's trial; reproduces conversations between members of He's laboratory as they tried to navigate the trial and their boss's ambitions; and details how He traveled across China and beyond to meet investors, academic advisers, and Chinese government officials to drum up support for his trial and other entrepreneurial endeavors connected to it.

The Mutant Project provides readers with an intriguing picture of the events, ambitions, and deceptions that led up to the twins' birth, but these insights are only part of what makes it such a fascinating read. At its core, the book is a complex analysis of the global culture in which the project of HHGE is now developing.

Kirksey begins his story in November 2018 at the Second International Summit on Human Genome Editing in Hong Kong, where news of the twins' birth first broke. Readers follow the author, who was scheduled to speak at the conference on a panel on research ethics, as he first hears the news and tries to gather more information. From here, the story expands across both time and space.

Kirksey recounts details from the First International Summit on Human Gene Editing in 2015, where Western leaders of gene-editing technology mixed and mingled with government officials and corporate lobbyists. He travels to Philadelphia and the San Francisco Bay Area, where he meets patient-activists who took part in early gene-editing trials and fight for equal access to new medical treatments. He travels to Indonesia, where he introduces readers to a transgender artist who explores possible genetic futures and interrogates questions of race and beauty. And he delves into the realm of do-it-yourself (DIY) biologists, who try to harness the power of new gene-editing technologies to modify themselves and others.

Out of this rich tapestry of stories, some central themes emerge, including the question of how profit-driven medicine and dreams of personal and national glory shape the complex landscape of modern genetic medicine. Kirksey shows how these are global phenomena that unfold their power in the West and the East, in mainstream academia, and at the fringes of DIY biology. He is also concerned about how cultural ideas about race, beauty, and health become intertwined with new genetic technologies and with the global financial network. These issues raise important questions about how we deal with stigma and inequality and about who benefits from new genetic treatments and at what cost.

Kirksey does not attempt to answer all of these questions, so those looking for a neat set of solutions might be disappointed. But they will also be missing a key point of this book. The author describes *The Mutant Project* as a "mosaic portrait," but it is more than this. It also traces the ever-shifting nature of the moral and ethical lines surrounding new genetic possibilities. Kirksey's analysis reminds us that the lines we draw are always in flux. In doing so, it forces us to slow down and think deeply about what questions we ask—and how we answer them—when seeking to determine what is acceptable when it comes to genome editing. ■

10.1126/science.abf2842

The reviewer is at the London School of Economics, London WC2A 2AE, UK. Email: s.m.guettinger@lse.ac.uk

A return to natural forest fire cycles could help prevent massive wildfires.

PHOTO: STANTON SHARPE/SOPA IMAGES/SIPA USA/AP IMAGES

PODCAST

What Would Nature Do? A Guide for Our Uncertain Times
Ruth DeFries
Columbia University Press, 2020. 264 pp.

The challenges facing humanity are increasingly complex, making it difficult to know where to begin when one seeks to chart an effective path forward. This week on the *Science* podcast, Ruth DeFries recommends looking to nature for fundamental and time-tested strategies that have long helped life thrive in uncertain times.
https://scim.ag/3k3V3U4

10.1126/science.abf4324

Global agreements aim to make the development of space safe and sustainable.

PHOTO: NASA

Edited by **Jennifer Sills**

U.S. space policy: Multilateral mining

In their Policy Forum "U.S. policy puts the safe development of space at risk" (9 October, p. 174), A. Boley and M. Byers claim that U.S. plans to purchase a small amount of commercial Moon resources put the safe development of space at risk. To the contrary, U.S. actions are poised to help jumpstart a new age of space innovation that has the potential to bring scientific and economic benefits to the whole world. There is no global consensus on international regulation of space mining. The Moon Agreement and other related recent attempts have been widely rejected. The United States has led negotiations on the Artemis Accords on a bilateral basis, but it will be a multilateral agreement, with Japan, Canada, the United Kingdom, Italy, Australia, Luxembourg, and the United Arab Emirates joining in October (*1*). The Accords can form a foundation for a future global framework, which ensures all nations can access the resources of outer space.

Environmental management and standards are essential to space mining but are not guaranteed by international regulatory models. International space law already requires, or encourages, national regulation with regard to planetary protection, avoiding harmful contamination, and discouraging harmful interference (*2*). U.S. leadership in the space environment can help ensure that sustainability is engrained in the space mining industry from the start. Early national regulation can facilitate sustainable space mining innovation; it does not inhibit it.

Boley and Byers point to the management of the deep seabed by the United Nations Convention on the Law of the Sea as an alternative global model for sustainable resource use. However, decades later, that model has still not been successful. No commercial deep-sea mining has occurred in the High Seas, the benefit sharing mechanism is not established, and the system has struggled to define environmental management standards. A global approach may eventually emerge, but it is better to begin by stimulating the market and implementing imperfect regulation than to wait.

Alexander Gilbert*, Morgan Bazilian, George Sowers
Colorado School of Mines, Golden, CO 80401, USA.
*Corresponding author. Email: gilbeaq@gmail.com

REFERENCES AND NOTES

1. National Aeronautics and Space Administration (NASA), "NASA, international partners advanced cooperation with first signings of the Artemis Accords" (2020).
2. S. W. Anderson, K. Christensen, J. LaManna, *J. Energ. Nat. Resourc. Law* **37**, 227 (2019).

10.1126/science.abf2456

U.S. space policy: An international model

In their Policy Forum "U.S. policy puts the safe development of space at risk" (9 October, p. 174), A. Boley and M. Byers paint an inaccurate picture of the National Aeronautics and Space Administration's (NASA's) Artemis program and the agreements (the so-called Artemis Accords) that NASA is using to partner with other countries in this historic mission. In contrast to the authors' claims, the Artemis Accords are not intended to supplant the need for an international agreement on the issue of space resource activity.

The United States has raised no objections to the scheduled "general exchange of views" regarding space resource activity in the United Nations' Committee on the Peaceful Uses of Outer Space (*1*). Moreover, the U.S. government actively encouraged the work of the Hague International Space Resources Governance Working Group, which recently completed their work creating "building blocks" for a future international framework to regulate this new activity (*2*). In fact, many of the concepts in the Artemis Accords have their roots in the discussions of the Hague Working Group. This allegation of unilateralism is particularly unfair given that a primary purpose of the Artemis Accords is to ensure that this historic mission is carried out by a coalition of international partners who explicitly commit to comply with the existing principles of international law.

Boley and Byers imply that NASA intends to unilaterally create "safety zones" around their areas of activity that will effectively "exclude other actors." The safety zones proposed by NASA do no such thing. The Outer Space Treaty clearly protects the right of all space actors to enjoy free access to all parts of the Moon. Creating keep-out zones (areas to which other countries are prohibited entry) would be a patent violation of

the treaty, and this is not the intent or the effect of the proposed safety zones (nor will NASA be the "gatekeeper" of the Moon, as the authors suggest). The safety zones are purely informational in order to permit other actors on the surface of the Moon to properly discharge their duty to act with "due regard to the corresponding interests" (*3*) of other actors under the Outer Space Treaty and avoid harmful interference with existing activities. NASA has no right to create keep-out zones and has no intention of doing so.

NASA is not trying to "redirect international space cooperation in favor of short-term U.S. commercial interests." Rather, NASA is working to ensure that the Artemis Program becomes a model of international cooperation that is carried out with strict adherence to international law. NASA is fully supportive of UN initiatives to explore the possibility of new law governing lunar activity, but the UN process will likely take a decade or more to produce an agreement of any significance. In the meantime, NASA is moving forward by building a team of international partners who, through the Artemis Accords, promise to observe existing space law as a condition of joining the venture. This is how responsible space actors behave.

Mark J. Sundahl
Cleveland-Marshall College of Law and Global Space Law Center, Cleveland State University, Cleveland, OH 44115, USA. Email: m.sundahl@csuohio.edu

REFERENCES AND NOTES

1. Committee on the Peaceful Uses of Outer Space, Legal Subcommittee, Fifty-Ninth Session, Annotated Provisional Agenda, A/AC.105/C.2/L.312/Rev.1 (2020), p. 3.
2. O. de O. Bittencourt Neto *et al*., Eds., *Building Blocks for the Development of an International Framework for the Governance of Space Resource Activities* (2020), p. 111; https://boeken.rechtsgebieden.boomportaal.nl/publicaties/9789462361218#152.
3. "United Nations Treaties and Principles on Outer Space" (United Nations, New York, 2002), p. 6.

COMPETING INTERESTS

M.J.S. is a member of the NASA Regulatory and Policy Advisory Committee.

10.1126/science.abf2324

Response

We agree with Gilbert *et al*. that "environmental management and standards are essential to space mining." However, we do not agree that national legislation is sufficient for regulating space mining, nor do we inherently expect national laws to help "facilitate sustainable space mining innovation."

In the case of terrestrial mining, differing national laws and regulations, along with weak enforcement, have led to human rights abuses, environmental damage, and adverse health impacts, particularly in the Global South. National regulation of space activity with little to no international involvement has already led to a low Earth orbit debris crisis (*1*, *2*), which will likely be exacerbated by the construction of mega-constellations of communications satellites, again under national regulation (*3*). Gilbert *et al*. argue that the Artemis Accords set the stage for global cooperation and access to space resources for all countries, but these bilateral agreements with just seven close U.S. allies have so far left out the vast majority of nations.

The four treaties that make up the core of international space law, to which Gilbert *et al*. positively refer, were negotiated through a multilateral process. The U.S. approach leaves out Russia, China, and India—all powerful space actors. As for deep seabed mining, the greatest challenge has not been multilateralism, but prohibitively high costs. So far, the International Seabed Authority created under the UN Convention on the Law of the Sea has issued exploratory contracts for over 1.3 million square kilometers of the high seas (*4*).

Sundahl focuses exclusively on the Artemis Accords without acknowledging the multiple components we discussed that now make up U.S. Space Policy, including the September 2020 announcement that NASA would purchase lunar regolith from a private company. Such a purchase would set a precedent for the U.S. view that commercial space mining is legal—a view not supported by most other nations. Our Policy Forum examines the concerted U.S. effort to secure widespread acceptance of the U.S. preferred interpretation of space law as the international interpretation of space law.

We are pleased that NASA partner states ultimately chose not to make the Artemis Accords legally binding. As explicitly non-binding political documents, their effect on existing international law will be limited. At the same time, this means that—legally speaking—the United States might continue to act unilaterally in space, for instance, by pursuing the planned purchase of regolith. Although the Accords make multiple references to an eventual multilateral process, this is not guaranteed and would take place against the backdrop of any U.S. unilateralism. As for the safety zones foreseen in the Artemis Accords, we appreciate that the zones are only meant to be advisory; however, notifying the United Nations of any such zone will have no legitimizing effect in the absence of a multilateral regime for space mining. It also raises the question of how China and other spacefaring nations might regard any safety zones that they choose to establish, now that the United States has opened this door.

Aaron Boley[1]* and Michael Byers[2]
[1]Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada. [2]Department of Political Science, University of British Columbia, Vancouver, BC, Canada. *Corresponding author. Email: acboley@phas.ubc.ca

REFERENCES AND NOTES

1. NASA, Orbital Debris Program Office (https://orbitaldebris.jsc.nasa.gov/reference-documents/).
2. J.-C. Liou, N. L. Johnson, *Adv. Space Res.* **41**, 1046 (2008).
3. J. O'Callaghan, "SpaceX's Starlink could cause cascades of space junk," *Scientific American* (2019).
4. M. Lodge, "The International Seabed Authority and deep seabed mining," *UN Chronicle* (www.un.org/en/chronicle/article/international-seabed-authority-and-deep-seabed-mining).

10.1126/science.abf4916

TECHNICAL COMMENT ABSTRACTS

Comment on "Forest microclimate dynamics drive plant responses to warming"

Romain Bertrand, Fabien Aubret, Gaël Grenouillet, Alexandre Ribéron, Simon Blanchet

Zellweger *et al*. (Reports, 15 May 2020, p. 772) claimed that forest plant communities' response to global warming is primarily controlled by microclimate dynamics. We show that community thermophilization is poorly explained by the underlying components of microclimate and that global warming primarily controls the climatic lag of plant communities. Deconstructing the underlying components of microclimate provides insights for managers.

Full text: dx.doi.org/10.1126/science.abd3850

Response to Comment on "Forest microclimate dynamics drive plant responses to warming"

Florian Zellweger, Pieter De Frenne, Jonathan Lenoir, Pieter Vangansbeke, Kris Verheyen, Markus Bernhardt-Römermann, Lander Baeten, Radim Hédl, Imre Berki, Jörg Brunet, Hans Van Calster, Markéta Chudomelová, Guillaume Decocq, Thomas Dirnböck, Tomasz Durak, Thilo Heinken, Bogdan Jaroszewicz, Martin Kopecký, František Máliš, Martin Macek, Marek Malicki, Tobias Naaf, Thomas A. Nagel, Adrienne Ortmann-Ajkai, Petr Petřík, Remigiusz Pielech, Kamila Reczyńska, Wolfgang Schmidt, Tibor Standovár, Krzysztof Świerkosz, Balázs Teleki, Ondřej Vild, Monika Wulf, David Coomes

Bertrand *et al*. question our interpretation about warming effects on the thermophilization in forest plant communities and propose an alternative way to analyze climatic debt. We show that microclimate warming is a better predictor than macroclimate warming for studying forest plant community responses to warming. Their additional analyses do not affect or change our interpretations and conclusions.

Full text: dx.doi.org/10.1126/science.abd6193

Cite as: R. Bertrand *et al.*, *Science* 10.1126/science.abd3850 (2020).

Comment on "Forest microclimate dynamics drive plant responses to warming"

Romain Bertrand[1]*, Fabien Aubret[2], Gaël Grenouillet[1,3], Alexandre Ribéron[1], Simon Blanchet[1,2]

[1]Laboratoire Évolution et Diversité Biologique, UMR5174, Université de Toulouse III Paul Sabatier, CNRS, IRD, Toulouse, France. [2]Station d'Écologie Théorique et Expérimentale, UMR5321, CNRS, Université de Toulouse III Paul Sabatier, Moulis, France. [3]Institut Universitaire de France, Paris, France.

***Corresponding author. Email: romain.bertrand2@univ-tlse3.fr**

Zellweger *et al*. (Reports, 15 May 2020, p. 772) claimed that forest plant communities' response to global warming is primarily controlled by microclimate dynamics. We show that community thermophilization is poorly explained by the underlying components of microclimate, and that global warming primarily controls the climatic lag of plant communities. Deconstructing the underlying components of microclimate provides insights for managers.

The debate about the efficiency of canopy cover in mitigating or amplifying the effect of global warming on forest plant communities has important implications for forest and environmental management (*1*–*3*). Zellweger *et al.* (*4*) tested the relationship between reshuffling in plant communities and changes in subcanopy microclimate conditions statistically inferred from canopy cover trends over the past few decades. They concluded that microclimate dynamics are pertinent in understanding the responses of forest plant communities to global warming. Although we agree that microclimate is an important driver, we stress that the claim that "thermophilization and the climatic lag in forest plant communities are primarily controlled by microclimate" simplifies a more complex regulation of plant communities by global warming. Microclimate stems from global climatic trends [macroclimate in (*4*)] and local habitat characteristics [canopy cover in (*4*)] that buffer or amplify the effect of macroclimate. Because these two components act simultaneously, their relative influence on biodiversity needs to be assessed explicitly (*2*). By not doing so, Zellweger *et al.* underestimated the effect of macroclimate on biodiversity. The explicit consideration of these two components provides crucial information to managers and policymakers on which component (local and/or global) to focus management efforts for effective mitigation of global change impacts on biodiversity. Here, we provide an alternative statistical approach that explicitly tests for the relative effects of each component of microclimate on forest plant communities. Beyond an alternative interpretation of Zellweger *et al.*'s dataset, we hope to contribute to a consensus on how to assess the effects of the underlying components of microclimate on biodiversity.

Our alternative analytical approach aims at comparing and assessing the relative effects of local and global drivers of microclimate explicitly (*2*). This comparison is necessary to interpret the effect of microclimate change on plant communities, as the latter can be largely driven by global warming. On the basis of the same linear mixed-effect models framework as Zellweger *et al.*, we propose the following model: thermophilization or microdebt = macroTC + ΔTbuff + macroTC*ΔTbuff + (1|region), where macroTC and ΔTbuff are the global and local components of microclimate, respectively, and macroTC*ΔTbuff is their interaction (*5*). This model assumes that change in macroclimate temperature (macroTC), change in temperature buffering (ΔTbuff), and the two-term interaction likely explain the microclimatic debt (microdebt) or the community thermophilization.

First, Zellweger *et al.* showed that thermophilization of plant communities (i.e., changes in communities toward more warm-affinity species assemblages) increased with microclimate warming (*4*). This is correct, but we wish to point out that microclimate explained only 1.1% of thermophilization. Our analysis further showed that canopy closure had cooled local climate conditions within forest stands and mitigated the thermophilization of plant communities as macroclimate warmed (Table 1). Although significant, change in temperature buffering due to canopy openness and closure and its interplay with global warming explained only 0.5% and 0.2% of the community thermophilization, respectively. Therefore, we cautiously conclude that community thermophilization is not primarily explained by microclimate change and these underlying local or global drivers. Instead, the reshuffling of plant communities might be the result of species persistence (through

adaptive and nonadaptive evolutionary pathways), species dispersal, species interaction, and other components of global change (*2*, *6*–*9*), likely explaining why community thermophilization does not strictly follow microclimate change.

Second, we showed that underlying local and global drivers of microclimate explained 77.1% of the microclimatic debt (i.e., the difference between the thermophilization rate of plant communities and actual changes in microclimate conditions, which quantifies the lag in plant community responses to warming; Table 1). Change in macroclimate explained a much greater portion of the variance in microclimatic debt than did change in temperature buffering (44.6% versus 31.1%; Fig. 1A). Global warming also amplified the microclimatic debt to a higher degree than change in temperature buffering (mean effect of macroTC = 0.252 versus mean effect of ΔTbuff = −0.210; Table 1). Moreover, the effect of ΔTbuff on microclimatic debt was modulated by the magnitude of global warming (mean effect of the interaction term = −0.05; Table 1). This showed that the amplifying effect of global warming on microclimatic debt increased as the temperature buffering decreased as a result of canopy openness, with an effect of global warming that was greater by a factor of ~6.5 when canopy opened rather than closed over time (Fig. 1B). The amplifying effect of global warming on microclimatic debt was significant even when the temperature buffering effect was maximal (mean effect of macroTC = 0.073 when ΔTbuff = 2.8°C; $P < 0.001$), which suggests a consistent influence of global warming on plant communities, regardless of the change in canopy cover.

Our reanalysis provides a more precise understanding of the role of microclimate dynamics in plant biodiversity response to climate change by highlighting the interplay between the local and global components of microclimate. Deconstructing the underlying components of microclimate change is required to formulate appropriate management avenues and to mitigate the effect of climate change on forest biodiversity. The reshuffling of understory plant communities toward warm-affinity assemblages is therefore not primarily controlled by microclimate change per se. The climatic lag observed in forest plant communities is mostly driven by a synergetic effect between global warming and changes in temperature buffering due to local habitat change (*3*), with a global warming effect persisting irrespective of the magnitude of canopy closure. If our results confirm that forest management should limit canopy openness (*10*), we stress that relying solely on canopy cover management to mitigate the effect of global warming on plants, and more broadly on forest ecosystems biodiversity, is not a long-term solution. Global warming is expected to increase tree mortality in the foreseeable future (*11*, *12*), with the effect of opening up forest canopies over vast parts of the planet (*13*). To preserve forest biodiversity, one must first control global warming.

REFERENCES AND NOTES

1. P. De Frenne, F. Rodríguez-Sánchez, D. A. Coomes, L. Baeten, G. Verstraeten, M. Vellend, M. Bernhardt-Römermann, C. D. Brown, J. Brunet, J. Cornelis, G. M. Decocq, H. Dierschke, O. Eriksson, F. S. Gilliam, R. Hédl, T. Heinken, M. Hermy, P. Hommel, M. A. Jenkins, D. L. Kelly, K. J. Kirby, F. J. G. Mitchell, T. Naaf, M. Newman, G. Peterken, P. Petrík, J. Schultz, G. Sonnier, H. Van Calster, D. M. Waller, G.-R. Walther, P. S. White, K. D. Woods, M. Wulf, B. J. Graae, K. Verheyen, Microclimate moderates plant responses to macroclimate warming. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18561–18565 (2013). doi:10.1073/pnas.1311190110 Medline
2. R. Bertrand, G. Riofrío-Dillon, J. Lenoir, J. Drapier, P. de Ruffray, J.-C. Gégout, M. Loreau, Ecological constraints increase the climatic debt in forests. *Nat. Commun.* **7**, 12643 (2016). doi:10.1038/ncomms12643 Medline
3. J. J. Lembrechts, I. Nijs, Microclimate shifts in a dynamic world. *Science* **368**, 711–712 (2020). doi:10.1126/science.abc1245 Medline
4. F. Zellweger, P. De Frenne, J. Lenoir, P. Vangansbeke, K. Verheyen, M. Bernhardt-Römermann, L. Baeten, R. Hédl, I. Berki, J. Brunet, H. Van Calster, M. Chudomelová, G. Decocq, T. Dirnböck, T. Durak, T. Heinken, B. Jaroszewicz, M. Kopecký, F. Máliš, M. Macek, M. Malicki, T. Naaf, T. A. Nagel, A. Ortmann-Ajkai, P. Petřík, R. Pielech, K. Reczyńska, W. Schmidt, T. Standovár, K. Świerkosz, B. Teleki, O. Vild, M. Wulf, D. Coomes, Forest microclimate dynamics drive plant responses to warming. *Science* **368**, 772–775 (2020). doi:10.1126/science.aba6880 Medline
5. Materials and methods, data, and R scripts of the analysis conducted here are available at https://figshare.com/s/07a39af48116a3126689.
6. F. Essl, S. Dullinger, W. Rabitsch, P. E. Hulme, P. Pyšek, J. R. U. Wilson, D. M. Richardson, Delayed biodiversity change: No time to waste. *Trends Ecol. Evol.* **30**, 375–378 (2015). doi:10.1016/j.tree.2015.05.002 Medline
7. R. Bertrand, Unequal contributions of species' persistence and migration on plant communities' response to climate warming throughout forests. *Ecography* **42**, 211–212 (2019). doi:10.1111/ecog.03591
8. I. R. Staude, D. M. Waller, M. Bernhardt-Römermann, A. D. Bjorkman, J. Brunet, P. De Frenne, R. Hédl, U. Jandt, J. Lenoir, F. Máliš, K. Verheyen, M. Wulf, H. M. Pereira, P. Vangansbeke, A. Ortmann-Ajkai, R. Pielech, I. Berki, M. Chudomelová, G. Decocq, T. Dirnböck, T. Durak, T. Heinken, B. Jaroszewicz, M. Kopecký, M. Macek, M. Malicki, T. Naaf, T. A. Nagel, P. Petřík, K. Reczyńska, F. H. Schei, W. Schmidt, T. Standovár, K. Świerkosz, B. Teleki, H. Van Calster, O. Vild, L. Baeten, Replacements of small- by large-ranged species scale up to diversity loss in Europe's temperate forest biome. *Nat. Ecol. Evol.* **4**, 802–808 (2020). doi:10.1038/s41559-020-1176-8 Medline
9. J. Lenoir, R. Bertrand, L. Comte, L. Bourgeaud, T. Hattab, J. Murienne, G. Grenouillet, Species better track climate warming in the oceans than on land. *Nat. Ecol. Evol.* **4**, 1044–1059 (2020). doi:10.1038/s41559-020-1198-2 Medline
10. D. Thom, A. Sommerfeld, J. Sebald, J. Hagge, J. Müller, R. Seidl, Effects of disturbance patterns and deadwood on the microclimate in European beech forests. *Agric. For. Meteorol.* **291**, 108066 (2020). doi:10.1016/j.agrformet.2020.108066
11. N. G. McDowell, C. D. Allen, Darcy's law predicts widespread forest mortality under climate warming. *Nat. Clim. Change* **5**, 669–672 (2015). doi:10.1038/nclimate2641
12. M. J. P. Sullivan, S. L. Lewis, K. Affum-Baffoe, C. Castilho, F. Costa, A. C. Sanchez, C. E. N. Ewango, W. Hubau, B. Marimon, A. Monteagudo-Mendoza, L. Qie, B. Sonké, R. V. Martinez, T. R. Baker, R. J. W. Brienen, T. R. Feldpausch, D. Galbraith, M. Gloor, Y. Malhi, S.-I. Aiba, M. N. Alexiades, E. C. Almeida, E. A. de Oliveira, E. Á. Dávila, P. A. Loayza, A. Andrade, S. A. Vieira, L. E. O. C. Aragão, A. Araujo-Murakami, E. J. M. M. Arets, L. Arroyo, P. Ashton, G. Aymard C, F. B.

Baccaro, L. F. Banin, C. Baraloto, P. B. Camargo, J. Barlow, J. Barroso, J.-F. Bastin, S. A. Batterman, H. Beeckman, S. K. Begne, A. C. Bennett, E. Berenguer, N. Berry, L. Blanc, P. Boeckx, J. Bogaert, D. Bonal, F. Bongers, M. Bradford, F. Q. Brearley, T. Brncic, F. Brown, B. Burban, J. L. Camargo, W. Castro, C. Céron, S. C. Ribeiro, V. C. Moscoso, J. Chave, E. Chezeaux, C. J. Clark, F. C. de Souza, M. Collins, J. A. Comiskey, F. C. Valverde, M. C. Medina, L. da Costa, M. Dančák, G. C. Dargie, S. Davies, N. D. Cardozo, T. de Haulleville, M. B. de Medeiros, J. Del Aguila Pasquel, G. Derroire, A. Di Fiore, J.-L. Doucet, A. Dourdain, V. Droissart, L. F. Duque, R. Ekoungoulou, F. Elias, T. Erwin, A. Esquivel-Muelbert, S. Fauset, J. Ferreira, G. F. Llampazo, E. Foli, A. Ford, M. Gilpin, J. S. Hall, K. C. Hamer, A. C. Hamilton, D. J. Harris, T. B. Hart, R. Hédl, B. Herault, R. Herrera, N. Higuchi, A. Hladik, E. H. Coronado, I. Huamantupa-Chuquimaco, W. H. Huasco, K. J. Jeffery, E. Jimenez-Rojas, M. Kalamandeen, M. N. K. Djuikouo, E. Kearsley, R. K. Umetsu, L. K. Kho, T. Killeen, K. Kitayama, B. Klitgaard, A. Koch, N. Labrière, W. Laurance, S. Laurance, M. E. Leal, A. Levesley, A. J. N. Lima, J. Lisingo, A. P. Lopes, G. Lopez-Gonzalez, T. Lovejoy, J. C. Lovett, R. Lowe, W. E. Magnusson, J. Malumbres-Olarte, Â. G. Manzatto, B. H. Marimon Jr., A. R. Marshall, T. Marthews, S. M. de Almeida Reis, C. Maycock, K. Melgaço, C. Mendoza, F. Metali, V. Mihindou, W. Milliken, E. T. A. Mitchard, P. S. Morandi, H. L. Mossman, L. Nagy, H. Nascimento, D. Neill, R. Nilus, P. N. Vargas, W. Palacios, N. P. Camacho, J. Peacock, C. Pendry, M. C. Peñuela Mora, G. C. Pickavance, J. Pipoly, N. Pitman, M. Playfair, L. Poorter, J. R. Poulsen, A. D. Poulsen, R. Preziosi, A. Prieto, R. B. Primack, H. Ramírez-Angulo, J. Reitsma, M. Réjou-Méchain, Z. R. Correa, T. R. de Sousa, L. R. Bayona, A. Roopsind, A. Rudas, E. Rutishauser, K. Abu Salim, R. P. Salomão, J. Schietti, D. Sheil, R. C. Silva, J. S. Espejo, C. S. Valeria, M. Silveira, M. Simo-Droissart, M. F. Simon, J. Singh, Y. C. Soto Shareva, C. Stahl, J. Stropp, R. Sukri, T. Sunderland, M. Svátek, M. D. Swaine, V. Swamy, H. Taedoumg, J. Talbot, J. Taplin, D. Taylor, H. Ter Steege, J. Terborgh, R. Thomas, S. C. Thomas, A. Torres-Lezama, P. Umunay, L. V. Gamarra, G. van der Heijden, P. van der Hout, P. van der Meer, M. van Nieuwstadt, H. Verbeeck, R. Vernimmen, A. Vicentini, I. C. G. Vieira, E. V. Torre, J. Vleminckx, V. Vos, O. Wang, L. J. T. White, S. Willcock, J. T. Woods, V. Wortel, K. Young, R. Zagt, L. Zemagho, P. A. Zuidema, J. A. Zwerts, O. L. Phillips, Long-term thermal sensitivity of Earth's tropical forests. *Science* **368**, 869–874 (2020). doi:10.1126/science.aaw7578 Medline

13. N. G. McDowell, C. D. Allen, K. Anderson-Teixeira, B. H. Aukema, B. Bond-Lamberty, L. Chini, J. S. Clark, M. Dietze, C. Grossiord, A. Hanbury-Brown, G. C. Hurtt, R. B. Jackson, D. J. Johnson, L. Kueppers, J. W. Lichstein, K. Ogle, B. Poulter, T. A. M. Pugh, R. Seidl, M. G. Turner, M. Uriarte, A. P. Walker, C. Xu, Pervasive shifts in forest dynamics in a changing world. *Science* **368**, eaaz9463 (2020). doi:10.1126/science.aaz9463 Medline

ACKNOWLEDGMENTS

We thank Zellweger and colleagues for making their analyses and data openly available, and for discussions. We thank R. Michniewicz for assistance with English. **Funding:** Supported by the Agence Nationale de la Recherche (ISOLAPOP ANR-18-CE03-0011) (R.B., S.B., and A.R.); European INTERREG POCTEFA projects (ECTOPYR no. EFA031/15 and OPCC ADAPYR no. EFA346/19) (R.B. and F.A.); and the TULIP Laboratory of Excellence (ANR-10-LABX-41). **Author contributions:** R.B. conducted the analysis and wrote the first draft with contributions from S.B.; all co-authors contributed substantially to improve it. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data used in the study are original data of Zellweger *et al.* (2020) available at https://doi.org/10.5061/dryad.r7sqv9s83. R scripts of the analysis conducted in the present Comment are available at https://figshare.com/s/07a39af48116a3126689.

25 June 2020; accepted 6 November 2020
Published online 27 November 2020
10.1126/science.abd3850

Table 1. Summary statistics for the models explaining variation of thermophilization and microclimatic debt observed in forest plant communities (n = 2943 observations). Means and 95% confidence intervals of estimates, as well as the significance term for each fixed effect, were computed from 10,000 bootstrapped models. Boldface P values highlight significant effect (considering a threshold $\alpha < 0.05$).

		Thermophilization			Microclimatic debt		
	Variable	**Mean**	**95% CI**	***P***	**Mean**	**95% CI**	***P***
Fixed effect	Intercept	0.012	[–0.004; 0.027]	0.066	0.43	[0.409; 0.451]	**<0.001**
	MacroTC	–0.002	[–0.015; 0.011]	0.372	0.252	[0.237; 0.267]	**<0.001**
	ΔTbuff	–0.012	[–0.018; –0.005]	**<0.001**	–0.21	[–0.217; –0.203]	**<0.001**
	MacroTC × ΔTbuff	–0.007	[–0.014; 0]	**0.025**	–0.05	[–0.057; –0.043]	**<0.001**
Variation explained	Marginal R^2	0.008	[0.002; 0.018]		0.771	[0.749; 0.79]	
	Conditional R^2	0.105	[0.066; 0.149]		0.812	[0.797; 0.825]	

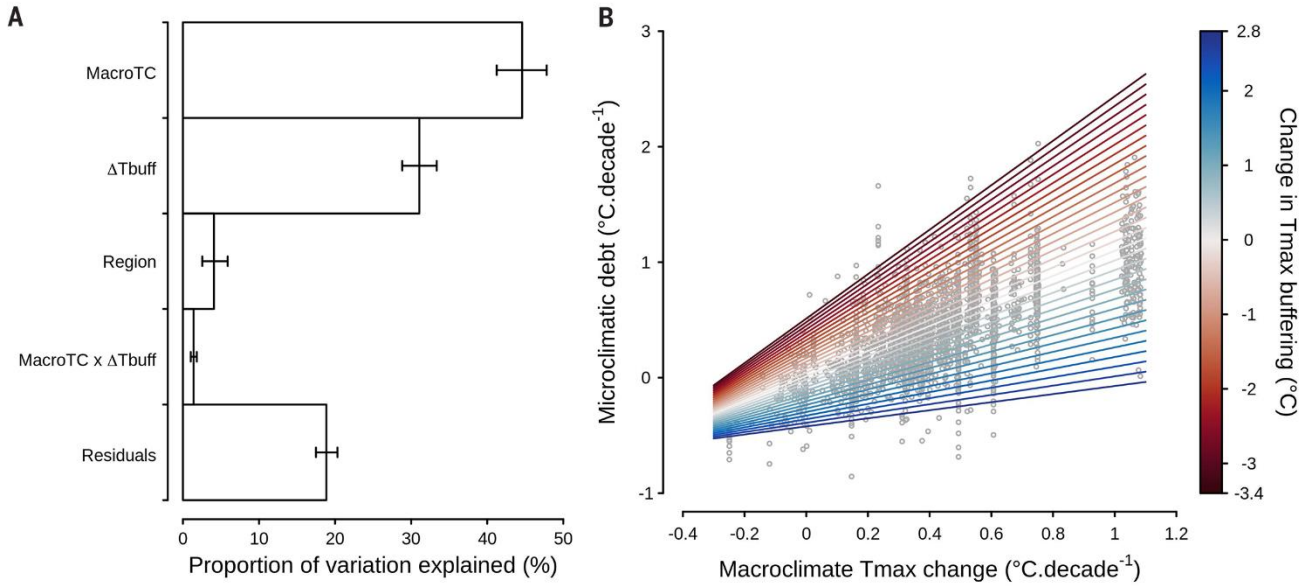


Fig. 1. Changes in macroclimate temperature (MacroTC) and in temperature buffering by canopy closure and openness (ΔTbuff) both explain the microclimatic debt in forest plant communities. (A) Proportion of the microclimatic debt variation explained by fixed effects (MacroTC and ΔTbuff), by random intercept (Region), and unexplained (Residuals). Error bars represent 95% confidence intervals computed from 10,000 bootstrap iterations. (B) The microclimatic debt rises as global warming increases and temperature buffering decreases as a result of canopy opening. Negative values on the x axis represent macroclimate cooling; positive values represent macroclimate warming. Dots are observations (n = 2943).

Cite as: F. Zellweger *et al.*, *Science* 10.1126/science.abd6193 (2020).

Response to Comment on "Forest microclimate dynamics drive plant responses to warming"

Florian Zellweger[1,2]*, Pieter De Frenne[3], Jonathan Lenoir[4], Pieter Vangansbeke[3], Kris Verheyen[3], Markus Bernhardt-Römermann[5], Lander Baeten[3], Radim Hédl[6,7], Imre Berki[8], Jörg Brunet[9], Hans Van Calster[10], Markéta Chudomelová[11], Guillaume Decocq[4], Thomas Dirnböck[12], Tomasz Durak[13], Thilo Heinken[14], Bogdan Jaroszewicz[15], Martin Kopecký[16,17], František Máliš[18,19], Martin Macek[16], Marek Malicki[20], Tobias Naaf[21], Thomas A. Nagel[22], Adrienne Ortmann-Ajkai[23], Petr Petřík[16], Remigiusz Pielech[24], Kamila Reczyńska[25], Wolfgang Schmidt[26], Tibor Standovár[27], Krzysztof Świerkosz[28], Balázs Teleki[29,30], Ondřej Vild[11], Monika Wulf[21], David Coomes[1]

[1]Forest Ecology and Conservation Group, Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. [2]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, 8903 Birmensdorf, Switzerland. [3]Forest and Nature Lab, Department of Environment, Faculty of Bioscience Engineering, Ghent University, Melle-Gontrode, Belgium. [4]UR "Ecologie et Dynamique des Systèmes Anthropisés" (EDYSAN, UMR 7058 CNRS-UPJV), Université de Picardie Jules Verne, 80037 Amiens Cedex 1, France. [5]Institute of Ecology and Evolution, Friedrich Schiller University Jena, D-07743 Jena, Germany. [6]Institute of Botany of the Czech Academy of Sciences, CZ-602 00 Brno, Czech Republic. [7]Department of Botany, Faculty of Science, Palacký University in Olomouc, CZ-78371 Olomouc, Czech Republic. [8]Institute of Environmental and Earth Sciences, University of Sopron, H-9400 Sopron, Hungary. [9]Swedish University of Agricultural Sciences, Southern Swedish Forest Research Centre, Box 49, 230 53 Alnarp, Sweden. [10]Research Institute for Nature and Forest (INBO), B-1000 Brussels, Belgium. [11]Institute of Botany of the Czech Academy of Sciences, CZ-602 00 Brno, Czech Republic. [12]Environment Agency Austria, A-1090 Vienna, Austria. [13]Department of Plant Physiology and Ecology, University of Rzeszów, PL-35-959 Rzeszów, Poland. [14]General Botany, Insitute of Biochemistry and Biology, University of Potsdam, 14469 Potsdam, Germany. [15]Białowieża Geobotanical Station, Faculty of Biology, University of Warsaw, 17-230 Białowieża, Poland. [16]Institute of Botany of the Czech Academy of Sciences, CZ-252 43 Průhonice, Czech Republic. [17]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, CZ-165 21 Prague 6 - Suchdol, Czech Republic. [18]Faculty of Forestry, Technical University in Zvolen, SK-960 01 Zvolen, Slovakia. [19]National Forest Centre, SK-960 01 Zvolen, Slovakia. [20]Department of Botany, Faculty of Biological Sciences, University of Wrocław, Wrocław, Poland. [21]Leibniz Centre for Agricultural Landscape Research (ZALF), D-15374 Muencheberg, Germany. [22]Department of Forestry and Renewable Forest Resources, Biotechnical Faculty, University of Ljubljana, Ljubljana 1000, Slovenia. [23]Department of Hydrobiology, Institute of Biology, University of Pécs, H-7624 Pécs, Hungary. [24]Department of Forest Biodiversity, Faculty of Forestry, University of Agriculture, Kraków, Poland. [25]Department of Botany, Institute of Environmental Biology, University of Wrocław, PL-50-328 Wrocław, Poland. [26]Department of Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, Göttingen, Germany. [27]Department of Plant Systematics, Ecology and Theoretical Biology, Institute of Biology, L. Eötvös University, H-1117 Budapest, Hungary. [28]Museum of Natural History, University of Wrocław, PL-50-335 Wroclaw, Poland. [29]Institute for Regional Development, University of Pécs, H-7100 Szekszárd, Hungary. [30]Department of Ecology, University of Debrecen, H-4032 Debrecen, Hungary.

*Corresponding author. Email: florian.zellweger@wsl.ch

Bertrand *et al.* question our interpretation about warming effects on the thermophilization in forest plant communities and propose an alternative way to analyze climatic debt. We show that microclimate warming is a better predictor than macroclimate warming for studying forest plant community responses to warming. Their additional analyses do not affect or change our interpretations and conclusions.

Canopy structure and composition are important determinants of spatial and temporal variation in forest microclimate (*1*, *2*). Accordingly, we inferred the microclimate from macroclimate data and the modulating effect of the canopy layer [figure 1B in Zellweger *et al.* (*3*)]. Bertrand *et al.* (*4*) used our data to separately analyze the effects of macroclimate warming and temperature buffering on community thermophilization and its associated microclimatic debt (i.e., the difference between the rate of forest microclimate warming and the rate of community thermophilization). We argue that it is ecologically more meaningful to analyze community responses to global warming based on the warming rates that the forest understory organisms actually experience—that is, the microclimate warming rates—applying a "plant's-eye view" approach on biodiversity responses to climate change (*5*, *6*). For that specific reason, in our report (*3*), we directly incorporated the widely acknowledged temperature buffering effect that canopy cover has on macroclimate to assess the rate of microclimate warming, which we subsequently related to community thermophilization. We therefore believe that Bertrand *et al.*'s slightly different model to deconstruct the microclimate warming rate experienced by forest-dwelling organisms into macroclimate warming and temperature buffering hardly provides additional insights into the effects of microclimate dynamics on forest understory plant responses to warming.

Because we did not separate macroclimate warming from temperature buffering, Bertrand *et al.* claim that we "underestimated the effect of macroclimate on biodiversity." This would imply a shortcoming of our paper that is not supported by their additional analysis. In fact, macroclimate warming is an integral part of microclimate warming, and microclimate warming happens in part because of macroclimate warming. The main point drawn from our study is that the rate of forest microclimate warming is simply a better and more natural predictor variable for studying forest understory plant community responses to climate change than the rate of macroclimate warming. Arguably, from a forest floor organism's view, the microclimate is more relevant than the macroclimate. Microclimate change rates should therefore be preferred over macroclimate data in future assessments of biodiversity responses to climate change.

As discussed in our paper, our results indeed have implications for forest management and biodiversity conservation. For example, as shown in figure 3A of (*3*), temporal changes in canopy cover can mitigate or amplify the need for forest plant communities to respond to warming, a finding that is of interest to foresters and policymakers engaged in forest biodiversity conservation. This finding is further supported by figure S5 of (*3*), where we show the relationship between microclimate warming and canopy cover changes over time. We therefore disagree with Bertrand *et al.*'s claim that removing macroclimate change from calculations of microclimate change provides crucial information to managers and policymakers. However, we entirely agree with Bertrand *et al.* that a sole focus on canopy cover management does not represent a sustainable long-term solution to counteract the pressures imposed on forest biodiversity by global warming, and we have never argued otherwise.

With regard to Bertrand *et al.*'s comments on our thermophilization analysis, we stress that we studied directional shifts in species composition as a result of temperature changes by analyzing community thermophilization. This allowed us to focus explicitly on warming effects on plant communities. Our results show that the thermophilization rates in forest plant communities are more related to microclimate warming than to macroclimate warming, because microclimate data integrate the temperature buffering information as well as macroclimate data. Bertrand *et al.*'s reanalysis actually supports this finding, but it is noteworthy that neither the microclimate warming rates nor the macroclimate warming rates were strongly related to the thermophilization rates. In the main text of our paper, we acknowledge the large variation and high degree of stochasticity in the thermophilization rates overserved in our continental dataset reaching back to 1934, and we are most certainly aware that other drivers—such as drought, changes in soil acidity and nutrients, or the colonization of invasive species—may also affect community turnover. We clearly refer to these factors in the supplementary materials and study these other drivers extensively in other work. However, in the context of warming temperatures, we are confident that our interpretation that thermophilization in forest plant communities is primarily controlled by microclimate warming, and not macroclimate warming, is sound and supported by our analyses (*3*). We also fully agree that macroclimate warming can contribute to community thermophilization, as other studies have found [e.g., (*7*, *8*)], and we reference these studies in our paper without contradicting them. However, in our dataset this relationship was statistically not significant. Please note that Bertrand *et al.*'s claim that their "analysis further showed that canopy closure had cooled local climate conditions within forest stands and mitigated the thermophilization of plant communities as macroclimate warmed" does not represent an additional finding, as we have already shown this in our paper [see figure 1C and figure 2 of (*3*)].

As outlined above, the two main explanatory variables in Bertrand *et al.*'s model (i.e., macroclimate warming and changes in temperature buffering) represent our definition of forest microclimate warming in this study. Following Occam's razor principle, a more parsimonious statistical model (fewer predictor variables than in Bertrand *et al.*) to analyze the effect of microclimate warming on the microclimatic debt can therefore be formulated as follows: microclimatic debt ~ microclimate warming. The resulting marginal R^2 and AIC values for this model are 79.1% and −2463. The respective values for Bertrand *et al.*'s model with our data, which includes a much more complex formula with two predictor variables and their interaction effect, are 77.1% and −2257. These model evaluations show that our conclusion that the climatic debt in forest plant communities is primarily controlled by microclimate is sound and supported by our analyses and results. However, it is important to note that microclimate debt and microclimate warming are not independent of each other, because the latter was used to calculate the former. Although we agree with Bertrand *et al.* that macroclimate warming can affect the microclimatic debt, we would like to point out that this is not surprising given the strong correlation between macroclimate warming and microclimate warming, as shown in figure 1C of (*3*). Our understanding of the relative importance of micro- and macroclimate change for driving community responses to climate change is still far from complete and warrants further research.

Finally, we agree with Bertrand *et al.*'s universal statement that preserving forest biodiversity requires controlling global warming. The critical focus with regard to our report

is how, and at what scale, warming temperatures need to be quantified in order to better understand the response of forest biodiversity to climate change. In this regard, forest microclimate ecology has a lot to teach us.

REFERENCES

1. R. Geiger, R. H. Aron, P. Todhunter, *The Climate Near the Ground* (Rowman and Littlefield, 2003).
2. P. De Frenne, F. Zellweger, F. Rodríguez-Sánchez, B. R. Scheffers, K. Hylander, M. Luoto, M. Vellend, K. Verheyen, J. Lenoir, Global buffering of temperatures under forest canopies. *Nat. Ecol. Evol.* **3**, 744–749 (2019). doi:10.1038/s41559-019-0842-1 Medline
3. F. Zellweger, P. De Frenne, J. Lenoir, P. Vangansbeke, K. Verheyen, M. Bernhardt-Römermann, L. Baeten, R. Hédl, I. Berki, J. Brunet, H. Van Calster, M. Chudomelová, G. Decocq, T. Dirnböck, T. Durak, T. Heinken, B. Jaroszewicz, M. Kopecký, F. Máliš, M. Macek, M. Malicki, T. Naaf, T. A. Nagel, A. Ortmann-Ajkai, P. Petřík, R. Pielech, K. Reczyńska, W. Schmidt, T. Standovár, K. Świerkosz, B. Teleki, O. Vild, M. Wulf, D. Coomes, Forest microclimate dynamics drive plant responses to warming. *Science* **368**, 772–775 (2020). doi:10.1126/science.aba6880 Medline
4. R. Bertrand, F. Aubret, G. Grenouillet, A. Ribéron, S. Blanchet, Comment on "Forest microclimate dynamics drive plant responses to warming". *Science* **370**, eabd3850 (2020).
5. V. Vandvik, O. Skarpaas, K. Klanderud, R. J. Telford, A. H. Halbritter, D. E. Goldberg, Biotic rescaling reveals importance of species interactions for variation in biodiversity responses to climate change. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 22858–22865 (2020). 10.1073/pnas.2003377117 Medline
6. J. Lenoir, Rethinking climate context dependencies in biological terms. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23208–23210 (2020). 10.1073/pnas.2016537117 Medline
7. B. Fadrique, S. Báez, Á. Duque, A. Malizia, C. Blundo, J. Carilla, O. Osinaga-Acosta, L. Malizia, M. Silman, W. Farfán-Ríos, Y. Malhi, K. R. Young, F. Cuesta C, J. Homeier, M. Peralvo, E. Pinto, O. Jadan, N. Aguirre, Z. Aguirre, K. J. Feeley, Widespread but heterogeneous responses of Andean forests to climate change. *Nature* **564**, 207–212 (2018). doi:10.1038/s41586-018-0715-9 Medline
8. P. De Frenne, F. Rodríguez-Sánchez, D. A. Coomes, L. Baeten, G. Verstraeten, M. Vellend, M. Bernhardt-Römermann, C. D. Brown, J. Brunet, J. Cornelis, G. M. Decocq, H. Dierschke, O. Eriksson, F. S. Gilliam, R. Hédl, T. Heinken, M. Hermy, P. Hommel, M. A. Jenkins, D. L. Kelly, K. J. Kirby, F. J. G. Mitchell, T. Naaf, M. Newman, G. Peterken, P. Petrík, J. Schultz, G. Sonnier, H. Van Calster, D. M. Waller, G.-R. Walther, P. S. White, K. D. Woods, M. Wulf, B. J. Graae, K. Verheyen, Microclimate moderates plant responses to macroclimate warming. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18561–18565 (2013). doi:10.1073/pnas.1311190110 Medline

15 July 2020; accepted 6 November 2020
Published online 27 November 2020
10.1126/science.abd6193

2020 AAAS Fellows approved by the AAAS Council

In October 2020, the Council of the American Association for the Advancement of Science elected 489 members as Fellows of AAAS. These individuals will be recognized for their contributions to science and technology during the 2021 AAAS Annual Meeting. Presented by section affiliation, they are:

Section on Agriculture, Food, and Renewable Resources

Ann M. Bartuska, *Resources for the Future*

Carl Bernacchi, *U.S. Department of Agriculture – Agricultural Research Service*

Amy O. Charkowski, *Colorado State Univ.*

Clarice J. Coyne, *U.S. Department of Agriculture – Agricultural Research Service*

Geoffrey E. Dahl, *Univ. of Florida*

Roch E. Gaussoin, *Univ. of Nebraska-Lincoln*

Patrick M. Hayes, *Oregon State Univ.*

Thomas J. (TJ) Higgins, *CSIRO Agriculture and Food (Australia)*

Nancy Collins Johnson, *Northern Arizona Univ.*

Shibu Jose, *Univ. of Missouri*

Daniel Kliebenstein, *Univ. of California, Davis*

Rosemary Loria, *Univ. of Florida*

Shailaja K. Mani, *Baylor College of Medicine*

Rafael Muñoz-Carpena, *Univ. of Florida*

David D. Myrold, *Oregon State Univ.*

K. Raja Reddy, *Mississippi State Univ.*

Jean Ristaino, *North Carolina State Univ.*

Jeanne Romero-Severson, *Univ. of Notre Dame*

Pablo Juan Ross, *Univ. of California, Davis*

Jennifer L. Tank, *Univ. of Notre Dame*

William F. Tracy, *Univ. of Wisconsin-Madison*

Section on Anthropology

Margaret W. Conkey, *Univ. of California, Berkeley*

Anne Grauer, *Loyola Univ. Chicago*

Debbie Guatelli-Steinberg, *The Ohio State Univ.*

Edward B. Liebow, *American Anthropological Association*

J. Terrence McCabe, *Univ. of Colorado Boulder*

Denise Fay-Shen Su, *Cleveland Museum of Natural History*

Section on Astronomy

Nancy Susan Brickhouse, *Harvard-Smithsonian Center for Astrophysics*

John E. Carlstrom, *Univ. of Chicago*

Sean Carroll, *California Institute of Technology*

Timothy Heckman, *Johns Hopkins Univ.*

Paul Martini, *The Ohio State Univ.*

Norman Murray, *Canadian Institute for Theoretical Astrophysics*

Joan R. Najita, *National Science Foundation's NOIRLab*

Liese van Zee, *Indiana Univ.*

Risa Wechsler, *Stanford Univ.*

Ellen G. Zweibel, *Univ. of Wisconsin-Madison*

Section on Atmospheric and Hydrospheric Science

Ghassem R. Asrar, *Universities Space Research Association*

Elizabeth Boyer, *Pennsylvania State*

Deborah Bronk, *Bigelow Laboratory for Ocean Sciences*

Rong Fu, *Univ. of California, Los Angeles*

Isaac Held, *Princeton Univ. Atmospheric and Oceanic Sciences Program*

Forrest M. Hoffman, *Oak Ridge National Laboratory*

William K. M. Lau, *Univ. of Maryland*

Zhengyu Liu, *The Ohio State Univ.*

Natalie Mahowald, *Cornell Univ.*

Sally McFarlane, *U.S. Department of Energy*

Jerry Schubel, *Aquarium of the Pacific (Retired)*

Patricia L. Wiberg, *Univ. of Virginia*

Section on Biological Sciences

Mary Catherine Aime, *Purdue Univ.*

Suresh K. Alahari, *Louisiana State Univ. Health Sciences Center School of Medicine*

Gladys Alexandre, *Univ. of Tennessee, Knoxville*

Craig Reece Allen, *Univ. of Nebraska-Lincoln*

Sonia M. Altizer, *Univ. of Georgia*

Swathi Arur, *The Univ. of Texas MD Anderson Cancer Center*

Alison M. Bell, *Univ. of Illinois at Urbana-Champaign*

Elizabeth T. Borer, *Univ. of Minnesota*

Lisa Brooks, *National Human Genome Research Institute*

John Michael Burke, *Univ. of Georgia*

George A. Calin, *The Univ. of Texas MD Anderson Cancer Center*

Andrew G. Campbell, *Brown Univ.*

Alice Y. Cheung, *Univ. of Massachusetts Amherst*

Anita S. Chong, *Univ. of Chicago*

Gregory P. Copenhaver, *Univ. of North Carolina at Chapel Hill*

Leah E. Cowen, *Univ. of Toronto (Canada)*

Dana Crawford, *Case Western Reserve Univ.*

Charles F. Delwiche, *Univ. of Maryland, College Park*

Diana M. Downs, *Univ. of Georgia*

Jeffrey Dukes, *Purdue Univ.*

Peter Dunn, *Univ. of Wisconsin-Milwaukee*

Jonathan Eisen, *Univ. of California, Davis*

Eva Engvall, *Sanford Burnham Prebys Medical Discovery Institute*

Valerie Eviner, *Univ. of California, Davis*

Philip Martin Fearnside, *INPA - National Institute of Amazonian Research (Brazil)*

Gloria Cruz Ferreira, *Univ. of South Florida*

J. Patrick Fitch, *Los Alamos National Laboratory*

John W. Fitzpatrick, *Cornell Univ.*

Christopher Francklyn, *Univ. of Vermont*

Serita Frey, *Univ. of New Hampshire*

Andrea L. Graham, *Princeton Univ.*

Michael William Gray, *Dalhousie Univ. (Canada)*

Karen Jeanne Guillemin, *Univ. of Oregon*

Paul Hardin, *Texas A&M Univ.*

Stacey Lynn Harmer, *Univ. of California, Davis*

Jessica Hellmann, *Univ. of Minnesota*

Nancy Marie Hollingsworth, *Stony Brook Univ.*

Charles Hong, *Univ. of Maryland School of Medicine*

Laura Foster Huenneke, *Northern Arizona Univ.*

Mark O. Huising, *Univ. of California, Davis*

Travis Huxman, *Univ. of California, Irvine*

Kenneth D. Irvine, *Rutgers, The State Univ. of New Jersey*

Ursula Jakob, *Univ. of Michigan*

Janet K. Jansson, *Pacific Northwest National Laboratory*

Susan Kaech, *Salk Institute for Biological Studies*

Patricia Kiley, *Univ. of Wisconsin-Madison*

Joan Kobori, *Agouron Institute*

Barbara N. Kunkel, *Washington Univ. in St. Louis*

Armand Michael Kuris, *Univ. of California, Santa Barbara*

Pui-Yan Kwok, *Univ. of California, San Francisco*

Douglas Landis, *Michigan State Univ.*

Samuel J. Landry, *Tulane Univ. School of Medicine*

Eaton Edwards Lattman, *Univ. at Buffalo, the State Univ. of New York (Retired)*

Rodney L. Levine, *National Heart, Lung, and Blood Institute/NIH*

Han Liang, *The Univ. of Texas MD Anderson Cancer Center*

Senjie Lin, *Univ. of Connecticut*

Published by AAAS

Hiten D. Madhani, *Univ. of California, San Francisco*

Jennifer B. H. Martiny, *Univ. of California, Irvine*

John McCutcheon, *Arizona State Univ.*

Rima McLeod, *Univ. of Chicago*

Paula McSteen, *Univ. of Missouri-Columbia*

Matthew Meyerson, *Dana-Farber Cancer Institute/ Harvard Medical School*

Constance Millar, *U.S. Forest Service*

Lisa A. Miller, *Univ. of California, Davis*

Beronda L. Montgomery, *Michigan State Univ.*

Tuli Mukhopadhyay, *Indiana Univ.*

Katsuhiko (Katsu) Murakami. *Pennsylvania State Univ.*

William J. Murphy, *Texas A&M Univ.*

Rama Natarajan, *City of Hope National Medical Center*

Nicholas E. Navin, *The Univ. of Texas MD Anderson Cancer Center*

Anthony V. Nicola, *Washington State Univ.*

Basil Nikolau, *Iowa State Univ.*

E. Michael Ostap, *Univ. of Pennsylvania Perelman School of Medicine*

Franklin Wayne Outten, *Univ. of South Carolina*

Abraham Palmer, *Univ. of California, San Diego*

Maria C. Pellegrini, *W. M. Keck Foundation*

Len Pennacchio, *Lawrence Berkeley National Laboratory*

Philip S. Perlman, *Howard Hughes Medical Institute.*

James Pinckney, *Univ. of South Carolina*

Judith A. Potashkin, *Rosalind Franklin Univ. of Medicine and Science*

P. Hemachandra Reddy, *Texas Tech Univ. Health Sciences Center*

William S. Reznikoff, *Marine Biological Laboratory*

Joan T. Richtsmeier, *Pennsylvania State Univ.*

Isidore Rigoutsos, *Thomas Jefferson Univ.*

Charles Rock, *St. Jude Children's Research Hospital*

Antonis Rokas, *Vanderbilt Univ.*

Jeffrey Ross-Ibarra, *Univ. of California, Davis*

James A. Roth, *Iowa State Univ.*

Daniel Schaid, *Mayo Clinic*

G. Eric Schaller, *Dartmouth Univ.*

Jeremiah Scharf, *Massachusetts General Hospital*

Karen Sears, *Univ. of California, Los Angeles*

Mark Seielstad, *Univ. of California, San Francisco*

Peter Setlow, *UConn Health*

Sally Shaywitz, *Yale Univ.*

Alan Shuldiner, *Univ. of Maryland School of Medicine*

Nathan Michael Springer, *Univ. of Minnesota*

Jason E. Stajich, *Univ. of California, Riverside*

James V. Staros, *Univ. of Massachusetts Amherst*

David Johnston Stewart, *Cold Spring Harbor Laboratory*

Joel A. Swanson, *Univ. of Michigan Medical School*

Rick L. Tarleton, *Univ. of Georgia*

Nektarios Tavernarakis, *Foundation for Research and Technology-Hellas/Univ. of Crete Medical (Greece)*

Eric W. Triplett, *Univ. of Florida*

Geoffrey C. Trussell, *Northeastern Univ.*

Walter Reinhart Tschinkel, *Florida State Univ.*

Kan Wang, *Iowa State Univ.*

Joy Ward, *Univ. of Kansas*

Vassie Ware, *Lehigh Univ.*

Stephen T. Warren, *Emory Univ. School of Medicine*

Wyeth W. Wasserman, *BC Children's Hospital/ Univ. of British Columbia (Canada)*

Daniel J. Wozniak, *The Ohio State Univ.*

Jin-Rong Xu, *Purdue Univ.*

Soojin Yi, *Georgia Institute of Technology*

Havva Fitnat Yildiz, *Univ. of California, Santa Cruz*

Fanxiu Zhu, *Florida State Univ.*

Lee Zou, *Massachusetts General Hospital Cancer Center/Harvard Medical School*

Section on Chemistry

José R. Almirall, *Florida International Univ.*

Rohit Bhargava, *Univ. of Illinois at Urbana-Champaign*

Paul V. Braun, *Univ. of Illinois at Urbana-Champaign*

Joan Blanchette Broderick, *Montana State Univ.*

Felix (Phil) N. Castellano, *North Carolina State Univ.*

David E. Chavez, *Los Alamos National Laboratory*

Kelsey D. Cook, *National Science Foundation*

Yi Cui, *Stanford Univ.*

Wibe A. de Jong, *Lawrence Berkeley National Laboratory*

William Dichtel, *Northwestern Univ.*

Vishva Dixit, *Genentech, Inc.*

Paul J. Dyson, *Swiss Federal Institute of Technology Lausanne*

Laura Gagliardi, *Univ. of Chicago*

Jiaxing Huang, *Northwestern Univ.*

Prashant K. Jain, *Univ. of Illinois at Urbana-Champaign*

Phillip E. Klebba, *Kansas State Univ.*

Kenneth L. Knappenberger, *Pennsylvania State Univ.*

Yamuna Krishnan, *Univ. of Chicago*

Jason S. Lewis, *Memorial Sloan Kettering Cancer Center*

Hongbin Li, *Univ. of British Columbia (Canada)*

David R. Liu, *Harvard Univ.*

Tianbo Liu, *Univ. of Akron*

Tadeusz (Ted) Franciszek Molinski, *Univ. of California, San Diego*

Janet R. Morrow, *Univ. at Buffalo, the State Univ. of New York*

John W. Olesik, *The Ohio State Univ.*

Nicola Pohl, *Indiana Univ.*

Daniel Raftery, *Univ. of Washington*

Michael D. Sevilla, *Oakland Univ.*

David S. Sholl, *Georgia Institute of Technology*

Sara E. Skrabalak, *Indiana Univ.*

Brian Space, *North Carolina State Univ.*

Raymond C. Stevens, *Univ. of Southern California*

James M. Takacs, *Univ. of Nebraska-Lincoln*

Chuanbing Tang, *Univ. of South Carolina*

H. Holden Thorp, Science *family of journals*

Gregory Tschumper, *Univ. of Mississippi*

Christopher D. Vanderwal, *Univ. of California, Irvine*

Nathalie A. Wall, *Univ. of Florida*

Rory Waterman, *Univ. of Vermont*

Charles Weschler, *Rutgers, The State Univ. of New Jersey*

Robert F. Williams, *Los Alamos National Laboratory*

Frankie Wood-Black, *Northern Oklahoma College*

Karen Wooley, *Texas A&M Univ.*

Peidong Yang, *Univ. of California, Berkeley*

Section on Dentistry and Oral Health Sciences

Renny Theodore Franceschi, *Univ. of Michigan*

Dennis F. Mangan, *Chalk Talk Science Project*

Frank C. Nichols, *Univ. of Connecticut School of Dental Medicine*

Stefan Hans-Klaus Ruhl, *Univ. at Buffalo, the State Univ. of New York*

Section on Education

James Bell, *Center for Advancing of Informal Science Education.*

Michael J. Dougherty, *GenomEducation Consulting/ Univ. of Colorado School of Medicine*

John Kermit Haynes, *Morehouse College*

Henry Vincent Jakubowski, *College of St. Benedict/ St. John's Univ.*

Stacey Kiser, *Lane Community College*

Richard L. Kopec, *St. Edward's Univ.*

Xiufeng Liu, *Univ. at Buffalo, the State Univ. of New York*

David J. Marcey, *California Lutheran Univ.*

Marsha Lakes Matyas, *Evaluation for Excellence*

Linda Nicholas-Figueroa, *Iḷisaġvik College*

Dee Unglaub Silverthorn, *The Univ. of Texas at Austin Dell Medical School*

Edward J. Smith, *Virginia Polytechnic Institute and State Univ.*

David W. Szymanski, *Bentley Univ.*

Edna Tan, *Univ. of North Carolina at Greensboro*

Stephen Young, *TriCore Reference Laboratories*

Hinda Zlotnik, *Retired*

Section on Engineering

Mohammad S. Alam, *Texas A&M Univ.*

Laura Albert, *University of Wisconsin-Madison*

William R. Bickford, *L'Oréal, Inc.*

L. Catherine Brinson, *Duke Univ.*

Ruben G. Carbonell, *North Carolina State Univ.*

Michael L. Chabinyc, *Univ. of California, Santa Barbara*

Manish Chhowalla, *Univ. of Cambridge (U.K.)*

Edwin K. P. Chong, *Colorado State Univ.*

Kristen P. Constant, *Iowa State Univ.*

Susan Daniel, *Cornell Univ.*

Angela Phillips Diaz, *Univ. of California, San Diego*

Elizabeth C. Dickey, *North Carolina State Univ.*

Peter S. Fedkiw, *North Carolina State Univ.*

Eric M. Furst, *Univ. of Delaware*

Sharon Gerecht, *Johns Hopkins Univ.*

Richard D. Gitlin, *Univ. of South Florida*

Michael C. Jewett, *Northwestern Univ.*

Vistasp M. Karbhari, *The Univ. of Texas at Arlington*

Michael R. Kessler, *North Dakota State Univ.*

Behrokh Khoshnevis, *Univ. of Southern California*

Kristi L. Kiick, *Univ. of Delaware*

Catherine Klapperich, *Boston Univ.*

Gerhard Klimeck, *Purdue Univ.*

Sanjay Kumar, *Univ. of California, Berkeley*

Ju Li, *Massachusetts Institute of Technology*

JoAnn Slama Lighty, *Boise State Univ.*

Ivan M. Lorković, *Raytheon Vision Systems*

Laura Marcu, *Univ. of California, Davis*

Sudip K. Mazumder, *Univ. of Illinois at Chicago*

Triantafillos (Lakis) Mountziaris, *Univ. of Massachusetts Amherst*

Uday B. Pal, *Boston Univ.*

Ah-Hyung (Alissa) Park, *Columbia Univ.*

Hridesh Rajan, *Iowa State Univ.*

Gintaras Reklaitis, *Purdue Univ.*

Robert Oliver Ritchie, *Univ. of California, Berkeley*

J. Paul Robinson, *Purdue Univ.*

Nancy R. Sottos, *Univ. of Illinois at Urbana-Champaign*

Georgia (Gina) D. Tourassi, *Oak Ridge National Laboratory*

Paul J. Turinsky, *North Carolina State Univ.*

John L. Volakis, *Florida International Univ.*

Qing Wang, *Pennsylvania State Univ.*

Lan Yang, *Washington Univ. in St. Louis*

Section on General Interest in Science and Engineering

Nan Broadbent, *Seismological Society of America*

Tinsley Davis, *National Association of Science Writers*

Linda D. Harrar, *WGBH Educational Foundation*

James H. Lambert, *Univ. of Virginia*

Andrew D. Maynard, *Arizona State Univ.*

Jeremy B. Searle, *Cornell Univ.*

Ronald M. Thom, *Pacific Northwest National Laboratory (Emeritus)*

Cliff Wang, *U.S. Army Research Office/North Carolina State Univ.*

Nan Yao, *Princeton Univ.*

Section on Geology and Geography

Li An, *San Diego State Univ.*

David Cairns, *Texas A&M Univ.*

Richard Walter Carlson, *Carnegie Institution for Science*

Charles B. (Chuck) Connor, *Univ. of South Florida*

Peter B. de Menocal, *Woods Hole Oceanographic Institution*

Andrea Donnellan, *NASA Jet Propulsion Laboratory*

Joshua S. Fu, *Univ. of Tennessee, Knoxville*

George Helz, *Univ. of Maryland, College Park*

Tessa M. Hill, *Univ. of California, Davis*

David A. Hodell, *Univ. of Cambridge (U.K.)*

(Max) Qinhong Hu, *The Univ. of Texas at Arlington*

Hitoshi Kawakatsu, *The Univ. of Tokyo (Japan)*

Sheryl Luzzadder-Beach, *The Univ. of Texas at Austin*

Vicki McConnell, *Geological Society of America*

Carolyn Olson, *U.S. Geological Survey*

Lewis A. Owen, *North Carolina State Univ.*

David Sandwell, *Scripps Institution of Oceanography*

Nathan Dale Sheldon, *Univ. of Michigan*

May Yuan, *The Univ. of Texas at Dallas*

Section on History and Philosophy of Science

Colin Allen, *Univ. of Pittsburgh*

Rachel Ankeny, *Univ. of Adelaide (Australia)*

David Cassidy, *Hofstra Univ.*

Marsha L. Richmond, *Wayne State Univ.*

Section on Industrial Science and Technology

Suresh K. Bhargava, *RMIT Univ. (Australia)*

Aaron Dominguez, *Catholic Univ. of America*

Johney B. Green, *National Renewable Energy Laboratory*

James D. Kindscher, *Univ. of Kansas Medical Center*

Daniela Rus, *Massachusetts Institute of Technology*

Steven Suib, *Univ. of Connecticut*

Erik B. Svedberg, *National Academies of Sciences, Engineering, and Medicine*

Section on Information, Computing, and Communication

James Allen, *Univ. of Rochester/Institute for Human and Machine Cognition*

James Hampton Anderson, *Univ. of North Carolina at Chapel Hill*

Carla Brodley, *Northeastern Univ.*

Lorrie Cranor, *Carnegie Mellon Univ.*

Pedro Domingos, *Univ. of Washington*

Kenneth D. Forbus, *Northwestern Univ.*

Yolanda Gil, *Univ. of Southern California*

Leana Golubchik, *Univ. of Southern California*

Yuri Gurevich, *Univ. of Michigan*

Murat Kantarcioglu, *The Univ. of Texas at Dallas*

Maria Klawe, *Harvey Mudd College*

Peter M. Kogge, *Univ. of Notre Dame*

Patrick Drew McDaniel, *Pennsylvania State Univ.*

Debasis Mitra, *Columbia Univ.*

John Douglas Owens, *Univ. of California, Davis*

Timothy Mark Pinkston, *Univ. of Southern California*

William C. Regli, *Univ. of Maryland, College Park*

Munindar P. Singh, *North Carolina State Univ.*

Anuj Srivastava, *Florida State Univ.*

David Touretzky, *Carnegie Mellon Univ.*

Jeffrey S. Vetter, *Oak Ridge National Laboratory*

Toby Walsh, *Univ. of New South Wales - Sydney and CSIRO Data61 (Australia)*

Daniel S. Weld, *Univ. of Washington/Allen Institute for Artificial Intelligence*

Hui Xiong, *Rutgers, The State Univ. of New Jersey*

Section on Linguistics and Language Sciences

John Baugh, *Washington Univ. in St. Louis*

Bryan Gick, *Univ. of British Columbia (Canada)*

Colin Phillips, *Univ. of Maryland*

Published by AAAS

Joan A. Sereno, *Univ. of Kansas*

Matthew W. Wagers, *Univ. of California, Santa Cruz*

Section on Mathematics

Harold P. Boas, *Texas A&M Univ.*

Leslie Hogben, *Iowa State Univ./American Institute of Mathematics*

Kristin Lauter, *Microsoft Research*

Paul K. Newton, *Univ. of Southern California*

Esmond G. Ng, *Lawrence Berkeley National Laboratory*

Karen Hunger Parshall, *Univ. of Virginia*

Malgorzata Peszynska, *Oregon State Univ.*

Jack Xin, *Univ. of California, Irvine*

Section on Medical Sciences

Stephen B. Baylin, *Johns Hopkins Univ. School of Medicine*

Barbara D. Beck, *Gradient*

Yasmine Belkaid, *National Institute of Allergy and Infectious Diseases/NIH*

Barry B. Bercu, *Univ. of South Florida*

Keith C. Cheng, *Pennsylvania State Univ. College of Medicine*

Shi-Yuan Cheng, *Northwestern Univ. Feinberg School of Medicine*

Ronald W. Davis, *Stanford Univ.*

Catherine Drennan, *Massachusetts Institute of Technology*

Dongsheng Duan, *Univ. of Missouri*

Carol Fuzeti Elias, *Univ. of Michigan*

Hudson Freeze, *Sanford Burnham Prebys Medical Discovery Institute*

Marcia B. Goldberg, *Massachusetts General Hospital/Harvard T. H. Chan School of Public Health*

David H. Gutmann, *Washington Univ. School of Medicine in St. Louis*

Thomas H. Haines, *City College of New York (Retired)*

Raymond C. Harris, *Vanderbilt Univ. School of Medicine*

Jeffrey P. Krischer, *Univ. of South Florida*

Thomas E. Lane, *Univ. of California, Irvine*

W. Jonathan Lederer, *Univ. of Maryland School of Medicine*

Bruce T. Liang, *Univ. of Connecticut School of Medicine*

Jeffrey D. Lifson, *Frederick National Laboratory for Cancer Research*

Faina Linkov, *Duquesne Univ.*

Shan-Lu Liu, *The Ohio State Univ.*

Karl L. Magleby, *Univ. of Miami, School of Medicine*

Sendurai Mani, *The Univ. of Texas MD Anderson Cancer Center*

Douglas L. Mann, *Washington Univ. School of Medicine in St. Louis*

Rodger P. McEver, *Oklahoma Medical Research Foundation*

Ross Erwin McKinney Jr., *Association of American Medical Colleges*

Hiroyoshi Nishikawa, *National Cancer Center/ Nagoya Univ. (Japan)*

Richard M. Peek, *Vanderbilt Univ. Medical Center*

Sallie R. Permar, *Duke Univ.*

W. Kimryn Rathmell, *Vanderbilt Univ. Medical Center*

D. Nageshwar Reddy, *Asian Institute of Gastroenterology (India)*

John Jeffrey Reese, *Vanderbilt Univ. Medical Center*

Alan Saghatelian, *Salk Institute for Biological Studies*

Suzanne Scarlata, *Worcester Polytechnic Institute*

Brian Leslie Strom, *Rutgers, The State Univ. of New Jersey*

Jie Tian, *Chinese Academy of Sciences (China)*

Jerrold Ross Turner, *Brigham and Women's Hospital/ Harvard Medical School*

Matthew Bret Weinger, *Vanderbilt Univ. Medical Center*

Section on Neuroscience

Patrick Aebischer, *Swiss Federal Institute of Technology Lausanne*

Michel Baudry, *Western Univ. of Health Sciences*

Nicole Calakos, *Duke Univ.*

Gabriel Corfas, *Univ. of Michigan*

Aaron DiAntonio, *Washington Univ. School of Medicine in St. Louis*

Nita A. Farahany, *Duke Univ.*

Eva Lucille Feldman, *Univ. of Michigan*

Eberhard Erich Fetz, *Univ. of Washington*

Alan L. Goldin, *Univ. of California, Irvine*

Steve A. N. Goldstein, *Univ. of California, Irvine School of Medicine*

John Krystal, *Yale Univ.*

Debomoy (Deb) K. Lahiri, *Indiana Univ.*

Stephen G. Lisberger, *Duke Univ.*

Wendy Blair Macklin, *Univ. of Colorado Denver*

Stefan M. Pulst, *Univ. of Utah*

Nirao M. Shah, *Stanford Univ.*

Steven L. Small, *The Univ. of Texas at Dallas*

Paul Taghert, *Washington Univ. School of Medicine in St. Louis*

Rachel Tyndale, *Univ. of Toronto (Canada)/Centre for Addiction and Mental Health*

Linda Jo Van Eldik, *Univ. of Kentucky College of Medicine*

Section on Pharmaceutical Sciences

Patricia Babbitt, *Univ. of California, San Francisco*

Joseph R. Haywood, *Michigan State Univ.*

Julie A. Johnson, *Univ. of Florida*

Lyn H. Jones, *Dana-Farber Cancer Institute*

M. N. V. Ravi Kumar, *Texas A&M Univ.*

Susan L. Mooberry, *Univ. of Texas Health Science Center at San Antonio*

Walter H. Moos, *Univ. of California, San Francisco*

Raymond Felix Schinazi, *Emory Univ. School of Medicine*

Thomas D. Schmittgen, *Univ. of Florida*

Section on Physics

Charles H. Bennett, *IBM Thomas J. Watson Research Center*

Eberhard Bodenschatz, *Max Planck Institute for Dynamics and Self Organization (Germany)*

Steven E. Boggs, *Univ. of California, San Diego*

Jesse Brewer, *Univ. of British Columbia (Canada)*

Bulbul Chakraborty, *Brandeis Univ.*

Andre De Gouvea, *Northwestern Univ.*

Peter Fisher, *Massachusetts Institute of Technology*

Chris L. Fryer, *Los Alamos National Laboratory*

Alexandra Gade, *Michigan State Univ.*

Graciela Gelmini, *Univ. of California, Los Angeles*

Neil Gershenfeld, *Massachusetts Institute of Technology*

Tony Gherghetta, *Univ. of Minnesota*

Marcelo Jaime, *Los Alamos National Laboratory*

Spencer R. Klein, *Lawrence Berkeley National Laboratory/Univ. of California, Berkeley National Laboratory/Univ. of California, Berkeley*

Yuri V. Kovchegov, *The Ohio State Univ.*

Ying-Cheng Lai, *Arizona State Univ.*

Konrad W. Lehnert, *National Institute of Standards and Technology/Univ. of Colorado Boulder*

Manfred Lindner, *Max Planck Institute for Nuclear Physics (Germany)*

Eric Mazur, *Harvard Univ.*

Michael A. McGuire, *Oak Ridge National Laboratory*

Roberto Morandotti, *INRS - National Institute of Scientific Research (Canada)*

Jason Petta, *Princeton Univ.*

Thomas Proffen, *Oak Ridge National Laboratory*

Laura J. Pyrak-Nolte, *Purdue Univ.*

Talat Shahnaz Rahman, *Univ. of Central Florida*

Susan Seestrom, *Sandia National Laboratories*

Jonathan V. Selinger, *Kent State Univ.*

Arthur John Stewart Smith, *Princeton Univ.*

Christopher Stubbs, *Harvard Univ.*

AAAS NEWS & NOTES

Nandini Trivedi, *The Ohio State Univ.*

Yuhai Tu, *IBM Thomas J. Watson Research Center*

Clare Yu, *Univ. of California, Irvine*

Anvar A. Zakhidov, *The Univ. of Texas at Dallas*

Section on Psychology

Tammy D. Allen, *Univ. of South Florida*

Peter R. Finn, *Indiana Univ.*

Howard Goldstein, *Univ. of South Florida*

William (Bill) P. Hetrick, *Indiana Univ.*

Stefan Hofmann, *Boston Univ.*

Elaine Hull, *Florida State Univ.*

Blair T. Johnson, *Univ. of Connecticut*

Angeline S. Lillard, *Univ. of Virginia*

Raymond G. Miltenberger, *Univ. of South Florida*

Jay Myung, *The Ohio State Univ.*

Steven L. Neuberg, *Arizona State Univ.*

Barbara A. Wanchisen, *National Academies of Sciences, Engineering, and Medicine*

Cathy Spatz Widom, *John Jay College of Criminal Justice*

Section on Social, Economic, and Political Sciences

John Maron Abowd, *U.S. Census Bureau/Cornell Univ.*

Melissa S. Anderson, *Univ. of Minnesota*

Janet Box-Steffensmeier, *The Ohio State Univ.*

R. Alta Charo, *Univ. of Wisconsin-Madison*

Nazli Choucri, *Massachusetts Institute of Technology*

Elizabeth Cooksey, *The Ohio State Univ.*

Paul Allen David, *Stanford Univ.*

Joane P. Nagel, *Univ. of Kansas*

Kristen Olson, *Univ. of Nebraska-Lincoln*

Susan E. Short, *Brown Univ.*

Section on Societal Impacts of Science and Engineering

Gregg M. Garfin, *Univ. of Arizona*

Leah Gerber, *Arizona State Univ.*

Ramanan Laxminarayan, *Center for Disease Dynamics, Economics & Policy/Princeton Univ.*

Mary E. Maxon, *Lawrence Berkeley National Laboratory*

James Bradley Miller, *Smithsonian National Museum of Natural History*

Oladele (Dele) A. Ogunseitan, *Univ. of California, Irvine*

Lawrence J. Satkowiak, *Oak Ridge National Laboratory*

Vaughan Charles Turekian, *National Academies of Sciences, Engineering, and Medicine*

Section on Statistics

Sudipto Banerjee, *Univ. of California, Los Angeles*

David L. Banks, *Duke Univ.*

Deborah J. Donnell, *Fred Hutchinson Cancer Research Center*

Timothy C. Hesterberg, *Google, Inc.*

Qi Long, *Univ. of Pennsylvania*

Ying Lu, *Stanford Univ. School of Medicine*

Richard L. Smith, *Univ. of North Carolina at Chapel Hill*

Elizabeth A. Stuart, *Johns Hopkins Bloomberg School of Public Health*

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CLIMATOLOGY

A dangerous trend

How anthropogenically driven climate change is affecting heat waves and drought is one of the most important environmental issues facing societies around the globe. Zhang *et al.* present a 260-year-long record of temperature and soil moisture over inner East Asia that reveals an abrupt shift to hotter and drier conditions (see the Perspective by Zhang and Fang). Extreme episodes of hotter and drier climate over the past 20 years, which are unprecedented in the earlier records, are caused by a positive feedback loop between soil moisture deficits and surface warming and potentially represent the start of an irreversible trend. —HJS

Science, this issue p. 1095; see also p. 1037

East Asia, including the semiarid grasslands of Mongolia, has experienced a shift toward a hotter and drier climate in the past 20 years.

PHOTO: MAURITIUS IMAGES GMBH / ALAMY STOCK PHOTO

RIBOSOME STRUCTURE

Quality control in mitochondria

Human mitochondria have their own genome and ribosomes called mitoribosomes that respectively encode and synthesize essential subunits of complexes that use the energy from the oxidation of metabolites to drive the synthesis of adenosine triphosphate (ATP). These complexes are key to the health of the cell. Desai *et al*. studied a mitoribosome-associated quality control pathway that prevents aberrant translation. They purified mitoribosomes under conditions designed to induce stalling and determined the structures of two intermediates in the rescue pathway. These structures revealed two proteins that eject the unfinished polypeptide chain and peptidyl transfer RNA from the ribosome. Their cryo–electron microscopy dataset also revealed additional states that may correspond to intermediates in the mitochondrial translation elongation cycle. —VV

Science, this issue p. 1105

TOPOLOGICAL SYSTEMS

When all routes are exceptional

The exploration of non-Hermitian physics and parity-time symmetry have provided a route to develop a wealth of exotic physical effects. In such dissipative systems, the balance of gain and loss of the system lead to what are called exceptional points, or "sweet spots," which relate to optimal device operation or material properties. The ability to tune the gain and loss over a range of system properties leads to exceptional arcs. Tang *et al*. show that systems can be designed in which the tuning of multiple parameters leads to a crossroads, or nexus, of exceptional arcs. Illustrating the effect in an acoustic system, the same properties should be attainable over various types of dissipative systems and thus provide a versatile route to fine-tune optimal performance of materials and devices. —ISO

Science, this issue p. 1077

3D PRINTING

Driving the pores away

The formation of "keyholes" (vapor-filled depressions) during additive manufacturing leads to porosity, which degrades alloy performance, especially fracture properties, and remains a big challenge for the 3D printing of metals. Zhao *et al*. used high-speed x-ray imaging to take a detailed look at how keyhole formation connects to porosity in a titanium alloy. They found that instability at the keyhole tip drives pores away to get trapped in the solidification front. Understanding this process and the operating parameters under which it occurs provides a roadmap for avoiding porosity and building high-quality metal parts. —BG

Science, this issue p. 1080

SYNTHETIC BIOLOGY

A logic to cell-cell recognition

There has been exciting progress in the field of cancer immunotherapy, which harnesses a patient's own immune system to kill cancer cells. However, achieving precise recognition of cancer cells remains challenging. Cells engineered with synthetic Notch (synNotch) receptors bind to specific antigens, and binding induces the expression of defined genes. Williams *et al*. used synNotch modules as transcriptional connectors that daisy-chain

Published by AAAS

together multiple receptors. They engineered T cells that can recognize up to three target antigens expressed on or inside cancer cells and integrated these inputs to achieve NOT, AND, and OR logic. The engineered cells achieved precise recognition of targeted cancer cells. —VV

Science, this issue p. 1099

SEX DETERMINATION

Paternal factor specifies female wasps

Not all animals have specialized sex chromosomes to determine their sex. In hymenopteran insects, for example, unfertilized eggs become males and fertilized eggs become females. Prior work showed that the paternal genome provides instruction for female development. Zou *et al*. identified a sex determination instructor gene, *wasp overruler of masculinization*, with parent-of-origin effect, in the parasitoid *Nasonia vitripennis*. It is only transcribed from the paternally provided genome in fertilized eggs to initiate female development. This discovery provides insights into the molecular basis and evolution of sex determination. —BAP

Science, this issue p. 1115

MARINE ROBOTICS

Deep dive

Mapping the ocean's biochemistry will lead to a better understanding of Earth's biogeochemical cycles. Breier *et al*. developed an untethered autonomous underwater vehicle, which they named *Clio*, to map vertical profiles of the water column through high-resolution sensor surveys and sample-return missions. *Clio* could controllably descend through the water column at a rate of 0.8 meters per second with a maximum depth of 6000 meters. In 2019, *Clio* was deployed to map 1144 kilometers of the Sargasso Sea to a depth of 1000 meters. —MML

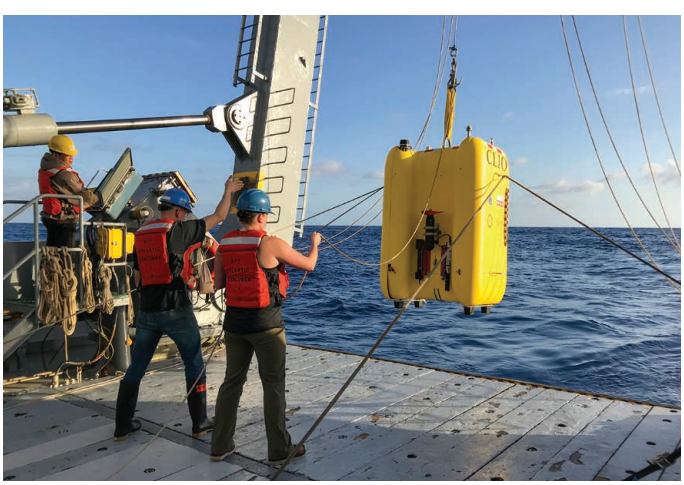
Sci. Robot. **5** abc7104 (2020).

DRUGS

Helping asthmatics breathe better

β_2-adrenoreceptor agonists used for acute relief of airway constriction in asthmatic patients become less effective with repeated use because of receptor desensitization. They also increase the risk of death. Using cells, lung slices, and mouse models of asthma, Wang *et al*. showed that osthole, a compound derived from a traditional Chinese medicine, promoted airway relaxation through mechanisms other than activating β_2-adrenoreceptors. These results suggest that variants of osthole could be developed to induce bronchodilation without desensitizing receptors or increasing the risk of death. —WW

Sci. Signal. **13**, eaax0273 (2020).



Researchers launch *Clio*, an autonomous deep-ocean exploration and sample-collection vehicle.

IN OTHER JOURNALS

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

As the climate changes, mixed-grass prairie becomes vulnerable to dust-bowl formation.

EDUCATION

Preaching outside of the choir

Workshops and training programs centered on sharing evidence-based instructional practices (EBIPs) for better engaging students in science, technology, engineering, and mathematics (STEM) learning are widely available. However, how and whether faculty share their new knowledge with their colleagues remain unclear. Lane *et al*. conducted a multi-institutional study to determine who EBIP users talked to about their teaching practices. Interview, survey, and social network analyses revealed that faculty who implemented EBIPs preferentially talked to each other, essentially slowing the pace of EBIP adoption. This analysis confirms that faculty teaching interactions are critical in promoting institutional change in higher education and that a systems approach to altering educational structures, environments, and policies is needed to engage the full range of faculty in teaching conversations. —MMc

Proc. Natl. Acad. Sci. U.S.A. **117**, 22665 (2020).

BIOTECHNOLOGY

PORE-cupine for RNA structures

Unlike DNA, RNA can fold into a staggering number of distinct structures that perform diverse functions. Some structures transfer genetic information, and others regulate their own activities and life spans to act as catalysts for crucial reactions inside the cell. Our ability to visualize or predict the RNA structures that a cell can make is limited, hindering exploitation for RNA therapeutics. Aw *et al*. improved on a technology that uses nanopore technology to sequence RNA by measuring the changing current as the RNA nucleotides are pulled through the nanopore. This "PORE-cupine" technology incorporates structural probing by chemical modification and machine learning to capture structural

PHOTO: CPAUL LETHABY BERMUDA INSTITUTE OF OCEAN SCIENCES

PHOTO: JOHN SIRLIN/ALAMY STOCK PHOTO

PLANT SCIENCE

Photosynthesis and the Dust Bowl

In the 1930s, the central United States turned into a dust bowl, which disrupted agriculture, economies, social structures, and ecosystems and may hold lessons for a future of increasing climate instability. Knapp *et al*. replicated some of the Dust Bowl conditions in Kansas and Wyoming to determine why a grass characterized by the C3 photosynthetic pathway replaced seemingly better-adapted C4 grasses during the drought despite its preference for cooler temperature and higher soil moisture. The key lies in the timing of precipitation. Drier summer months and wetter cool months during this 4-year period favored the growth of the C3 grasses, which outcompeted the C4 grasses. —PJH

Proc. Natl. Acad. Sci. U.S.A. **117**, 22249 (2020).

information on a genome scale. The technique can detect structural differences between closely related RNAs encoded by the same genes. —DJ

Nat. Biotechnol. 10.1038/s41587-020-0712-z (2020).

NEUROSCIENCE

How a brain controls a computer

Most brain-computer interfaces (BCIs) that use neuron recordings have analyzed the activity of those neurons that contribute directly to the decoded BCI output. Liu and Schieber found that although only a few primary motor cortex (M1) units controlled a closed-loop BCI, substantial numbers of non-BCI units were likewise modulated in relation to the task, not only in frontal motor areas (area M1 and the dorsal and ventral premotor cortex) but also in parietal areas (somatosensory cortex and the anterior intraparietal area). All of these cortical areas thus participated both in natural control of voluntary limb movement and in a more general system for closed-loop control of an effector being moved to a visual target. Harnessing the activity of units from multiple cortical areas might help in the development of next-generation BCIs. —PRS

eNeuro **7**, ENEURO.0376-20.2020 (2020).

NANOMATERIALS

Kinetic control of hierarchical growth

Sequential seeding of nanocrystals can be used to control and develop complex shapes during growth. Smith *et al*. explored the effects of limiting adatom addition versus diffusion for gold or palladium grown on gold seeds. Fast growth conditions promote branched nanocrystals along the <111> directions. The addition of hydrochloric acid lowered the pH and increased the halide concentration, slowing the rate of metal precursor reduction. Fast growth promoted concave shapes, whereas slower growth favored convex ones. By using the production of one growth cycle as the seed for the next one, 85 distinct hierarchical shapes were grown. —MSL

ACS Nano 10.1021/acsnano.0c07384 (2020).

KIDNEY DISEASE

Weighing in on chronic kidney disease

Previous studies have reported correlations between adiposity and chronic kidney disease, but it is not clear whether this is an independent association or if there is a causal relationship. To address this question systematically, Zhu *et al*. analyzed prospectively collected data from almost 300,000 participants in the UK Biobank. By taking advantage of Mendelian randomization, the authors showed that diabetes and high blood pressure, which are known kidney disease risk factors, played an even larger role than expected. However, adiposity itself was also an independent causative factor for the development of chronic kidney disease. —YN

J. Am. Soc. Nephrol. 10.1681/ASN.2020050679 (2020).

STELLAR POPULATIONS

Too hot for helium to handle

Helium has a high ionization potential of 24.6 electron volts. Galaxies that have emission lines of He II in their spectra must contain sufficiently hot sources, such as extremely massive stars, to ionize interstellar helium. Pérez-Montero *et al*. modeled multiple emission lines in the spectra of a sample of galaxies that each show He II emission. They found that the effective temperatures of the radiation fields were higher than can be explained by conventional stellar population models. The authors argue that populations of Wolf-Rayet stars with low abundances of elements heavier than helium, surrounded by clumpy gas, can explain the observations. —KTS

Astron. Astrophys. **643**, A80 (2020).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

SYNTHETIC BIOLOGY

Engineering immune responses

Cytokines mediate various immune responses and have potential as drugs to treat cancer and other diseases. However, cytokines have diverse and competing roles; although some cytokine therapies are approved, they come with side effects and their efficacy can be limited. In a Perspective, Li and Lim discuss the synthetic biology approaches to engineering better cytokine-based drugs. They discuss improving on naturally occurring cytokines, engineering cytokine-signaling circuits for use with engineered cell therapies, and the possibility of designing cytokines that do not exist naturally. If efficacious, such approaches could improve the potency of cytokines as drugs and add to our understanding of cytokine-based communication. —GKA

Science, this issue p. 1034

PROTEIN MODELING

Understanding what drives proteins

Computational molecular physics (CMP) aims to leverage the laws of physics to understand not just static structures but also the motions and actions of biomolecules. Applying CMP to proteins has required either simplifying the physical models or running simulations that are shorter than the time scale of the biological activity. Brini *et al*. reviewed advances that are moving CMP to time scales that match biological events such as protein folding, ligand unbinding, and some conformational changes. They also highlight the role of blind competitions in driving the field forward. New methods such as deep learning approaches are likely to make CMP an increasingly powerful tool in describing proteins in action. —VV

Science, this issue p. 1056

NEUROGENOMICS

An in vivo analysis of autism risk genes

CRISPR targeting in vivo, especially in mammals, can be difficult and time consuming when attempting to determine the effects of a single gene. However, such studies may be required to identify pathological gene variants with effects in specific cells along a developmental trajectory. To study the function of genes implicated in autism spectrum disorders (ASDs), Jin *et al*. applied a gene-editing and single-cell–sequencing system, Perturb-Seq, to knock out 35 ASD candidate genes in multiple mice embryos (see the Perspective by Treutlein and Camp). This method identified networks of gene expression in neuronal and glial cells that suggest new functions in ASD-related genes. —LMZ

Science, this issue p. 1057; see also p. 1038

CORONAVIRUS

Profiling coronaviruses

Among the coronaviruses that infect humans, four cause mild common colds, whereas three others, including the currently circulating severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), result in severe infections. Shrock *et al*. used a technology known as VirScan to probe the antibody repertoires of hundreds of coronavirus disease 2019 (COVID-19) patients and pre–COVID-19 era controls. They identified hundreds of antibody targets, including several antibody epitopes shared by the mild and severe coronaviruses and many specific to SARS-CoV-2. A machine-learning model accurately classified patients infected with SARS-CoV-2 and guided the design of an assay for rapid SARS-CoV-2 antibody detection. The study also looked at how the antibody response and viral exposure history differ in patients with diverging outcomes, which could inform the production of improved vaccine and antibody therapies. —VV

Science, this issue p. 1058

TRANSCRIPTION

Dephosphorylating RNA polymerase II

Transcription in metazoans requires coordination of multiple factors to control the progression of polymerases and the integrity of their RNA products. Zheng *et al*. identified a new dual-enzyme complex called INTAC, which is composed of protein phosphatase 2A (PP2A) core enzyme and the multisubunit RNA endonuclease Integrator. Structural and functional studies show that INTAC functions as a noncanonical PP2A holoenzyme that dephosphorylates the C-terminal domain of RNA polymerase II to attenuate transcription. This study provides a direct connection between PP2A-mediated dephosphorylation and transcriptional regulation, two fundamental cellular processes. —DJ

Science, this issue p. 1059

CELL BIOLOGY

Tension where multiple cells meet

Cells exist in varying environments and must respond to specific stimuli. During development, epithelial cells need to rapidly reorganize under tension without compromising epithelial integrity. Yu *et al*. demonstrate that *Drosophila* epithelial cells achieve this by transiently stabilizing adhesion at tricellular junctions where three cells meet (see the Perspective by Raghavan and Vasioukhin). The conserved adhesion regulator Canoe/Afadin is recruited to tricellular junctions under tension within seconds and dissociates when tension is released through a mechanism that requires Abl-dependent tyrosine phosphorylation. These results identify an in vivo mechanotransduction pathway that dynamically couples tricellular adhesion with physiological forces, allowing cells to rapidly modulate their behavior in response to mechanical changes in their environment. —BAP

Science, this issue p. 1060; see also p. 1036

FUNDAMENTAL PHYSICS

Testing physics using the hydrogen atom

Discrepancy between the proton radius determined from hydrogen and muonic hydrogen spectroscopy data, the so-called "proton radius puzzle," has been a focus of the physics community for more than a decade now. Using two-photon ultraviolet frequency comb spectroscopy below 1 kilohertz, Grinin *et al*. report a high-precision measurement of the 1S-3S transition frequency in atomic hydrogen (see the Perspective by Ubachs). Combining this measurement with the data for the 1S-2S transition, the authors obtained the Rydberg constant with improved accuracy and an independent value for the proton charge radius that favors the data from muonic hydrogen. However, the present frequency value differs from the one obtained previously using a different spectroscopic technique, leaving the puzzle still unresolved. —YS

Science, this issue p. 1061; see also p. 1033

FOREST ECOLOGY

Limits to the growing season

The length of the growing season in temperate forests has been increasing under recent climate change because of earlier leaf emergence and later leaf senescence. However, Zani *et al*.

show that this trend might be reversed as increasing photosynthetic productivity begins to drive earlier autumn leaf senescence (see the Perspective by Rollinson). Using a combination of experimental, observational, and modeling studies based on European forest trees, the researchers conclude that leaf senescence will advance by 3 to 6 days by the end of the 21st century rather than lengthening by 1 to 3 weeks as current phenological models have predicted. In turn, this predicted phenological pattern will limit the capacity of temperate forests to mitigate climate change through carbon uptake. —AMS

Science, this issue p. 1066; see also p. 1030

CHEMICAL PHYSICS

Roaming dynamics in real time

Roaming is distinct from conventional reaction channels because of the unusual geometries that chemical systems use to bypass the minimum energy pathway. It is a relatively new phenomenon that is usually determined in experiments through spectroscopic characterization of the roaming products. Using a combination of time-resolved Coulomb explosion imaging and quasiclassical trajectory analysis, Endo *et al*. report real-time observation of individual fragments of the prototypical reaction of deuterated formaldehyde (D_2CO) dissociation as they roam on ultrafast time scales. They show that roaming not only occurs several orders of magnitude earlier than previously expected but also that it can terminate in a radical (D + DCO) rather than the well-known molecular (D_2 + CO) product channel. —YS

Science, this issue p. 1072

CONSERVATION GENETICS

Survival of the most variable

As more species become highly threatened because of human activity, there has been an increasing push to understand how best to reintroduce or translocate individuals from wild or captive populations. Suggestions have varied from choosing individuals from the most environmentally similar regions to choosing those that might have the best ability to adapt to new environments. Scott *et al*. used long-term data collected during translocations of Mojave Desert tortoises, including animals formerly kept as pets, to test these questions. Although the overall rates of survival for all tortoises at the site (both reintroduced and native) were extremely low, translocated individuals with the highest heterozygosity survived at much higher rates than those determined to be similar to the target population. —SNV

Science, this issue p. 1086

CORONAVIRUS

Structure of a vaccine candidate

Much effort is being targeted at developing vaccines that will provide protection against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). A trimeric spike protein that decorates the virus is a primary target of the host immune system and the focus of vaccine development. Bangaru *et al*. present the structure of a leading vaccine candidate: a full-length spike protein with some modifications aimed at enhancing stability that is formulated in polysorbate 80 detergent. The study confirms that the full-length immunogen is in a stable prefusion conformation and provides a basis for understanding immune responses to the vaccine. —VV

Science, this issue p. 1089

CORONAVIRUS

A beneficial cocktail

Since the start of the coronavirus disease 2019 (COVID-19) pandemic, considerable effort has gone into generating and characterizing neutralizing antibodies that could be used as therapeutics. Studies in humanized mice and convalescent humans led to the development of a cocktail of two potent antibodies that simultaneously bind to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein and prevent the virus from entering host cells. Baum *et al*. evaluated the efficacy of this cocktail, REGN-COV2, in rhesus macaques, which may model mild disease, and in golden hamsters, which present more severe symptoms. The antibody cocktail provided benefits in both models when administered either prophylactically or therapeutically and is currently in clinical trials. —VV

Science, this issue p. 1110

NEUROIMMUNOLOGY

Regulatory IgA response

Multiple sclerosis (MS) is associated with the production of oligoclonal immunoglobulin G (IgG) antibodies within the central nervous system (CNS), but less is known about the status of IgA-producing cells. Pröbstel *et al*. detected clonally expanded IgA^+ cells in cerebrospinal fluid and tissue adjacent to areas of active MS-associated demyelination. Recombinant IgA antibodies assembled from the cloned antibody genes bound to surface antigens found on multiple bacterial phyla but did not cross-react with brain tissue. Recruitment of gut-associated IgA^+ B cells and plasma cells to the CNS is a potential biomarker of disease activity in MS. —IRW

Sci. Immunol. **5**, eabc7191 (2020).

ALZHEIMER'S DISEASE

Understanding protein trafficking defects

Alzheimer's disease (AD) is characterized by abnormal protein accumulation in the extracellular milieu in the brain. The endosomal pathway is responsible for proteolytic cleavage and secretion of proteins, and recent data have shown that a defective endosomal pathway contributes to AD pathogenesis. However, which proteins are affected by defective endosomal pathways is still unclear. Simoes *et al*. performed proteomic analysis in cerebrospinal fluid from mice with defective endosomal trafficking and in patients with AD and showed that the expression of two transmembrane proteins, APLP1 and CHL, was increased and correlated with tau expression. The results provide valuable insights regarding the effect of endosomal trafficking alterations in AD. —MM

Sci. Transl. Med. **12**, eaba6334 (2020).

WATER RESOURCES

Missing freshwater found off Hawai'i

Freshwater resources are limited on oceanic islands, and their scarcity complicates sustainability efforts. Attias *et al*. identified an important new freshwater resource in the submarine environment offshore of the island of Hawai'i. Previous studies suggested that as much as 40% of water input to the onshore aquifer is missing and unavailable for use. The salinity of seawater makes it more electrically conductive than freshwater, and electromagnetic imaging surveys off the coast of Hawai'i revealed large quantities of freshwater in deep submarine sediments just offshore. These alternative water resources improve the prospects of sustainable water extraction for Hawai'i, and the authors speculate that the same mechanism may operate on other volcanic islands. —KVH

Sci. Adv. 10.1126/sciadv.abd4866 (2020).

REVIEW SUMMARY

PROTEIN MODELING

Protein storytelling through physics

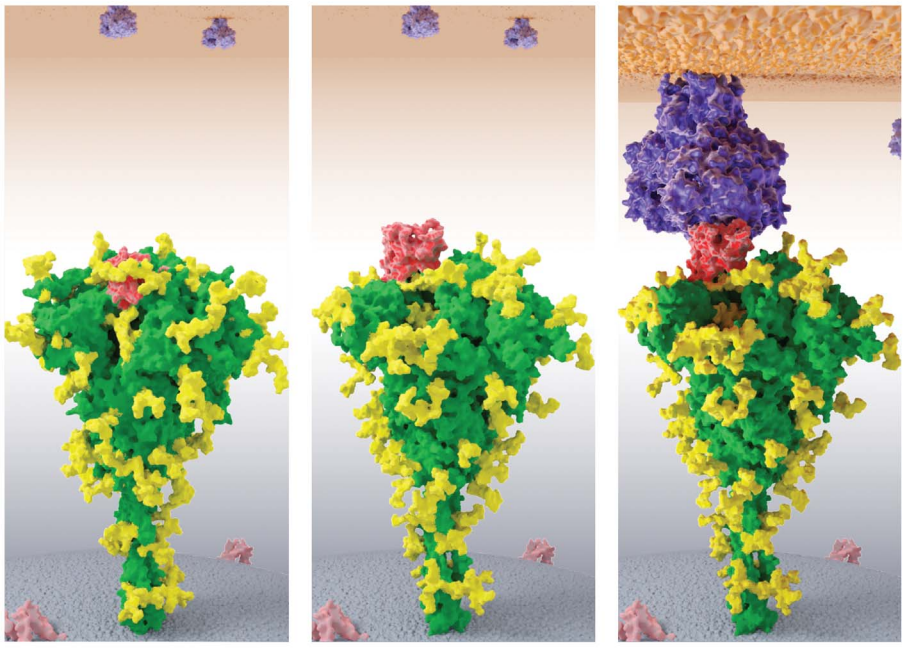
Emiliano Brini, Carlos Simmerling, Ken Dill*

BACKGROUND: Understanding biology, particularly at the level of actionable drug discovery, is often a matter of developing accurate stories about how proteins work. This requires understanding the physics of the system, and physics-based computer modeling is a prime tool for that. However, the computational molecular physics (CMP) of proteins has previously been much too expensive and slow. A large fraction of public supercomputing resources worldwide is currently running CMP simulations of biologically relevant systems. We review here the history and status of this large and diverse scientific enterprise. Among other things, protein modeling has driven major computer hardware advances, such as IBM's Blue Gene and DE Shaw's Anton computers. Further, protein modeling has advanced rapidly over 50 years, even slightly faster than Moore's law. We also review an interesting scientific social construct that has arisen around protein modeling: community-wide blind competitions. They have transformed how we test, validate, and improve our computational models of proteins.

ADVANCES: For 50 years, two approaches to computer modeling have been mainstays for developing stories about protein molecules and their biological actions. (i) Inferences from structure-property relations: Based on the principle that a protein's action depends on its shape, it is possible to use databases of known proteins to learn about unknown proteins. (ii) Computational molecular physics uses force fields of atom-atom interactions, sampled by molecular dynamics (MD), to develop biological action stories that satisfy principles of chemistry and thermodynamics. CMP has traditionally been computationally costly, limited to studying only simple actions of small proteins. But CMP has recently advanced enormously. (i) Force fields and their corresponding solvent models are now sufficiently accurate at capturing the molecular interactions, and conformational searching and sampling methods are sufficiently fast, that CMP is able to model, fairly accurately, protein actions on time scales longer than microseconds, and sometimes milliseconds. So, we are now accessing important biological events, such as protein folding, unbinding, allosteric change, and assembly. (ii) Just as car races do for auto manufacturers, communal blind tests such as protein structure-prediction events are giving protein modelers a shared evaluation venue for improving our methods. CMP methods are now competing and often doing quite well. (iii) New methods are harnessing external information—like experimental structural data—to accelerate CMP, notably, while preserving proper physics.

What are we learning? For one thing, a long-standing hypothesis is that proteins fold by multiple different microscopic routes, a story that is too granular to learn from experiments alone. CMP recently affirmed this principle while giving accurate and testable microscopic details, protein by protein. In addition, CMP is now contributing to physico-chemical drug design. Structure-based methods of drug discovery have long been able to discern what small-molecule drug candidates might bind to a given target protein and where on the protein they might bind. However, such methods don't reveal some all-important physical properties needed for drug discovery campaigns—the affinities and the on- and off-rates of the ligand binding to the protein. CMP is beginning to compute these properties accurately. A third example is shown in the figure. It shows the spike protein of severe acute respiratory syndrome coronavirus 2(SARS-CoV-2), the causative agent of today's coronavirus disease 2019 (COVID-19) pandemic. A large, hinge-like movement of this sizable protein is the critical action needed for the virus to enter and infect the human cell. The only way to see the details of this motion—to attempt to block it with drugs—is by CMP. The figure shows CMP simulation results of three dynamical states of this motion.

OUTLOOK: A cell's behavior is due to the actions of its thousands of different proteins. Every protein has its own story to tell. CMP is a granular and principled tool that is able to discover those stories. CMP is now being tested and improved through blind communal validations. It is attacking ever larger proteins, exploring increasingly bigger and slower motions, and with ever more accurate physics. We are reaching a physical understanding of biology at the microscopic level as CMP reveals causations and forces, step-by-step actions in space and time, conformational distributions along the way, and important physical quantities such as free energies, rates, and equilibrium constants. ■



CMP modeling of COVID-19 infecting the human cell. SARS-CoV-2 spike glycoprotein (green, with its glycan shield in yellow) attaching to the human angiotensin-converting enzyme 2 (ACE2) receptor protein (purple) through its spike receptor-binding domain (red). (Left) The receptor binding domain (RBD) is hidden. (Middle) The RBD is open and accessible. (Right) The RBD binds human ACE2 receptor. This is followed by a cascade of larger conformational changes in the spike protein, leading to viral fusion to the human host cell.

CREDIT: LUCY FALLON

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dill@laufercenter.org
Cite this article as E. Brini *et al.*, *Science* **370**, eaaz3041 (2020). DOI: 10.1126/science.aaz3041

REVIEW

PROTEIN MODELING

Protein storytelling through physics

Emiliano Brini[1], Carlos Simmerling[1,2], Ken Dill[1,2,3]*

Every protein has a story—how it folds, what it binds, its biological actions, and how it misbehaves in aging or disease. Stories are often inferred from a protein's shape (i.e., its structure). But increasingly, stories are told using computational molecular physics (CMP). CMP is rooted in the principled physics of driving forces and reveals granular detail of conformational populations in space and time. Recent advances are accessing longer time scales, larger actions, and blind testing, enabling more of biology's stories to be told in the language of atomistic physics.

Science…is the organized systematic enterprise that gathers knowledge about the world and condenses [it] into testable laws and principles.

—E. O. Wilson, *Consilience*

Scientists are storytellers. Biologists tell stories about biomolecules and their actions in the cell. Even the simplest cell has thousands of types of proteins. Based on the premise that structure determines function—the same principle that helps you discern how a knife and fork work—protein shapes are often the starting points for protein storytelling. Like a necklace of beads, a protein is a molecule of amino acids (the beads) chained together. The 20 different types of amino acids make up an alphabet for stringing the beads into different sequences that fold into different native structures (the compact shapes that proteins adopt in the cell). These shapes are known at atomic detail for more than 150,000 proteins and available in a public resource called the Protein DataBank (PDB) (*1*), thanks to x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo–electron microscopy (cryo-EM) experiments from a large community of structural biologists over the past 60 years.

Computer modeling plays a big role in molecular storytelling. For one thing, two proteins having similar sequences often have similar shapes and perform similar actions. After a protein folds, it can go to work in the cell through binding to other molecules, or its motions, or working together in complex assemblies. Some computer algorithms are designed to get insights about poorly understood proteins by looking at proteins with known structures. For another thing, computing can leverage the laws of physics for understanding the motions and actions of biomolecules. The former approach is called structural bioinformatics (SBI) and the latter, computational molecular physics (CMP). Today's computational modeling is often a combination of both. For protein storytelling, what's usually asked of SBI is to reveal a native structure. But the native structure is a single static snapshot, often insufficient to tell the story of how the protein works. What's missing in SBI is the physics—the driving forces, the causes and effects, the intermediate steps—the how's, not just the what's. What's asked of CMP is to reveal the forces, motions, binding, and actions, in addition to structures. Atomistic physical modeling with force fields, often sampled by molecular dynamics (MD), draws upon our knowledge of covalent and noncovalent bonding in molecules, how those interactions are affected by the solvent water molecules, the considerable role of entropies, and how it all manifests in the complex environment inside the protein.

CMP helps to write the stories of biomolecules

Molecular physics aims to accomplish the following: (i) Reveal causalities, by allowing investigations of the driving forces, dynamics, and motions. (ii) Give the intermediate steps in time and space, angstrom by angstrom, nanosecond by nanosecond—narratives of molecular actions such as folding, binding, and rearrangements. (iii) Give conformational distributions, not just a single averaged structure. Proteins writhe and deform in ways that are needed for matching our stories to experiments. Drugs often don't bind "key-in-lock" to preexisting protein cavities; they push the protein into a new shape that we seek to learn. (iv) Systematize, giving a common language for our storytelling. In principle, physical modeling is transferable: to different proteins, different ligands, different binding situations, or different experimental conditions, including those not yet measured. And physics-based models are less susceptible to errors of so-called hasty generalization (i.e., incorrect inferences from too little data), so their parameters can be systematically improved. (v) Go beyond limitations of databases. Physical modeling can be applied to classes of biomolecules that are sparsely represented in databases, such as membrane proteins, or molecules that are intrinsically disordered, large, or complexed. These are not well populated in databases because experiments are difficult. But they obey the same laws of physics.

CMP needs large computations. In practice, it means sacrificing some accuracy in the physical model, or limiting the time scale simulated to those shorter than the biological actions of interest. To reach for the deepest truths, we need to root our stories in our deepest understanding of nature. And though CMP modeling entails simplifications and approximations, it nevertheless provides today's best achievable description of the underlying physics consistent with achievable computational costs.

CMP uses physical potentials and Boltzmann sampling

The properties of molecules are governed by quantum mechanics (QM). But QM computations are much too costly for large, flexible systems like proteins in water. So, in biomolecular modeling, true forces are approximated using force fields. Such models treat molecules as having preferred bond lengths and angles and describe other interatomic interactions as a combination of van der Waals and Coulomb forces, including the impact of solvation (*2*)—typically in water—and ultimately of proteins' preferential conformations. Molecular force fields originated in the 1960s and '70s (*3*–*13*); they were pioneered by Allinger, Lifson, Kollman, and Scheraga, in addition to Levitt, Warshel, and Karplus, whose achievements were recognized by the 2013 Nobel Prize in Chemistry. They continue to undergo systematic improvement by a large community (*14*–*20*).

But knowing accurate physical potentials is only part of the problem. To generate faithful protein stories, modeling must respect other physics: the nature of movement and change and the laws of thermodynamics. Dynamical processes must follow Newton's laws of motion. Equilibrium modeling must give the Boltzmann distribution of conformational populations—and thus free energies, which are where models meet experiments. The latter can only be fulfilled by using special sampling approaches, such as Monte Carlo (MC) or MD, sometimes accelerated by enhanced sampling (*21*). MD computations are expensive because they require femtosecond timesteps, to avoid violating Newton's laws (*22*). And even though such methods have now become relatively efficient, they still entail compromises, and the stories that are most faithful to nature are of the simplest proteins, the shortest time scales, or the smallest actions.

[1]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA. [2]Department of Chemistry, Stony Brook University, Stony Brook, NY 11794, USA. [3]Department of Physics and Astronomy, Stony Brook University, Stony Brook, New NY 11794, USA.
*Corresponding author. Email: dill@laufercenter.org

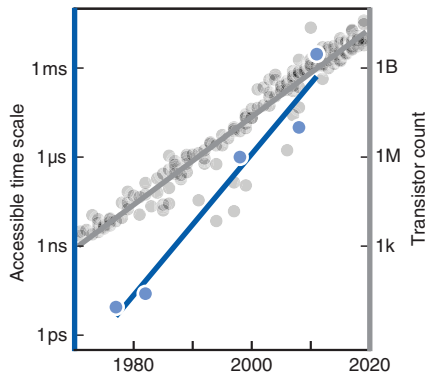


Fig. 1. Molecular simulations have improved faster than Moore's law. Blue: MD simulations have accessed exponentially longer time scales of molecular motions over the past 50 years. Gray: For reference, Moore's law of increased densities of transistors on microchips (*136*).

A key consideration is always speed versus accuracy. To gain speed, you can use coarse-grained models. Coarse-grained modeling lumps together atoms into larger rigid units (*23*–*26*). This approach is useful when you know in advance which degrees of freedom are relevant to the problem at hand. Conversely, some questions, involving mechanisms of enzyme reactions or spectroscopic observables, for example, require quantum-mechanical details, including the electrons on all or part of the protein. Including such details increases the cost of the calculation and is tractable only when protein movements are small. Although important advances have been made with both of these approaches, atomistic CMP modeling remains a popular compromise between speed and accuracy for much of biology and drug discovery and is the focus of this review.

CMP has been both a driver and beneficiary of many advances

As computer power has grown exponentially, so has CMP modeling power (Fig. 1). The first stories told were of how proteins fold and how they bind small molecules, both of which have relevance to drug discovery. The earliest computational physics of proteins, in the 1960s and '70s (*27*), was done largely on central mainframes. Then came labclusters in the '90s and supercomputing. Among the first protein-folding stories, in 1998, was one from van Gunsteren and co-workers on peptides (*28*) and one from Duan and Kollman and their colleagues, who applied supercomputing to attempt to fold the villin headpiece, one of the smallest foldable proteins (*29*).

Protein modeling is now a major activity of public supercomputers. The demand for better CMP modeling of biomolecules has driven advances in high-performance computing, including (i) the IBM Blue Gene computer (*30*); (ii) Folding@Home, a distributed grid network developed by Pande and Shirts at Stanford (*31*); and (iii) D. E. Shaw and colleagues' special purpose Anton and Anton 2 supercomputer (*32*, *33*).

CMP modeling has advanced at Moore's law rates

Advances have come not only from computer hardware. CMP has also advanced from systematic improvements in atomistic force fields, solvation, sampling, and workflows. The first MD simulations, in 1977, sampled motions of a small protein representing a real time of only about 3 ps and necessarily left out an essential component, namely, the water solvent (*27*). Today's simulations are run with quality force fields and models of surrounding solvent or membrane, and over much longer time scales, through developments described below. An important milestone, which was reached about a decade ago, was simulating time scales of milliseconds and longer. This has been transformative because these are the time scales on which large-scale, biologically interesting motions appear.

Enhanced sampling methods help to tell bigger stories

Today's storytelling is often limited to small proteins, simple actions, or short time scales. To tell bigger stories, we need faster and more efficient conformational searching and sampling. Advances are coming from enhanced sampling methods, such as metadynamics (*34*), replica exchange molecular dynamics (REMD) (*35*), simulated annealing (*36*), adaptive force biasing (*37*), and umbrella sampling (*38*) [for reviews, see (*39*, *40*), and for the basic concepts see (*21*, *41*)]. Also important has been systematic improvement in measuring and controlling errors, such as through the multi-state Bennett acceptance ratio estimator (MBAR) (*42*) and the weighted histogram analysis method (WHAM) (*43*).

Protein structure modeling can benefit from harnessing external information, an approach called integrative modeling (*44*). Computer algorithms can be accelerated by leveraging additional SBI insights about protein structures. For example, Rosetta (*45*), Quark (*46*), Dock 1 (*47*), Cluspro (*48*), Haddock (*49*), MODELLER (*50*), and IMP (*51*) do this to outstanding effect for certain structure-based computations. However, methods for leveraging external information often require sacrificing the ability to give Boltzmann populations, free energies, or dynamics.

An alternative integrative approach is MELD (modeling employing limited data), an accelerator of MD that also retains the advantages of the physics (*52*, *53*). MELD-accelerated MD (MELD x MD) melds external information or directives about the end state of interest with CMP, in a way that preserves Boltzmann populations. It uses Bayesian inference to leverage noisy, combinatoric, corrupted, sparse, or ambiguous information, to accelerate MD, often by orders of magnitude. It adds value in protein structure prediction from sequences (*52*, *54*, *55*), native structure determination in conjunction with problematic experimental data (*53*, *56*), ligand binding to proteins (*57*–*59*), and finding dynamical routes of conformational change processes (*60*).

CMP can tell dynamical stories

Some protein stories are about dynamics and motions. In those cases, we seek the sequences of events (i.e., the pathways or routes) and their speeds (i.e., the transition rates). Pathways can be studied in CMP by using long simulations of individual trajectories (*61*). But long trajectories are expensive to compute and hard to converge, owing to many stochastic meanderings. Often, the story line is better obtained by dividing the possible routes into multiple short trajectories that can be computed in parallel, each of which traverses fewer states. Increasingly, studies of processes are done by stitching together many short simulations in parallel.

One strategy for computational parallelization is Markov state modeling (*62*–*68*). Markov state models (MSMs) use short trajectories in parallel to efficiently compute the rates and routes of protein motions and actions (*69*), protein-folding pathways (*70*), motions and dynamics (*71*), and ligand binding (*72*) and unbinding (*73*). MSMs are best when a pathway has a few dominant routes ("superhighways" in conformational space).

Another parallelization strategy is milestoning, which is best when a process has a dispersed diffusional route structure (*74*). In milestoning, a path along a predefined reaction coordinate is metaphorically "fenced off into different time zones," called milestones. Each milestone boundary is a starting point for spawning independent parallel trajectories to the next milestone. First-passage times are computed, then combined, to give the time evolution of the whole process. Milestoning is effective in simulating binding (*75*), membrane dynamics (*76*), and transitions between protein conformations (*77*). Milestoning relies on the assumptions that the transition events between successive milestones and the time lags between these transitions are statistically independent.

The weighted ensemble method (*78*) also fences off conformational space into zones but then spawns new daughter trajectories whenever a trajectory hits a fence that leads to an "interesting" area of the phase space. This enhances the sampling of interesting events while providing a rigorous nonequilibrium

reweighting scheme to learn the unbiased time evolution of the trajectory (*79*).

Blind competitions are validating our storytelling tools

How do we know if our stories are true? This question applies to both types of stories—those from structural biology and those from CMP. It has long been difficult to model the native structures—and tell the stories—of most proteins. It still is. But the field was advanced enormously in 1994 when John Moult introduced a new type of community-wide blind-test event (*80*). In CASP (critical assessment of protein structure prediction), a core team of assessors releases various amino acid sequences onto a website. Modelers then have a short, fixed time to submit their predictions of that sequence's three-dimensional (3D) structure. Subsequently, the true experimentally determined 3D structures are released. The assessors then compare and evaluate all the predictions. This event has brought considerable value. Blind testing helps eliminate our biases, accelerate our learning, and enhance our communal interests.

Until recently, CMP modeling has been too slow to compete, so CASP has given insights mostly about SBI modeling. Because SBI methods draw inferences from other known protein structures, SBI predictors are most successful when a closely related PDB template protein can be found. And SBI successes at CASP have benefited from (i) rapid growth of the PDB database; (ii) algorithms for improved alignments of the target with the PDB template protein (such as PSIBLAST (*81*)); (iii) using protein fragments rather than individual amino acids as units of structure; (iv) using information fromsequence coevolution (*82*) to identify native contacts; and recently, (v) deep learning (DL) modeling; see Fig. 2.

CASP has spawned other protein computational competitions. Among the newest is the EM validation challenge, started in 2019, which tests protein structures computed from limited EM data (*83*). Other events concern binding actions [CAPRI (Critical Assessment of Prediction of Interactions], which began in 2001, addresses multiple proteins binding to each other. SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands), initiated in 2008 (*84*), is about predicting solvation free energies, binding affinities, host-guest conformational sampling, and pK_a's (acid dissociation constants) of small molecules. D3R (Drug Design Data Resource), first run in 2015, evaluates small-molecule and drug-like ligands bound to proteins (*85*).] And still other events focus on the functions of proteins [CAFA (Critical Assessment of protein Function Annotation), initiated in 2013, assesses protein function predicted from amino acid sequences (*86*) and CAGI (Critical Assessment of Genome Interpretation), now in its fifth iteration, assesses effects of mutations on the stabilities and functions of proteins (*87*).] DREAM (Dialogue on Reverse Engineering Assessment and Methods), established in 2006, is crowdsourcing competitions in computational biomedicine for disease and drug discovery (*88*). These communal blind events have advanced our understanding of methodology through systematic comparative evaluations. They provide benchmarks of the year-to-year progress in computational protein modeling, illuminating the best methods and indicating where improvements are needed.

CMP is now becoming competitive in blind communal tests

Over the past half-decade, physics-based methods have become fast enough to enter the protein modeling competitions (*89*, *90*). Physical methods predict more than is tested by these competitions, but these events provide quantitative touchstones that any storytelling tool must get right. If our stories don't have the correct ending (native structure), how can we trust the rest of the story? And, whereas blind competitions help advance CMP, conversely, CMP adds value to protein structure prediction. CMP can circumvent an Achilles heel of SBI, namely, the need for structure databases, homology modeling, and sequence alignments. CMP can tackle proteins that don't resemble known proteins. Some successes from the 2018 events are reported below.

Predicting native structures, assisted by data

Experimental structural biology provides data, but computations are needed to leverage that data to give an atomically detailed structure. CASP's structure refinement category asks: Are approximately correct structures improvable? Early CASP results showed that trying to improve one SBI algorithm by another SBI method was mostly unsuccessful (*91*). However, recently, Feig and others have shown good success in improving SBI predictions by CMP approaches; see Fig. 3A (i) (*92*). In one CASP event category, predictors are given sparse NMR data from nuclear Overhauser effect spectroscopy and dipolar couplings. On its own, these data are not sufficient to determine a native structure. CASP asks whether computations can augment this data to give correct structures. The best predictor in this CASP category in 2018 was a CMP method, MELD x MD; see Fig. 3A (ii) (*56*, *93*). The 2019 first EM validation challenge provided data from cryo-EM for predicting the structure of a protein at different EM map resolutions, and the structure of a protein-ligand complex at a single resolution (*94*). By several metrics, including root mean square deviation (RMSD), all predictors gave fairly accurate structures. But a CMP method (MELD x MDFF) also gave the best-fitting map, at a resolution of 2.3 Å; see Fig. 3A (iii). Structure 3AJO in Fig. 3B shows good agreement between the predicted (blue) and true (orange) native structure.

Predicting a native structure from sequence alone

More challenging than predicting a structure by using ancillary data is CASP's T0 category, for predicting structure from the amino acid sequence alone. For a few small, simple proteins, CMP methods succeed at these ab initio folding tests, from amino acid sequence alone (Fig. 3B). Each image in that figure gives its CASP target ID number (Txxx) and the RMSD error between the predicted (blue) and true (orange) native structure. For protein target T0958, a CMP model was the best prediction in CASP among all submitted (black triangle, Fig. 2). Although CMP modeling in CASP does not require externally supplied structural knowledge, the result in Fig. 3B was accelerated by seeding SBI server–predicted structures as initial structures for the simulation.

Predicting the tightness of binding

In the 2018 D3R grand challenge 4, predictors were given a target protein, cathepsin S, and 39 ligands that bind to it, spanning three orders of magnitude in binding affinity. The challenge was to compute the binding affinities (*95*). In this test, CMP free-energy calculations

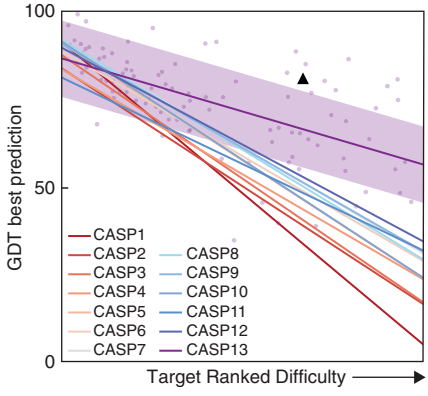


Fig. 2. CASP success rates versus difficulty of the target protein, over the years. "Difficulty" is defined by how similar a template sequence that can be found in the PDB is to the target protein being predicted. Data points show the best predictions for each target protein in CASP 13; the purple shading shows the variance. Lines show the mean for each CASP. Main conclusions: (i) SBI requires good templates (all lines slope down). (ii) Predictions are improving over time (lines are higher in later events). (iii) Coevolutionary data and deep learning are adding value (step from CASP 12 to 13). (iv) CMP is now competitive in CASP (black triangle).

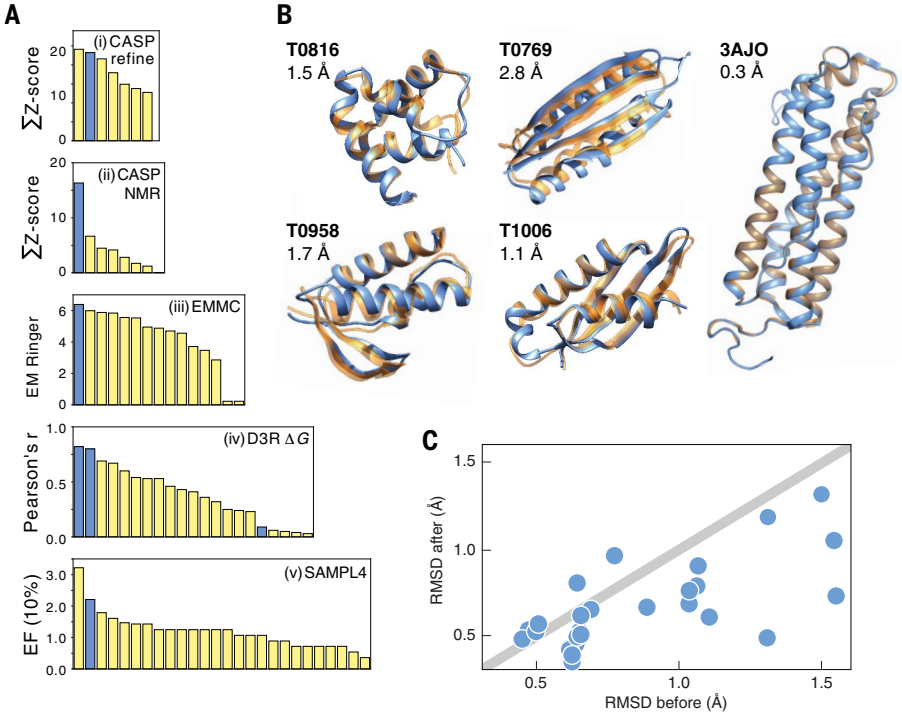


Fig. 3. Successes of CMP in blind communal events. (**A**) Comparisons with other predictions, in five events. Blue bars, CMP predictions; yellow bars, other. Left to right: Most- to least-successful predictions, based on each *y* axis metric. From top to bottom: CASP 13 structure refinement, CASP 13 leveraging NMR data to determine protein structures, Cryo-Electron Microscopy Model Challenge (2019 EMMC), D_3R binding affinity of 300 ligands to cathepsin S, SAMPL 4 virtual screening of ligands to HIV integrase. (**B**) Predicted versus true structures from CMP in CASP 13 ab initio folding and 2019 EMMC cryo-EM refinement. (**C**) CMP-based protein-protein docking structures, in the D3R Grand Challenge 4 stage 2b event (*59*), after a rigid-body ClusPro first step. *y* axis, RMSD error of MELD + ClusPro refined protein docked structures; *x* axis, RMSD error of ClusPro prediction alone. Points below the line indicate successful refinements.

outperformed all other methods (including DL) to achieve an RMSE (root mean square energy error) of 0.49 kcal/mol over all ligands; see Fig. 3A (iv) (*96*). A second test, of protein BACE1 bound to a challenging set of ligands, which have multiple scaffolds and high chemical and structural diversity, was not successful (*95*). However, CMP was partially successful in finding correct docking orientations; see Fig. 3C (*59*). A 2013 SAMPL 4 challenge sought small-molecule inhibitors of the HIV integrase catalytic core domain (*97*). CMP performed significantly better than control and null models (*98*); see Fig. 3A (v).

Predicting the shape fitting of two proteins binding together

It is currently challenging to predict the structure and relative orientations of two proteins that are bound to each other, starting from knowledge of only their amino acid sequences. If successful, it would pave the way to atomistic modeling of whole biochemical pathways, the elementary units of which are pairs of interacting proteins. Today's successes require much more input knowledge. If the conformation of the bound form of both proteins are already known, then existing algorithms can often find the right docked structure by rigid body rotations of one relative to the other (*99*–*101*). A step in the direction of the grander challenge was protein T121 in CAPRI round 38: seeking the binding conformation of a protein of known shape to a peptide that is unstructured prior to binding. The challenge is to find the induced structure of the peptide in its bound state. In this case, combined with a rigid-body front-end step, MELD x MD gives a medium-quality model (*58*, *101*).

Predicting conformational populations, foretelling when structures are right

Here is a common problem in predictions: An algorithm outputs several plausible native structures, but we don't know which one is right. SBI methods don't provide a principle for choosing and have traditionally not been able to tell (*89*). The physically principled way to choose is to know the free energies (i.e., the relative populations). CMP methods can compute these. Indeed, as MD modeling has recently entered CASP, its predictions of relative populations have been shown to correctly foretell when its methods have found the true native structure (*52*–*54*).

MD can predict structures without assistance from homology models

For some proteins, there are no templates in the PDB that are good starting points for bioinformatics predictions of native structures. Because CMP methods make predictions from physical principles and don't require starting structural knowledge, they can often predict protein structures that database methods cannot (*55*). At the moment, however, the greatest power is still achieved by combining CMP with external structural information (*52*–*54*, *56*). CMP plus information is still CMP, provided it satisfies the physics of Newton and Boltzmann.

In short, CMP methods have now reached two key mileposts. (i) They are computationally fast enough to enter communal blind competitions, and (ii) they add value in some cases. However, CMP is still computationally expensive; does not yet have turn-key reliability across challenges; is best when some constraints are known; and is limited to small, simple proteins and ligands. But these are early days. Below are some stories being told; see also (*21*, *102*).

A few of the stories being told by CMP

CMP modeling has several roles. It helps to turn NMR, x-ray, and cryo-EM data into protein structures; to compute binding affinities that can be actionable in discovering new drugs; and to establish detailed principled narratives about biomolecule behaviors and their actions in the cell. Below are a few examples.

Resolving the paradox of protein-folding route heterogeneity

A long-standing puzzle entails the routes of protein folding. Is there a general kinetic principle by which all proteins fold from their unfolded state to each one's unique native state? The issue has hinged on the meaning of the word "pathway." For experimentalists, a pathway is often defined on the experimental macroscale as a single predominant sequence of events. For theorists, a pathway is often defined by the simulational microscale watching the individual wiggling of each trajectory. State-of-the-art MD simulations over the past decade have shown that individual molecules reflect the heterogeneity expected from statistical mechanical theories, yet with average behaviors that reflect the routes that experimentalists see (*103*–*105*), resolving such paradoxes. An important side consequence of modeling pathways is the demonstration that the molecular physics alone can accurately predict the pathway end states, namely, the native structures of (small, simple) proteins,

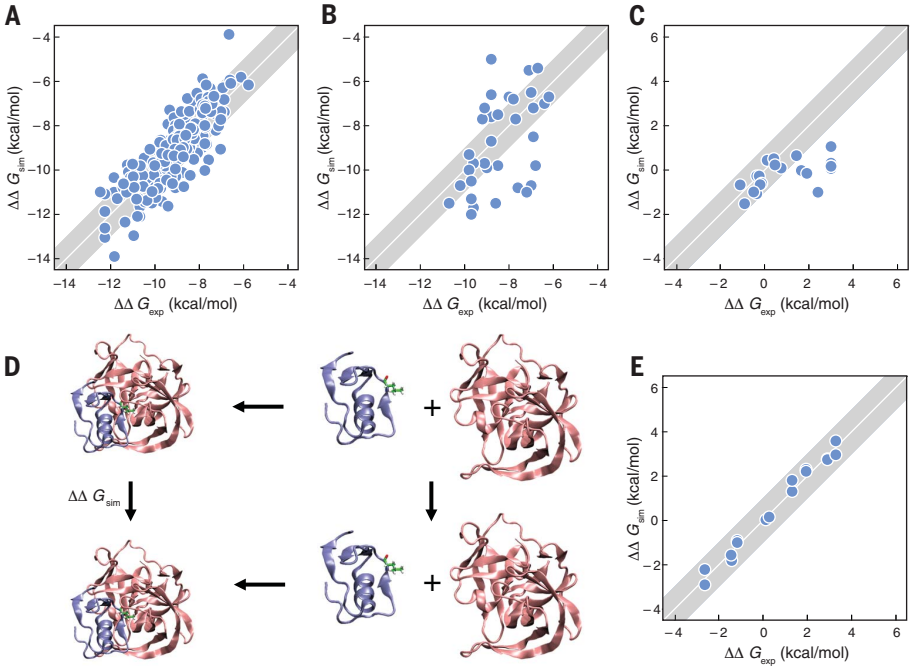


Fig. 4. CMP can predict relatively accurate experimental binding affinities, across multiple ligands and proteins. The diagonal line represents perfect agreement with experiments. The shaded area indicate 1 kcal-mol^{-1} error bars. (**A**) Schrodinger free-energy perturbation calculations of 200 ligands in eight proteins (*108*). (**B**) Algedhi *et al*. prediction of binding affinities of ligands across protein families (*111*). (**C**) MELD x MD relative binding affinities of various P53 mutant peptides, which are highly flexible, to the MDM2/x protein (*57*). (**D**) The thermodynamic cycle of Zou *et al*. for computing mutational effects on protein-protein binding (*112*). (**E**) Predicted affinities versus experiments for (D) (*112*).

without the need for homology modeling of other reference proteins (*55*, *104*, *105*).

Computing drug–protein affinities

A major objective has been to compute how tightly a chemical (natural ligand or drug, for example) can bind to a protein. So-called docking methods can often estimate a pose, what protein crevice the molecule might fit, and how it orients inside it. But more crucial for understanding biological mechanisms and protein stories is how tightly it binds. This is becoming increasingly possible through CMP. Here are some of the advances (*106*–*110*): (i) A test of small ligands in different proteins. In 2015, the Schrodinger group modeled 200 ligands in eight proteins (*108*), giving the results in Fig. 4A. (ii) Computing selectivities. Aldeghi *et al*. computed the selectivities of ligand binding across a family of related proteins (*111*); see Fig 4B. (iii) Complex flexible ligands. Morrone *et al*. (*57*) computed the relative binding affinities of highly flexible peptide ligands (P53 mutants) to MDM2 and MDMx proteins; see Fig. 4C. (iv) Mutations in protein-protein binding. Free energies of binding one protein to another can change upon mutation. Some are now accurately captured by CMP (*112*); see Fig. 4E. The average errors are very small, 0.27 kcal/mol. Similarly small average errors are found for CMP predictions of mutational effects on protein stabilities (*113*). The high accuracy likely reflects the relative maturity of today's force fields designed to handle proteins. Force fields for more diverse molecules, such as small-molecule drugs and ligands, are not as accurate, but communal efforts (*114*) are improving them. (v) Thermodynamic components. MD simulations are now also giving the thermodynamic components, enthalpy and entropy, of binding, which give additional insights for storytelling about molecular recognition (*115*, *116*).

Mechanisms of ligand dissociation and off-rate estimation

There is interest in knowing not just the affinity of a ligand molecule for a protein, but also its off-rate. The off-rate (k_{off}) is a measure of the residence time that the ligand spends in the site and is sometimes a better predictor of biological efficacy than is the equilibrium binding affinity ($K = k_{on}/k_{off}$) (*117*). Even though some off-rates are known to be as slow as minutes, advanced-sampling MD simulations give good agreement with experiments (*118*, *119*), including over these slow time scales (*120*). MD simulations identified the mechanisms underlying slow off-rates in a tuberculosis drug target, guiding the rational design of new ligands with even longer residence times. Subsequent synthesis, crystallography, and kinetic assays have confirmed the predictions (*121*, *122*).

Mechanisms of amyloid aggregation

No experiment is yet granular enough to show the molecular events as protein chains come together to form the amyloid fibers that are relevant to diseases of neurodegeneration. MD simulations are elucidating the kinetic steps (*123*–*126*) and giving insights into new drug discovery approaches (*127*).

Space here does not allow us to review the many other important emerging stories, about how transducing proteins convert the chemical energy of adenosine 5´-triphosphate hydrolysis into mechanical work and forces, how allosteric "action-at-a-distance" conformational changes transmit chemical signals across large molecules, how kinase proteins send signals in cancer cells, and how molecules cause biological motions by "walking" along tracks made of other proteins.

What lies ahead?

How can we model bigger proteins, bigger motions, and longer time scales?

We need continued improvements in force fields and solvent models. We need need faster and more targeted searching and sampling, to reduce the combinatorial nightmare of conformational space. One innovative idea is to estimate entropies by using ideas based on computer compression algorithms (*128*). One new accelerator of molecular dynamics is Boltzmann generators, which use deep learning methods to learn where to find the deepest free-energy wells on landscapes (*129*) and improve sampling (*130*). And quantum computing might ultimately help us tackle these big stochastic optimization challenges. But we also need more than better hardware alone, because the scaling to larger proteins is a bigger problem than even continued advances based on Moore's law can solve.

The role of deep learning

DL can bring insights into protein structures that can complement the CMP's insights into the physics. In December 2018, the CASP evaluators reported notable success from AlphaFold, a new DL method from Google's London-based DeepMind group and other groups (*131*). The superior ability of DL methods (*131*, *132*) in harnessing data relative to earlier statistical methods led to the best results in CASP's free modeling category (*90*). The advance from DL was about twofold that of the average natural biennial advance in CASP. DL is powerful when given either large databases to learn from, or when some rules—such as in games—can be used to generate input. AlphaFold learned from the protein structures in the

PDB. But DL does not know what it cannot see. In drug discovery, what's most needed is to generate diversity, i.e., new classes of molecules that are not already known drugs (*133*, *134*). Protein storytelling can benefit from both DL for insights from knowledge bases and from CMP for causation and driving forces, motions, and binding, where structures are heterogeneous, and where we need to understand the role of the protein's external environment.

Summary

CMP is an increasingly powerful tool for telling the stories of protein molecule actions. Systematic improvements in force fields, enhanced sampling methods, and accelerators have enabled CMP to reach time scales of important biological actions and opened the door to evaluation in communal blind events, like CASP. These time-scale successes were forecast a quarter-century ago (*135*). At this rate, in the next quarter-century, we'll be telling stories of protein molecules over the whole life span—tens of minutes—of a bacterial cell; perhaps the updated version of this article will be called "Cellular storytelling through physics." CMP is increasingly grounding our narrative stories of the biological actions of proteins in principled physics.

REFERENCES AND NOTES

1. H. M. Berman *et al*., The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899–907 (2002). doi: 10.1107/S0907444902003451; pmid: 12037327
2. E. Brini *et al*., How Water's Properties Are Encoded in Its Molecular Structure and Energies. *Chem. Rev.* **117**, 12385–12414 (2017). doi: 10.1021/acs.chemrev.7b00259; pmid: 28949513
3. R. A. Scott, H. A. Scheraga, Conformational Analysis of Macromolecules. III. Helical Structures of Polyglycine and Poly-L-Alanine. *J. Chem. Phys.* **45**, 2091–2101 (1966). doi: 10.1063/1.1727894
4. D. A. Brant, W. G. Miller, P. J. Flory, Conformational energy estimates for statistically coiling polypeptide chains. *J. Mol. Biol.* **23**, 47–65 (1967). doi: 10.1016/S0022-2836(67)80066-4
5. M. Levitt, S. Lifson, Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **46**, 269–279 (1969). doi: 10.1016/0022-2836(69)90421-5; pmid: 5360040
6. A. T. Hagler, E. Huler, S. Lifson, Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **96**, 5319–5327 (1974). doi: 10.1021/ja00824a004; pmid: 4851860
7. F. Momany, R. F. McGuire, A. Burgess, H. A. Scheraga, Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* **79**, 2361–2381 (1975). doi: 10.1021/j100589a006
8. B. R. Brooks *et al*., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983). doi: 10.1002/jcc.540040211
9. S. J. Weiner *et al*., A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984). doi: 10.1021/ja00315a051
10. W. F. van Gunsteren, H. J. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual* (Biomos, Groningen, Netherlands, 1987), pp. 1–221.
11. W. L. Jorgensen, J. Tirado-Rives, The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988). doi: 10.1021/ja00214a001; pmid: 27557051
12. W. D. Cornell *et al*., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995). doi: 10.1021/ja00124a002
13. D. A. Case *et al*., The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005). doi: 10.1002/jcc.20290; pmid: 16200636
14. C. Tian *et al*., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **16**, 528–552 (2020). doi: 10.1021/acs.jctc.9b00591; pmid: 31714766
15. K. T. Debiec *et al*., Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **12**, 3926–3947 (2016). doi: 10.1021/acs.jctc.6b00567; pmid: 27399642
16. L.-P. Wang *et al*., Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **121**, 4023–4039 (2017). doi: 10.1021/acs.jpcb.7b02320; pmid: 28306259
17. J. Huang *et al*., CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017). doi: 10.1038/nmeth.4067; pmid: 27819658
18. E. Harder *et al*., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **12**, 281–296 (2016). doi: 10.1021/acs.jctc.5b00864; pmid: 26584231
19. P. Robustelli, S. Piana, D. E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4758–E4766 (2018). doi: 10.1073/pnas.1800690115; pmid: 29735687
20. P. E. M. Lopes *et al*.., Polarizable Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **9**, 5430–5449 (2013). doi: 10.1021/ct400781b; pmid: 24459460
21. I. Bahar, R. L. Jernigan, K. Dill, *Protein actions: Principles and modeling* (Garland Science, 2017).
22. By *timestep*, we mean that inside the computer, over a discrete unit of time, the algorithm is representing a physical process happening over that unit of time. The computer is approximating Newton's laws and the timesteps must be extremely short to keep the errors acceptably small.
23. S. Izvekov, G. A. Voth, A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **109**, 2469–2473 (2005). doi: 10.1021/jp044629q; pmid: 16851243
24. S. P. Carmichael, M. S. Shell, A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly. *J. Phys. Chem. B* **116**, 8383–8393 (2012). doi: 10.1021/jp2114994; pmid: 22300263
25. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007). doi: 10.1021/jp071097f; pmid: 17569554
26. S. Kmiecik *et al*., Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **116**, 7898–7936 (2016). doi: 10.1021/acs.chemrev.6b00163; pmid: 27333362
27. J. A. McCammon, B. R. Gelin, M. Karplus, Dynamics of folded proteins. *Nature* **267**, 585–590 (1977). doi: 10.1038/267585a0; pmid: 301613
28. X. Daura *et al*., Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **38**, 236–240 (1999). doi: 10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M
29. Y. Duan, L. Wang, P. A. Kollman, The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9897–9902 (1998). doi: 10.1073/pnas.95.17.9897; pmid: 9707572
30. F. Allen *et al*., Blue Gene: A vision for protein science using a petaflop supercomputer. *IBM Syst. J.* **40**, 310–327 (2001). doi: 10.1147/sj.402.0310
31. M. Shirts, V. S. Pande, COMPUTING: Screen Savers of the World Unite! *Science* **290**, 1903–1904 (2000). doi: 10.1126/science.290.5498.1903; pmid: 17742054
32. D. E. Shaw *et al*., Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* **51**, 91–97 (2008). doi: 10.1145/1364782.1364802
33. D. E. Shaw *et al*., *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE, 2014), pp. 41–53.
34. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562–12566 (2002). doi: 10.1073/pnas.202427399; pmid: 12271136
35. Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999). doi: 10.1016/S0009-2614(99)01123-9
36. C. Tsallis, D. A. Stariolo, Generalized simulated annealing. *Physica A* **233**, 395–406 (1996). doi: 10.1016/S0378-4371(96)00271-3
37. E. Darve, A. Pohorille, Calculating free energies using average force. *J. Chem. Phys.* **115**, 9169–9183 (2001). doi: 10.1063/1.1410978
38. G. M. Torrie, J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977). doi: 10.1016/0021-9991(77)90121-8
39. R. C. Bernardi, M. C. Melo, K. Schulten, Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta Gen. Subj.* **1850**, 872–877 (2015). doi: 10.1016/j.bbagen.2014.10.019
40. A. Wang, Z. Zhang, G. Li, Advances in enhanced sampling molecular dynamics simulations for biomolecules. *Chin. J. Chem. Phys.* **32**, 277–286 (2019). doi: 10.1063/1674-0068/cjcp1905091
41. C. Chipot, A. Pohorille, *Free Energy Calculations* (Springer, 2007).
42. M. R. Shirts, J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008). doi: 10.1063/1.2978177; pmid: 19045004
43. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992). doi: 10.1002/jcc.540130812
44. M. P. Rout, A. Sali, Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384–1403 (2019). doi: 10.1016/j.cell.2019.05.016; pmid: 31150619
45. A. Leaver-Fay *et al*., ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011). doi: 10.1016/B978-0-12-381270-4.00019-6; pmid: 21187238
46. D. Xu, Y. Zhang, *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **80**, 1715–1735 (2012). doi: 10.1002/prot.24065; pmid: 22411565
47. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982). doi: 10.1016/0022-2836(82)90153-X; pmid: 7154081
48. D. Kozakov *et al*., The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, 255–278 (2017). doi: 10.1038/nprot.2016.169; pmid: 28079879
49. G. C. P. van Zundert *et al*., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016). doi: 10.1016/j.jmb.2015.09.014; pmid: 26410586
50. N. Eswar *et al*., Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **15**, 5.6.1–5.6.30 (2006). doi: 10.1002/0471250953.bi0506s15; pmid: 18428767
51. D. Russel *et al*., Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLOS Biol.* **10**, e1001244 (2012). doi: 10.1371/journal.pbio.1001244; pmid: 22272186
52. A. Perez, J. L. MacCallum, K. A. Dill, Accelerating molecular simulations of proteins using Bayesian inference on weak information. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11846–11851 (2015). doi: 10.1073/pnas.1515561112; pmid: 26351667
53. J. L. MacCallum, A. Perez, K. A. Dill, Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6985–6990 (2015). doi: 10.1073/pnas.1506788112; pmid: 26038552
54. A. Perez, J. A. Morrone, E. Brini, J. L. MacCallum, K. A. Dill, Blind protein structure prediction using accelerated free-energy simulations. *Sci. Adv.* **2**, e1601274 (2016). doi: 10.1126/sciadv.1601274; pmid: 27847872
55. J. C. Robertson, A. Perez, K. A. Dill, MELD × MD Folds Nonthreadables, Giving Native Structures and Populations. *J. Chem. Theory Comput.* **14**, 6734–6740 (2018). doi: 10.1021/acs.jctc.8b00886; pmid: 30407805
56. J. C. Robertson *et al*., NMR-assisted protein structure prediction with MELDxMD. *Proteins* **87**, 1333–1340 (2019). doi: 10.1002/prot.25788; pmid: 31350773
57. J. A. Morrone *et al*., Molecular Simulations Identify Binding Poses and Approximate Affinities of Stapled α-Helical Peptides to MDM2 and MDMX. *J. Chem. Theory Comput.* **13**,

863–869 (2017). doi: 10.1021/acs.jctc.6b00978; pmid: 28042965

58. A. Khramushin *et al*., Modeling beta-sheet peptide-protein interactions: Rosetta FlexPepDock in CAPRI rounds 38-45. *Proteins* **88**, 1037–1049 (2020). doi: 10.1002/prot.25871; pmid: 31891416
59. S. Kotelnikov *et al*., Sampling and refinement protocols for template-based macrocycle docking: 2018 D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **34**, 179–189 (2020). doi: 10.1007/s10822-019-00257-1; pmid: 31879831
60. A. Perez, F. Sittel, G. Stock, K. Dill, MELD-Path Efficiently Computes Conformational Transitions, Including Multiple and Diverse Paths. *J. Chem. Theory Comput.* **14**, 2109–2116 (2018). doi: 10.1021/acs.jctc.7b01294; pmid: 29547695
61. Y. Shan *et al*., How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **133**, 9181–9183 (2011). doi: 10.1021/ja202726y; pmid: 21545110
62. C. Schütte, A. Fischer, W. Huisinga, P. Deuflhard, A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.* **151**, 146–168 (1999). doi: 10.1006/jcph.1999.6231
63. W. C. Swope, J. W. Pitera, F. Suits, Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory †. *J. Phys. Chem. B* **108**, 6571–6581 (2004). doi: 10.1021/jp037421y
64. F. Noé, I. Horenko, C. Schütte, J. C. Smith, Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys.* **126**, 155102 (2007). doi: 10.1063/1.2714539; pmid: 17461666
65. J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, W. C. Swope, Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **126**, 155101 (2007). doi: 10.1063/1.2714538; pmid: 17461665
66. N.-V. Buchete, G. Hummer, Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **112**, 6057–6069 (2008). doi: 10.1021/jp0761665; pmid: 18232681
67. G. R. Bowman, X. Huang, V. S. Pande, Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **49**, 197–201 (2009). doi: 10.1016/j.ymeth.2009.04.013; pmid: 19410002
68. J.-H. Prinz *et al*., Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105 (2011). doi: 10.1063/1.3565032; pmid: 21548671
69. F. Noé, E. Rosta, Markov Models of Molecular Kinetics. *J. Chem. Phys.* **151**, 190401 (2019). doi: 10.1063/1.5134029
70. V. A. Voelz *et al*., Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. *J. Am. Chem. Soc.* **134**, 12565–12577 (2012). doi: 10.1021/ja302528z; pmid: 22747188
71. S. K. Sadiq, F. Noé, G. De Fabritiis, Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20449–20454 (2012). doi: 10.1073/pnas.1210983109; pmid: 23184967
72. I. Buch, T. Giorgino, G. De Fabritiis, Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10184–10189 (2011). doi: 10.1073/pnas.1103547108; pmid: 21646537
73. D. Huang, A. Caflisch, The free energy landscape of small molecule unbinding. *PLOS Comput. Biol.* **7**, e1002002 (2011). doi: 10.1371/journal.pcbi.1002002; pmid: 21390201
74. A. K. Faradjian, R. Elber, Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **120**, 10880–10889 (2004). doi: 10.1063/1.1738640; pmid: 15268118
75. L. W. Votapka, R. E. Amaro, Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. *PLOS Comput. Biol.* **11**, e1004381 (2015). doi: 10.1371/journal.pcbi.1004381; pmid: 26505480
76. A. E. Cardenas, R. Elber, Markovian and Non-Markovian Modeling of Membrane Dynamics with Milestoning. *J. Phys. Chem. B* **120**, 8208–8216 (2016). doi: 10.1021/acs.jpcb.6b01890; pmid: 27016332
77. B. Narayan *et al*., The transition between active and inactive conformations of Abl kinase studied by rock climbing and Milestoning. *Biochim. Biophys. Acta, Gen. Subj.* **1864**, 129508 (2020). doi: 10.1016/j.bbagen.2019.129508; pmid: 31884066
78. G. A. Huber, S. Kim, Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **70**, 97–110 (1996). doi: 10.1016/S0006-3495(96)79552-8; pmid: 8770190
79. D. M. Zuckerman, L. T. Chong, Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **46**, 43–57 (2017). doi: 10.1146/annurev-biophys-070816-033834; pmid: 28301772
80. J. Moult, J. T. Pedersen, R. Judson, K. Fidelis, A large-scale experiment to assess protein structure prediction methods. *Proteins* **23**, ii–v (1995). doi: 10.1002/prot.340230303; pmid: 8710822
81. S. F. Altschul *et al*., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). doi: 10.1093/nar/25.17.3389; pmid: 9254694
82. If a given type of protein has two amino-acid positions that co-vary together across many different organisms, it can indicate they are in contact in their similar native structures.
83. C. L. Lawson, W. Chiu, Comparing cryo-EM structures. *J. Struct. Biol.* **204**, 523–526 (2018). doi: 10.1016/j.jsb.2018.10.004; pmid: 30321594
84. A. Nicholls *et al*., Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J. Med. Chem.* **51**, 769–779 (2008). doi: 10.1021/jm070549+; pmid: 18215013
85. S. Gathiaka *et al*., D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. *J. Comput. Aided Mol. Des.* **30**, 651–668 (2016). doi: 10.1007/s10822-016-9946-8; pmid: 27696240
86. P. Radivojac *et al*., A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10**, 221–227 (2013). doi: 10.1038/nmeth.2340; pmid: 23353650
87. G. Andreoletti, L. R. Pal, J. Moult, S. E. Brenner, Reports from the fifth edition of CAGI: The Critical Assessment of Genome Interpretation. *Hum. Mutat.* **40**, 1197–1201 (2019). doi: 10.1002/humu.23876; pmid: 31334884
88. G. Stolovitzky, D. Monroe, A. Califano, Dialogue on reverse-engineering assessment and methods: The DREAM of high-throughput pathway inference. *Ann. N. Y. Acad. Sci.* **1115**, 1–22 (2007). doi: 10.1196/annals.1407.021; pmid: 17925349
89. L. A. Abriata, G. E. Tamò, B. Monastyrskyy, A. Kryshtafovych, M. Dal Peraro, Assessment of hard target modeling in CASP12 reveals an emerging role of alignment-based contact prediction methods. *Proteins* **86** (Suppl 1), 97–112 (2018). doi: 10.1002/prot.25423; pmid: 29139163
90. T. I. Croll, M. D. Sammito, A. Kryshtafovych, R. J. Read, Evaluation of template-based modeling in CASP13. *Proteins* **87**, 1113–1127 (2019). doi: 10.1002/prot.25800; pmid: 31407380
91. J. L. MacCallum *et al*., Assessment of protein structure refinement in CASP9. *Proteins* **79** (suppl. 10), 74–90 (2011). doi: 10.1002/prot.23131; pmid: 22069034
92. R. J. Read, M. D. Sammito, A. Kryshtafovych, T. I. Croll, Evaluation of model refinement in CASP13. *Proteins* **87**, 1249–1262 (2019). doi: 10.1002/prot.25794; pmid: 31365160
93. D. Sala *et al*., Protein structure prediction assisted with sparse NMR data in CASP13. *Proteins* **87**, 1315–1332 (2019). doi: 10.1002/prot.25837; pmid: 31603581
94. EMDataResource Validation Challenges, Em validation challenge, https://challenges.emdataresource.org/ (2019); accessed 22 February 2020.
95. C. D. Parks *et al*., D3R grand challenge 4: Blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **34**, 99–119 (2020). doi: 10.1007/s10822-020-00289-y; pmid: 31974851
96. J. Zou, C. Tian, C. Simmerling, Blinded prediction of protein-ligand binding affinity using Amber thermodynamic integration for the 2018 D3R grand challenge 4. *J. Comput. Aided Mol. Des.* **33**, 1021–1029 (2019). doi: 10.1007/s10822-019-00223-x; pmid: 31555923
97. D. L. Mobley *et al*., Blind prediction of HIV integrase binding from the SAMPL4 challenge. *J. Comput. Aided Mol. Des.* **28**, 327–345 (2014). doi: 10.1007/s10822-014-9723-5; pmid: 24595873
98. E. Gallicchio *et al*., Virtual screening of integrase inhibitors by large scale binding free energy calculations: The SAMPL4 challenge. *J. Comput. Aided Mol. Des.* **28**, 475–490 (2014). doi: 10.1007/s10822-014-9711-9; pmid: 24504704
99. J. Dapkūnas, K. Olechnovič, Č. Venclovas, Structural modeling of protein complexes: Current capabilities and challenges. *Proteins* **87**, 1222–1232 (2019). doi: 10.1002/prot.25774; pmid: 31294859
100. M. F. Lensink *et al*., Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins* **87**, 1200–1221 (2019). doi: 10.1002/prot.25838; pmid: 31612567
101. M. F. Lensink, N. Nadzirin, S. Velankar, S. J. Wodak, Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. *Proteins* **88**, 916–938 (2020). doi: 10.1002/prot.25870; pmid: 31886916
102. D. J. Huggins *et al*., Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **9**, e1393 (2019). doi: 10.1002/wcms.1393
103. D. E. Shaw *et al*., Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341–346 (2010). doi: 10.1126/science.1187409; pmid: 20947758
104. K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding proteins fold. *Science* **334**, 517–520 (2011). doi: 10.1126/science.1208351; pmid: 22034434
105. H. Nguyen, J. Maier, H. Huang, V. Perrone, C. Simmerling, Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* **136**, 13959–13962 (2014). doi: 10.1021/ja5032776; pmid: 25255057
106. D. L. Mobley *et al*., Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **371**, 1118–1134 (2007). doi: 10.1016/j.jmb.2007.06.002; pmid: 17599350
107. S. E. Boyce *et al*., Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **394**, 747–763 (2009). doi: 10.1016/j.jmb.2009.09.049; pmid: 19782087
108. L. Wang *et al*., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 (2015). doi: 10.1021/ja512751q; pmid: 25625324
109. G. D. R. Matos *et al*., Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database. *J. Chem. Eng. Data* **62**, 1559–1569 (2017). doi: 10.1021/acs.jced.7b00104; pmid: 29056756
110. E. P. Barros *et al*., Improving the Efficiency of Ligand-Binding Protein Design with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **15**, 5703–5715 (2019). doi: 10.1021/acs.jctc.9b00483; pmid: 31442033
111. M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, P. C. Biggin, Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J. Am. Chem. Soc.* **139**, 946–957 (2017). doi: 10.1021/jacs.6b11467; pmid: 28009512
112. J. Zou, C. Simmerling, D. P. Raleigh, Dissecting the Energetics of Intrinsically Disordered Proteins via a Hybrid Experimental and Computational Approach. *J. Phys. Chem. B* **123**, 10394–10402 (2019). doi: 10.1021/acs.jpcb.9b08323; pmid: 31702919
113. J. Zou, B. Song, C. Simmerling, D. Raleigh, Experimental and Computational Analysis of Protein Stabilization by Gly-to-D-Ala Substitution: A Convolution of Native State and Unfolded State Effects. *J. Am. Chem. Soc.* **138**, 15682–15689 (2016). doi: 10.1021/jacs.6b09511; pmid: 27934019
114. D. L. Mobley *et al*., Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **14**, 6076–6092 (2018). doi: 10.1021/acs.jctc.8b00640; pmid: 30351006
115. A. T. Fenley, H. S. Muddana, M. K. Gilson, Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20006–20011 (2012). doi: 10.1073/pnas.1213180109; pmid: 23150595
116. A. Li, M. K. Gilson, Protein-ligand binding enthalpies from near-millisecond simulations: Analysis of a preorganization paradox. *J. Chem. Phys.* **149**, 072311 (2018). doi: 10.1063/1.5027439; pmid: 30134726
117. R. A. Copeland, The drug-target residence time model: A 10-year retrospective. *Nat. Rev. Drug Discov.* **15**, 87–95 (2016). doi: 10.1038/nrd.2015.18; pmid: 26678621
118. S. Re, H. Oshima, K. Kasahara, M. Kamiya, Y. Sugita, Encounter complexes and hidden poses of kinase-inhibitor binding on the free-energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 18404–18409 (2019). doi: 10.1073/pnas.1904707116; pmid: 31451651
119. P. Tiwary, J. Mondal, B. J. Berne, How and when does an anticancer drug leave its binding site? *Sci. Adv.* **3**, e1700014 (2017). doi: 10.1126/sciadv.1700014; pmid: 28580424
120. S. D. Lotz, A. Dickson, Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *J. Am. Chem. Soc.* **140**, 618–628 (2018). doi: 10.1021/jacs.7b08572; pmid: 29303257
121. C.-T. Lai *et al*., Rational Modulation of the Induced-Fit Conformational Change for Slow-Onset Inhibition in

Mycobacterium tuberculosis InhA. *Biochemistry* **54**, 4683–4691 (2015). doi: 10.1021/acs.biochem.5b00284; pmid: 26147157
122. H.-J. Li *et al*., A structural and energetic model for the slow-onset inhibition of the *Mycobacterium tuberculosis* enoyl-ACP reductase InhA. *ACS Chem. Biol.* **9**, 986–993 (2014). doi: 10.1021/cb400896g; pmid: 24527857
123. D. K. Klimov, D. Thirumalai, Dissecting the assembly of Abeta16-22 amyloid peptides into antiparallel β sheets. *Structure* **11**, 295–307 (2003). doi: 10.1016/S0969-2126(03)00031-5; pmid: 12623017
124. D. Thirumalai, G. Reddy, J. E. Straub, Role of water in protein aggregation and amyloid polymorphism. *Acc. Chem. Res.* **45**, 83–92 (2012). doi: 10.1021/ar2000869; pmid: 21761818
125. L. E. Buchanan *et al*., Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient β-sheet. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19285–19290 (2013). doi: 10.1073/pnas.1314481110; pmid: 24218609
126. J. Nasica-Labouze *et al*., Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chem. Rev.* **115**, 3518–3563 (2015). doi: 10.1021/cr500638n; pmid: 25789869
127. T. Zhang, W. Xu, Y. Mu, P. Derreumaux, Atomic and dynamic insights into the beneficial effect of the 1,4-naphthoquinon-2-yl-L-tryptophan inhibitor on Alzheimer's Aβ1-42 dimer in terms of aggregation and toxicity. *ACS Chem. Neurosci.* **5**, 148–159 (2014). doi: 10.1021/cn400197x; pmid: 24246047
128. R. Avinery, M. Kornreich, R. Beck, Universal and Accessible Entropy Estimation Using a Compression Algorithm. *Phys. Rev. Lett.* **123**, 178102 (2019). doi: 10.1103/PhysRevLett.123.178102; pmid: 31702252
129. F. Noé, S. Olsson, J. Köhler, H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019). doi: 10.1126/science.aaw1147; pmid: 31488660
130. M. E. Tuckerman, Machine learning transforms how microstates are sampled. *Science* **365**, 982–983 (2019). doi: 10.1126/science.aay2568; pmid: 31488674
131. A. W. Senior *et al*., Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020). doi: 10.1038/s41586-019-1923-7; pmid: 31942072
132. J. Xu, Distance-based protein folding powered by deep learning. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16856–16865 (2019). doi: 10.1073/pnas.1821309116; pmid: 31399549
133. X. Jia *et al*., Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis. *Nature* **573**, 251–255 (2019). doi: 10.1038/s41586-019-1540-5; pmid: 31511682
134. L. M. Jarvis, Genentech's R&D chief Michael Varney on the future of drug discovery. *C&EN* **97** (2019); https://cen.acs.org/biological-chemistry/biotechnology/Genentechs-RD-chief-Michael-Varney/97/i31.
135. H. S. Chan, K. A. Dill, The Protein Folding Problem. *Phys. Today* **46**, 24–32 (1993). doi: 10.1063/1.881371
136. A. Perez, J. A. Morrone, C. Simmerling, K. A. Dill, Advances in free-energy-based simulations of protein folding and ligand binding. *Curr. Opin. Struct. Biol.* **36**, 25–31 (2016). doi: 10.1016/j.sbi.2015.12.002; pmid: 26773233

ACKNOWLEDGMENTS

We thank D. Case, M. Feig, M. Gilson, D. Kozakov, A. Kryshtafovych, A. MacKerrell, D. Mobley, J. Moult, R. Nassar, D. Padhorny, R. Read, and B. Sharma for valuable comments and insights. We are grateful to S. Bromberg for help with figures and her deeply insightful comments and editing. Thanks also to the anonymous referees, who provided excellent feedback and constructive criticism. **Funding:** We are grateful for the support from NIH grants 5R01GM107104, 5R01GM125813, and RM1GM135136 and the Stony Brook Laufer Center for Physical and Quantitative Biology. **Competing interests:** The authors have no competing interests to declare. **Data and materials availability:** Data presented in Figs. 1, 3C, and 4 and have been previously published in (*57*, *59*, *108*, *111*, *112*, *136*); data presented in Figs. 2 and 3, A and B, are publicly available at the competition websites.

10.1126/science.aaz3041

RESEARCH ARTICLE SUMMARY

NEUROGENOMICS

In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes

Xin Jin*, Sean K. Simmons, Amy Guo, Ashwin S. Shetty, Michelle Ko, Lan Nguyen, Vahbiz Jokhi, Elise Robinson, Paul Oyler, Nathan Curry, Giulio Deangeli, Simona Lodato, Joshua Z. Levin, Aviv Regev*†, Feng Zhang*†, Paola Arlotta*†

INTRODUCTION: Human genetic studies have revealed long lists of genes and loci associated with risk for many diseases and disorders, but to systematically evaluate their phenotypic effects remains challenging. Without any a priori knowledge, these risk genes could affect any cellular processes in any cell type or tissue, which creates an enormous search space for identifying possible downstream effects. New high-throughput approaches are needed to functionally dissect these large gene sets across a spectrum of cell types in vivo.

RATIONALE: Analysis of trio-based whole-exome sequencing has implicated a large number of de novo loss-of-function variants that contribute to autism spectrum disorder and developmental delay (ASD/ND) risk. Such de novo variants often have large effect sizes, thus providing a key entry point for mechanistic studies. We have developed in vivo Perturb-Seq to allow simultaneous assessment of the individual phenotypes of a panel of such risk genes in the context of the developing mouse brain.

RESULTS: Using CRISPR-Cas9, we introduced frameshift mutations in 35 ASD/ND risk genes in pools, within the developing mouse neocortex in utero, followed by single-cell transcriptomic analysis of perturbed cells from the early postnatal brain. We analyzed five broad cell classes—cortical projection neurons, cortical inhibitory neurons, astrocytes, oligodendrocytes, and microglia—and selected cells that had received only single perturbations. Using weighted gene correlation network analysis, we identified 14 covarying gene modules that represent transcriptional programs expressed in different classes of cortical cells.

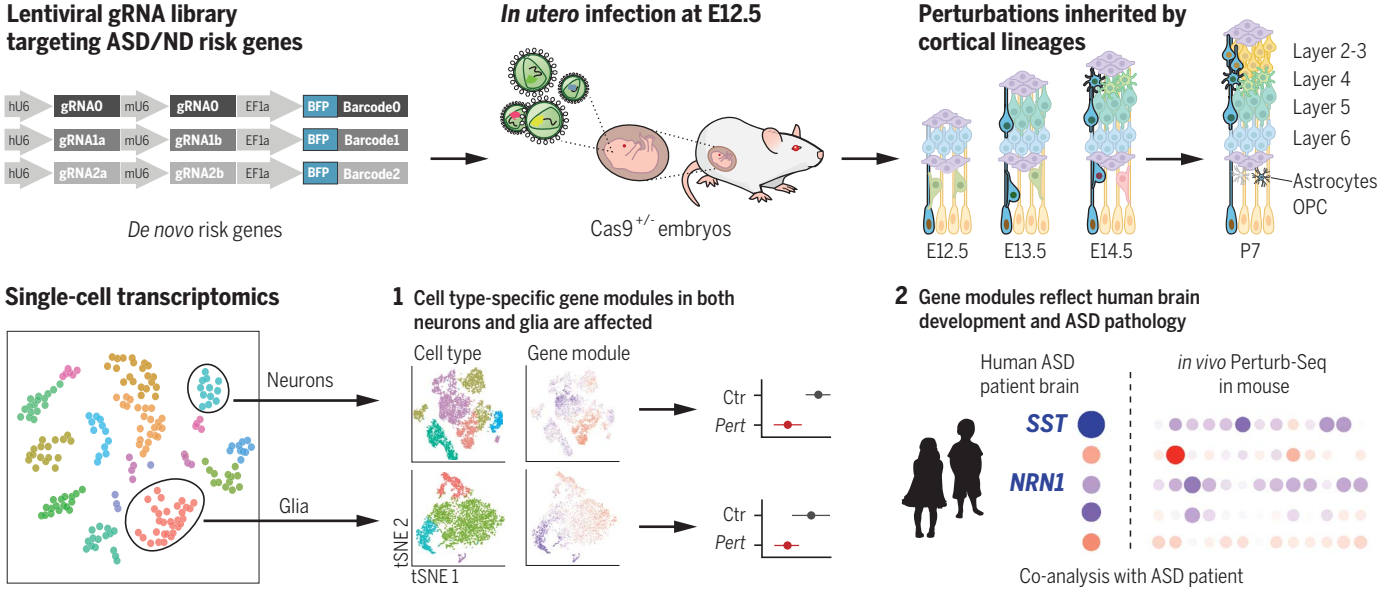
These modules included both those affecting common biological processes across multiple cell subsets and others representing cell type–specific features restricted to certain subsets. We estimated the effect size of each perturbation on each of the 14 gene modules by fitting a joint linear regression model, estimating how module gene expression in cells from each perturbation group deviated from their expression level in internal control cells. Perturbations in nine ASD/ND genes had significant effects across five modules across four cell classes, including cortical projection neurons, cortical inhibitory neurons, astrocytes, and oligodendrocytes. Some of these results were validated by using a single-perturbation model as well as a germline-modified mutant mouse model.

To establish whether the perturbation-associated gene modules identified in the mouse cerebral cortex are relevant to human biology and ASD/ND pathology, we performed co-analyses of data from ASD and control human brains and human cerebral organoids. Both gene expression and gene covariation ("modularity") of several of the gene modules identified in the mouse Perturb-Seq analysis are conserved in human brain tissue. Comparison with single-cell data from ASD patients showed overlap in both affected cell types and transcriptomic phenotypes.

CONCLUSION: In vivo Perturb-Seq can serve as a scalable tool for systems genetic studies of large gene panels to reveal their cell-intrinsic functions at single-cell resolution in complex tissues. In this work, we demonstrated the application of in vivo Perturb-Seq to ASD/ND risk genes in the developing brain. This method can be applied across diverse diseases and tissues in the intact organism. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: xinjin@fas.harvard.edu (X.J.); aregev@broadinstitute.org (A.R.); zhang@broadinstitute.org (F.Z.); paola_arlotta@harvard.edu (P.A.)
†These authors contributed equally to this work.
Cite this article as X. Jin *et al.*, *Science* **370**, eaaz6063 (2020). DOI: 10.1126/science.aaz60603

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aaz6063



In vivo Perturb-Seq identified neuron and glia-associated effects by perturbations of risk genes implicated in ASD/ND. De novo risk genes in this study were chosen from Satterstrom *et al.* (2018), and co-analysis with ASD patient data at bottom right is from Velmeshev *et al.* (2019); full citations for both are included in the full article online.

RESEARCH ARTICLE

NEUROGENOMICS

In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes

Xin Jin[1,2,3,4]*, Sean K. Simmons[3,5,6], Amy Guo[3], Ashwin S. Shetty[2,3,5], Michelle Ko[2], Lan Nguyen[3,6], Vahbiz Jokhi[2], Elise Robinson[3,5,7], Paul Oyler[2], Nathan Curry[2], Giulio Deangeli[2], Simona Lodato[8], Joshua Z. Levin[3,5,6], Aviv Regev[3,6,9,10]*†‡, Feng Zhang[3,4,10]*†, Paola Arlotta[2,3,5]*†

The number of disease risk genes and loci identified through human genetic studies far outstrips the capacity to systematically study their functions. We applied a scalable genetic screening approach, in vivo Perturb-Seq, to functionally evaluate 35 autism spectrum disorder/neurodevelopmental delay (ASD/ND) de novo loss-of-function risk genes. Using CRISPR-Cas9, we introduced frameshift mutations in these risk genes in pools, within the developing mouse brain in utero, followed by single-cell RNA-sequencing of perturbed cells in the postnatal brain. We identified cell type–specific and evolutionarily conserved gene modules from both neuronal and glial cell classes. Recurrent gene modules and cell types are affected across this cohort of perturbations, representing key cellular effects across sets of ASD/ND risk genes. In vivo Perturb-Seq allows us to investigate how diverse mutations affect cell types and states in the developing organism.

Human genetic studies have uncovered associations between genetic variants in tens of thousands of loci and complex human diseases (*1*–*3*). In particular, analysis of trio-based whole-exome sequencing (WES) has implicated a large number of de novo loss-of-function variants that contribute to the risk of neurodevelopmental pathologies, including autism spectrum disorders and neurodevelopmental delay (ASD/ND) (*4*, *5*). Compared with common variants identified through genome-wide association studies (GWASs), such de novo risk variants often have large effect sizes, are highly penetrant, and occur within a gene's coding region, thus providing a crucial entry point for disease modeling and mechanistic studies.

However, a major challenge remains in identifying the point of action of each of these many risk genes. For example, ASD/ND comprises a broad collection of neurodevelopmental disorders with highly heterogeneous genetic contributions, including hundreds of highly penetrant de novo risk variant genes (*6*). Moreover, there is substantial diversity in the function of the gene products that these genes encode, precluding a clear a priori prediction of the underlying brain cell types, developmental processes, and molecular pathways affected during neurodevelopment (*7*). Few of these risk genes have been studied in animal or cellular models, and thus, their function during brain development remains poorly defined. Because of the labor-intensive and time-consuming nature of generating and analyzing individual knockout animal models for functional investigation, it is crucial to develop phenotyping methods that are scalable, general-purpose, high-resolution, and high-content to identify tissue- and cell type–specific effects of genetic perturbations in vivo.

To address these challenges, we developed in vivo Perturb-Seq, a scalable genetic screen, to investigate the function of large sets of genes at single-cell resolution in complex tissue in vivo. We applied in vivo Perturb-Seq in utero in mice to study the effect of a panel of ASD/ND risk genes on brain development.

In vivo Perturb-Seq to assess the function of ASD/ND risk genes

We chose ASD/ND candidate genes from a recently published WES study of 11,986 cases with 6430 ASD/ND probands (table S1) (*4*). We initially prioritized 38 candidate genes (of which 35 were retained in the final analysis) (table S1) that harbor de novo variants specific to ASD/ND patients within the broader class of neurodevelopmental disability (fig. S1A and table S1). These ASD/ND risk genes are expressed in human brain tissue, as assessed by the BrainSpan bulk RNA-seq dataset (*8*); some are highly expressed at embryonic stages, and others are highly expressed from early postnatal to adult stages (fig. S1B). On the basis of mouse cortical single-cell RNA-sequencing (scRNA-seq) data, the orthologs of these ASD/ND risk genes are expressed in diverse cell types (fig. S2) [embryonic day 18.5 (E18.5) data are from the 10x Genomics public dataset (*9*); P7 data are from this work]. Thus, these ASD/ND genes could in principle act in many different cell types and temporal frames, requiring scalable methods to test gene function across a range of cell types and developmental events.

For in vivo Perturb-Seq, we used Cas9-mediated genome editing (*10*–*12*) in a pooled approach to introduce mutations in each of the ASD/ND risk genes within progenitor cells of the mouse developing forebrain in utero, followed by scRNA-seq at P7 to read out both a barcode identifying the perturbation and the expression profile of the perturbed cells (Fig. 1A). Specifically, we used a transgenic mouse line that constitutively expresses Cas9 (*13*) and delivered pools of guide RNAs (gRNAs) targeting the different risk genes through lentiviral infection into the lateral ventricles of the developing embryo in utero. Each lentiviral vector contained two different gRNAs targeting the 5′-end coding exons of one ASD/ND gene (to enhance knockout efficiency) and a blue fluorescent protein (BFP) reporter with a distinct barcode corresponding to the perturbation identity (*10*–*12*). To minimize vector recombination, we packaged each lentivirus separately and then pooled viruses at equal titers.

We injected a pool of lentiviruses with equal gRNA representation into the ventricles of the developing forebrain at E12.5 (Fig. 1A). In this approach, lentiviral injection leads to infection of neural progenitors lining the lateral ventricle of the developing forebrain, including progenitors of the neocortex and the ganglionic eminences. Because lentiviral vectors integrate into the genome, the progeny of the infected progenitors are labeled by BFP and carry a perturbation barcode corresponding to the target ASD/ND gene.

Both immunohistochemical analysis and scRNA-seq of BFP$^+$ cells at P7 showed that the Perturb-Seq vectors were expressed across a variety of neuronal and glial cell types in the cortex (Fig. 1, B and C, and fig. S3, A and B). Although microglia originate mostly from outside the targeted germinal zones, we detected lentiviral vector expression in cortical microglia, indicated by the presence of BFP as well as perturbation barcode expression, across multiple individual experiments (fig. S4, E and F). Although it is unclear how microglia are labeled in our experimental procedure, it is possible that the in utero injection could

[1]Society of Fellows, Harvard University, Cambridge, MA, USA. [2]Department of Stem Cell and Regenerative Biology, Harvard University, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]McGovern Institute of Brain Science, Department of Brain and Cognitive Science, Department of Biological Engineering, Massachussetts Institute of Technology (MIT), Cambridge, MA, USA. [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [8]Department of Biomedical Sciences and Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Humanitas Clinical and Research Center, Humanitas University, Milan, Italy. [9]Koch Institute of Integrative Cancer Research, Department of Biology, MIT, Cambridge, MA, USA. [10]Howard Hughes Medical Institute, Chevy Chase, MD, USA.
***Corresponding author. Email: xinjin@fas.harvard.edu (X.J.); aregev@broadinstitute.org (A.R.); zhang@broadinstitute.org (F.Z.); paola_arlotta@harvard.edu (P.A.)**
†These authors contributed equally to this work.
‡Present address: Genentech, 1 DNA Way, South San Francisco, CA.

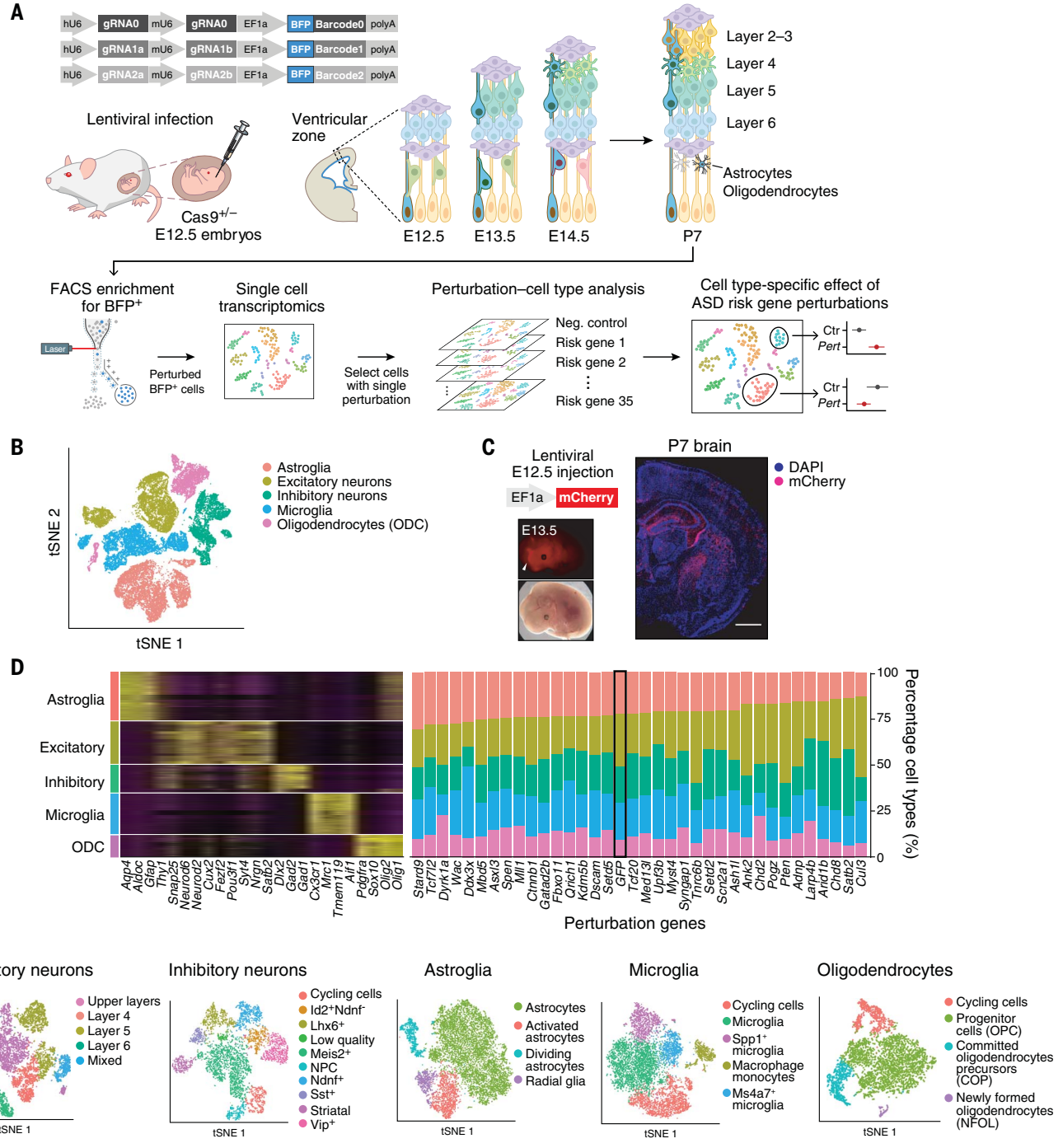


Fig. 1. In vivo Perturb-Seq to investigate functions of a panel of ASD/ND risk genes harboring de novo variants. (**A**) Schematics of the in vivo Perturb-Seq platform, which introduces mutations in individual genes in utero at E12.5, followed by transcriptomic profiling of the cellular progeny of these perturbed cells at P7 by means of scRNA-seq. (**B**) *t*-distributed stochastic neighbor embedding (*t*-SNE) of five major cell populations identified in the Perturb-Seq cells. (**C**) In vivo Perturb-Seq lentiviral vector carrying an mCherry reporter drives detectable expression within 24 hours and can sparsely infect brain cells across many brain regions. Scale bar, 1000 μm. (**D**) Cell-type analysis of in vivo Perturb-Seq of ASD/ND de novo risk genes. Canonical marker genes were used to identify (left) major cell clusters and (right) cell-type distribution in each perturbation group. Negative control (GFP) is indicated with a black rectangle. (**E**) *t*-SNEs showing the subclusters of each of the five major cell types, identified by reclustering each cell type separately.

have led to either local lesions that recruited and expanded the number of microglia along the injection tract or that microglia were labeled within the parenchyma along the same tract. Overall, our approach allowed us to examine the effects of each perturbation across a wide range of cell types from distinct brain regions (such as cortical projection neurons, interneurons, astroglia, and oligodendroglia), and under sparse labeling conditions in which less than 0.1% of cells in the cortex were perturbed, and thus development of individual perturbed cells was highly unlikely to be affected by perturbed neighboring cells (fig. S3, A to C).

In vivo Perturb-Seq targets diverse cell types without affecting overall cell type composition

We performed the experiment with 18 different cohorts of pregnant mice, for a total of 163 embryos, each subjected to the entire pool of perturbations. We microdissected and dissociated cortical tissues at P7, enriched the perturbed cells by means of fluorescence-activated cell sorting (FACS), and used droplet-based scRNA-seq to obtain each cell's expression profile along with its perturbation barcode. The cell survival rate after FACS was 78%, and we confirmed a 40 to 70% frameshift insertion or deletion for each gRNA target among the infected cells (fig. S3, D and E).

This multiplexed experimental design allowed us to test the cell-autonomous effect of all perturbations against the effect of a negative control construct targeting the endogenous green fluorescent protein (GFP) in the Rosa26 locus, thus controlling for effects related to viral infection and DNA double-strand break, among other confounders. To minimize batch-dependent variation, the control construct was included in the same pool as that of the perturbation vectors (fig. S3F). After quality control, we retained for further analysis a total of 46,770 neocortical cells across 17 high-quality experimental batches. We partitioned the cells into major cell classes using Louvain clustering (*14*) and annotated them by known marker gene expression (Fig. 1D) (*15*, *16*).

We focused on five broad cell populations from this cortical dataset for downstream analysis: cortical projection neurons (8450 cells), cortical inhibitory neurons (5532 cells), astrocytes (9526 cells), oligodendrocytes (4279 cells), and microglia/macrophages (8070 cells) (thus excluding vascular, endothelial, and contaminant hippocampal and striatal cells). We further filtered out some remaining low-quality cells in these five major cell classes, retaining 35,857 high-quality cells (median of 2436 detected genes per cell overall, and a median of 4084 genes in the projection neuron cluster, as expected from their large size and known high RNA content) (fig. S3G). We subclustered each of the five major cell types separately and annotated biologically meaningful subclusters (Fig. 1E and fig. S6).

From inspecting the perturbation barcodes from the lentiviral constructs, 92% of the cells (33,231 cells) in these five major cell classes had at least one perturbation read assigned to them, and 50% had barcodes for a single gene (18,044 cells) (fig. S4, A to C), reflecting the low multiplicity of infection (fig. S4D). Because it is rare for multiple ASD/ND loss-of-function risk gene mutations to co-occur in patients, we focused on the 18,044 cells that carried a single perturbation. We found a median of 338 cells per perturbation. After excluding perturbations with <70 perturbed cells, we retained 35 ASD/ND risk gene perturbations. BFP from the lentiviral vector was robustly detected as one of the most highly expressed genes in all retained cells (fig. S4E). The BFP detection rate in each cell type correlated with the average number of genes detected (fig. S4F), further supporting the reliability of the readout.

ASD/ND risk gene perturbations had a very modest effect on the presence and proportions of these five major cell types relative to the negative control (targeting the *GFP* gene). Only loss of *Dyrk1a* had a significant effect on cell type composition, increasing the proportion of oligodendrocytes and reducing the proportion of microglia and macrophages [false discovery rate (FDR)–corrected $P < 0.05$ using Poisson regression (*17*)] (Fig. 1D and fig. S5).

Covarying gene modules associate with cell states

To assess whether ASD/ND genetic perturbations caused molecular changes and alterations in cell states, we first sought to define gene modules that covary within each of the five broad cell classes. As previous work has shown (*10*–*12*, *18*), focusing on gene modules instead of individual genes provides more statistical power to detect biologically meaningful perturbation effects by using fewer cells than would be required for single-gene–level analysis and can capture diversity both within and across cell types.

We first tested whether the expression of known Gene Ontology (GO) gene sets (*19*) was affected by calculating a gene-set expression score for each cell and fitting a linear regression model to this score. After correcting for multiple hypothesis testing, no GO terms were significantly altered by any perturbation (table S8). However, this approach is limited by the large number of tests performed (one test per GO term per cell type per perturbation, for a total of 510,265 tests), as well as the limited number of GO terms relevant to the developing cortex.

We therefore sought to identify gene modules de novo in our data using two approaches: weighted gene correlation network analysis (WGCNA), which identifies "modules" of genes with correlated expression, and structural topic modeling (STM), which attempts to reduce the dimensionality of the gene expression matrix and returns "topics" corresponding to the components of this representation (Fig. 2A, figs. S6 to S8, and tables S2 and S3) (*20*, *21*). We performed these analyses for each of the five major cell clusters separately, to better identify effects associated with specific cell types; our nomenclature for the modules incorporates the cell cluster analysis it is derived from (for example, PN1 represents a module identified through analysis of projection neurons). Each of these analyses used the full set of perturbations to identify effects shared across multiple perturbations. We focused our subsequent analysis on the 14 modules identified with WGCNA because they were highly correlated with one or more topics returned by STM (fig. S7).

The 14 WGCNA modules comprised two broad categories. Some reflected common biological processes and were present across multiple cell subsets (such as cell cycle, differentiation, and maturation). For example, module PN2 is associated with genes involved in neurite development and varied across cells in multiple projection neuron subclusters (fig. S6A). Others represented cell type–specific features specific to only some subsets (such as subcluster-specific features of a neuronal subtype). For example, module PN1 is a module associated with two defined subclusters of neurons of layer 4 and layer 5 (fig. S6A).

ASD/ND gene perturbations affect cell states in multiple cell classes

Because the WGCNA analysis is expected to recover gene modules associated with many kinds of variation across the data, we next tested the association of each risk gene perturbation with the 14 individual WGCNA gene modules. We estimated the effect size of each perturbation on each gene module by fitting a joint linear regression model, estimating how module gene expression in cells from each perturbation group deviated from the GFP control cells (Fig. 2, A and B). To ensure that no single perturbation or batch dominated the linear model, we down-sampled the cells in each cell category so that no perturbation had more than two times the median number of cells over all perturbations. This linear regression analysis was performed on mean-centered and standard deviation–scaled module scores, so effect sizes can be interpreted in terms of standard deviations from the population mean (Fig. 2B). Our modeling approach assumes that module expression in individual cells is independent after conditioning on the experimental batch and that noise is normally distributed. To evaluate the effects of these assumptions, we also compared alternative approaches, including a linear mixed model–based approach and a permutation-based approach (fig. S9 and table S9).

Perturbations in nine ASD/ND genes (*Adnp*, *Ank2*, *Ash1l*, *Chd8*, *Gatad2b*, *Pogz*, *Scn2a1*, *Stard9*, and *Upf3b*) had significant effects across five modules (compared with the GFP control, FDR-corrected $P < 0.05$) (Fig. 2B, indicated circles, and table S4): a module associated with neurons of layers 4 and 5 (PN1, affected by perturbations in *Adnp*, *Ash1l*, *Scn2a1*, and *Stard9*); modules representing two distinct homeostatic signatures in astrocytes (Astro1 affected by perturbation of *Scn2a1*, and Astro3 affected by perturbations of *Chd8*, *Pogz*, and *Upf3b*); a module associated with oligodendrocyte progenitor cells (ODC1, affected by perturbations of *Chd8*

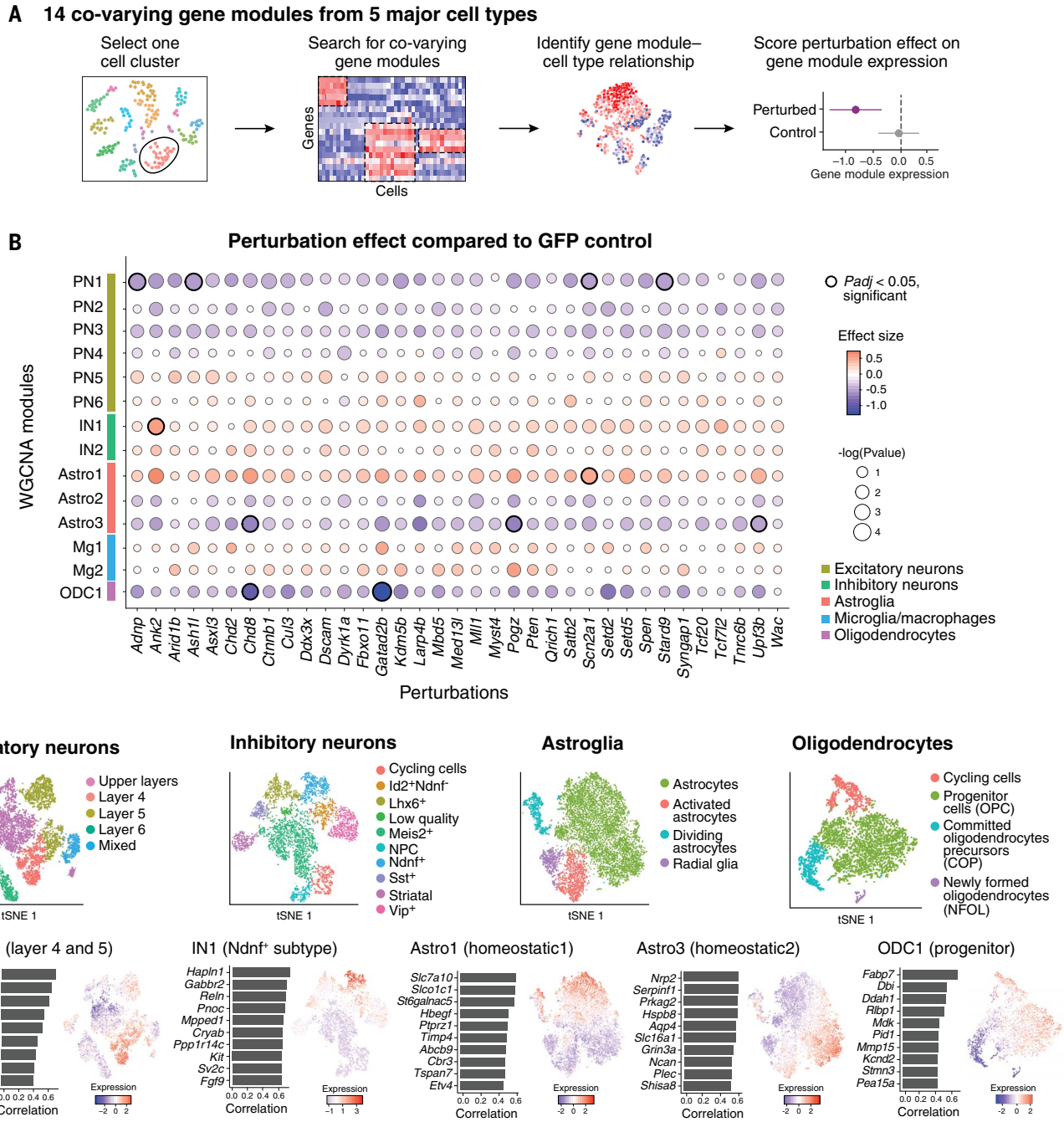


Fig. 2. In vivo Perturb-Seq reveals cell type–specific effects of ASD/ND risk gene perturbations. (**A**) Schematic illustration of the Perturb-Seq analysis pipeline. (**B**) ASD/ND risk gene perturbation effects in different WGCNA gene modules compared with GFP controls. Dot color corresponds to effect size, and dot size corresponds to negative base 10 log(*P* value). Module gene lists are presented in table S2. *P* values were calculated from linear modeling, and P_{adj} was calculated by means of Benjamini and Hochberg FDR correction. (**C**) The four cell types and five gene modules that were altered by ASD risk gene perturbations. (Top) Subcluster *t*-SNE of each cell class (repeated from Fig. 1E for ease of comparison). (Bottom) Feature plots of gene module expression scores and the top correlated genes within each module across the relevant cell class.

and *Gatad2b*); and a module associated with *Ndnf*$^+$ interneurons (IN1, affected by the perturbation of *Ank2*) (Fig. 2C and fig. S6).

The oligodendrocyte progenitor module (ODC1) also had a significant amount of its variation across the oligodendrocyte cell cluster explained by the perturbation state overall [van der Waerden test, a nonparametric alternative to analysis of variance (ANOVA), FDR-corrected $P < 0.05$] (fig. S5C), suggesting that this module represents most shared effects across different perturbed genes. Collectively, the data indicate that a selected group of perturbations was able to affect recurrent gene modules with cell-type specificity and point to some

convergent effects across diverse ASD/ND risk genes.

Single perturbation of *Ank2* confirms Perturb-Seq effect on an interneuron gene module

In our multiplex in vivo Perturb-Seq results, *Ank2* perturbation led to increased expression of an interneuron module (IN1) (FDR-corrected $P < 0.05$) (fig. S10). This module was strongly correlated with a subcluster of inhibitory interneurons expressing *Ndnf* (fig. S10, C and D) and contains genes such as *Kcnq5* (a voltage-gated potassium channel) and *Gabbr2* [γ-aminobutyric acid (GABA) receptor subunit] (fig. S6B and table S2). To validate our finding from the pooled Perturb-Seq experiment, we performed a single perturbation targeting either *Ank2* or *GFP* (control), followed by scRNA-seq of neocortical cells at P7, resulting in 2943 and 1716 high-quality cells, respectively.

The individual perturbation experiment confirmed the results from the pooled Perturb-Seq screen. *Ank2*-perturbed cells were present across all cell types, and overall proportions of cells were not significantly changed (fig. S10B). Within the *Ndnf*$^+$ interneurons, *Ank2* perturbation led to up-regulation of the IN1 module (FDR-corrected $P < 0.05$) (fig. S10E), confirming the Perturb-Seq result. This finding indicates that multiplexing perturbations in the pooled approach does not significantly distort the results observed for an individually perturbed gene.

Ank2 encodes an ankyrin protein and is expressed broadly in excitatory and inhibitory neurons as well as glial cells in the brain (*22*). Studies examining *Ank2* loss of function suggest that it is involved in axonal morphology, connectivity, and calcium signaling in excitatory neurons (*23*–*26*). Our Perturb-Seq data suggests a role of *Ank2* in the *Ndnf*$^+$ interneuron subtype during cortical development, in addition to its known roles in excitatory neurons.

The ASD/ND risk genes *Chd8* and *Gatad2b* alter gene modules in oligodendrocyte progenitors

In our Perturb-Seq experiment, *Chd8* and *Gatad2b* perturbations significantly decreased the expression of the ODC1 module in the oligodendrocyte cluster (FDR-corrected $P < 0.05$) [Fig. 3, A to D; an alternative measurement of effect size is provided in fig. S11A, estimated by log transcripts per million (TPM) gene expression differences]. The ODC1 module is highly expressed in cycling cells and oligodendrocyte precursor cells (OPCs) and lowly expressed in committed oligodendrocyte progenitor cells (COPs) and newly formed oligodendrocytes (NFOLs), suggesting that this module is linked to oligodendrocyte maturation (Fig. 3A) and therefore that perturbation in *Chd8* and *Gatad2b* might alter oligodendrocyte maturation. This is consistent with recent reports that *Chd8* loss of function potentiates an impaired OPC development phenotype caused by deletion of *Chd7* (*27*).

We further investigated and validated this result by examining oligodendrocyte development in a *Chd8* germline heterozygous mutant model [because homozygous mutation is embryonic lethal (*28*)], using several orthogonal methods. First, we used in situ hybridization for two canonical OPC markers that are involved in fate specification, *Cspg4* (a member of the ODC1 module) and *Pdgfra* (platelet-derived growth factor receptor A). Both were down-regulated in P7 *Chd8*$^{+/-}$ cortex (Fig. 3E and fig. S11, B to D), which is consistent with our in vivo Perturb-Seq results. Second, we used immunohistochemistry to examine a later developmental time point, P11. OPC cell number (such as PDGFRA$^+$ cells) did not show significant differences between the wild-type and *Chd8*$^{+/-}$ littermates, which is also consistent with in vivo Perturb-Seq; however, cells positive for myelin basic protein (MBP), a marker of myelinating oligodendrocytes, were increased in number and displayed increased MBP amounts in the *Chd8*$^{+/-}$ mutant (FDR-corrected $P < 0.05$, nonparametric ANOVA test) (Fig. 3F). In combination with the Perturb-Seq result that showed reduction in the signature of oligodendrocyte progenitor-expressed ODC1 module in *Chd8*-perturbed cells, this suggests that C*hd8* perturbation may result in acceleration of the increase in MBP amounts that occurs postnatally. These data further demonstrate that in vivo Perturb-Seq has the power to identify cell type–specific molecular changes similar to those observed in a single-gene, germline-modified mouse model.

Perturb-Seq gene modules are conserved between human and mouse

To establish whether the perturbed gene modules identified in the mouse cerebral cortex are conserved in human cells, we examined the expression of each module across multiple scRNA-seq and single-nuclei RNA-seq (snRNA-seq) datasets from human tissues: adult human cortex (*29*), ASD donor cortex with matched controls (*31*), fetal human cortex (*31*), and 3- and 6-month-old human brain organoids (Fig. 4, A and B) (*32*). In the fetal brain and the 3-month-old brain organoid samples, glial cell types were sparsely represented owing to the early developmental stages of the samples (fig. S12A). We identified human genes that had 1:1 orthologs to the mouse genes in each module and asked whether the modules were conserved, using two metrics: whether the orthologous genes were also expressed in the corresponding cell type in the human datasets and whether the expression of the genes in each module covaried across single cells (as estimated from correlation), reflecting the degree of "modularity" of these mouse gene programs in humans.

The expression of each module was largely conserved in all human datasets, with different modules showing distinct levels of conservation of expression in each dataset (Fig. 4A). Some modules—such as PN1, PN2, and PN5—displayed high levels of conservation of expression (with at least 75% of the genes in these modules being expressed by at least 5% of cells in the corresponding associated cell type) across all datasets. The proportions of the genes expressed in the corresponding cell types in human tissues were generally lower than in mouse tissues (fig. S12B).

We further calculated whether the covariation of expression of the genes in each module (their "modularity") was also comparable in humans. To do so, for each module and each dataset we calculated the average pairwise expression correlation coefficient between the genes in a given module and compared it with a module-specific null-distribution based on random gene sets with similar expression levels, to calculate both a *P* value for the correlation of our modules and a normalized correlation coefficient. Out of 14 modules, eight showed greater intramodule correlation than that of a comparable random gene set in the adult human brain dataset from Hodge *et al.* (*29*) (Fig. 4B). Correlation also increased with the age of the human samples across brain regions of the BrainSpan dataset (Fig. 4, C to E, and fig. S13) (*8*). As a control, we used the same approach to calculate the expression and modularity of each gene module in nonassociated cell types. We found that the modularity was decreased in nonassociated cell types (fig. S12, D and E), reflected by both the proportion of comparisons with significant correlation and by the strength of the significant correlations, suggesting that our modules reflect cell type–specific effects.

Altogether, our results suggest that expression and modularity of most gene modules in the mouse are conserved in human brain tissue, pointing at potential shared functions and further suggesting that processes identified as affected in our Perturb-Seq experiments are relevant to biological processes that may be developmentally regulated in the human brain.

Mouse Perturb-Seq results are correlated with expression changes in ASD patient brain tissues

Last, we explored whether the effects observed in mouse Perturb-Seq may be similar to changes observed in post mortem brains of ASD patients. To this end, we compared our data with a snRNA-seq dataset of post mortem ASD brain

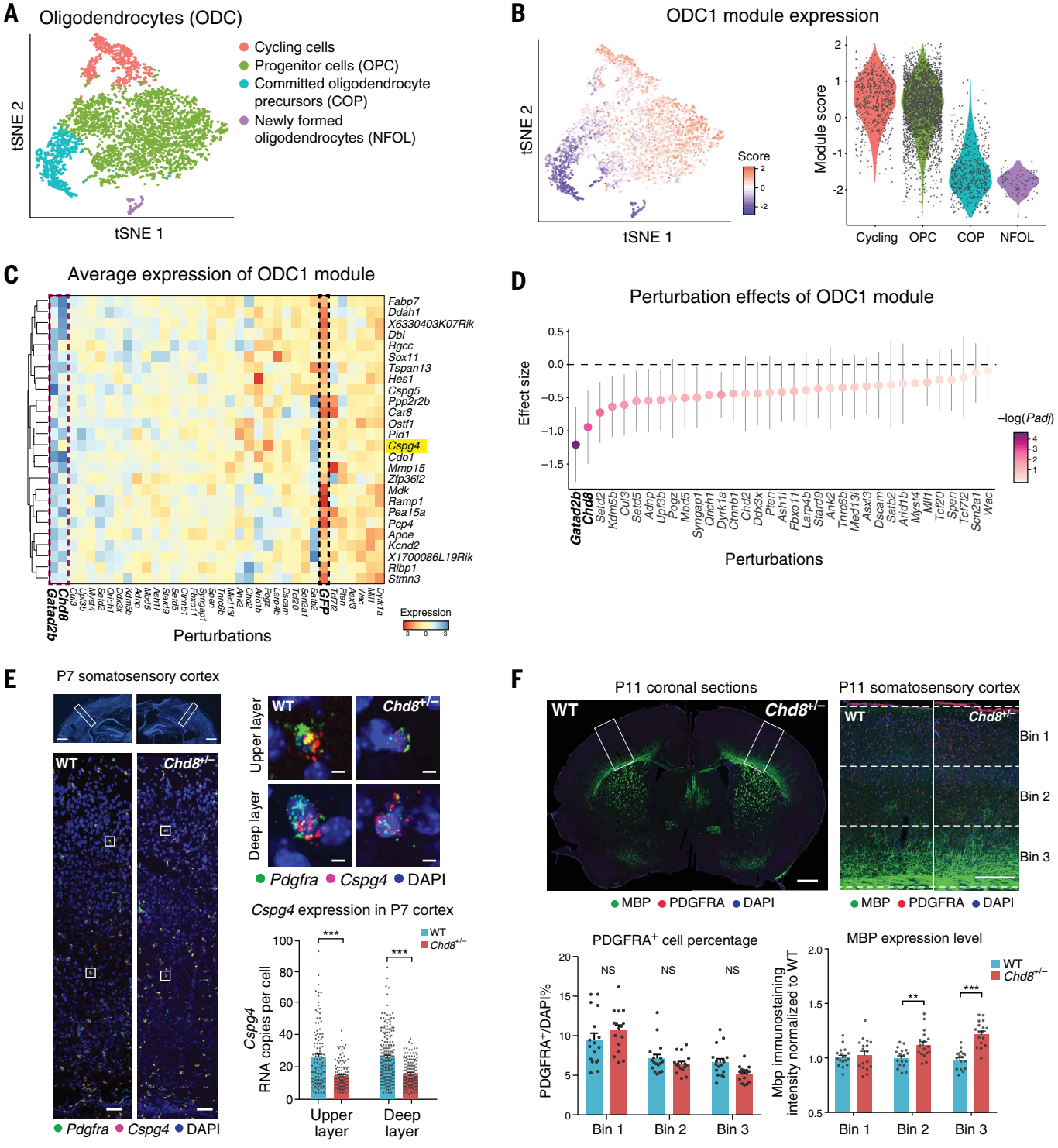


Fig. 3. Perturbation effect in oligodendrocytes and validation in the *Chd8*$^{+/-}$ mouse model. (**A**) *t*-SNE of oligodendrocyte subtypes from the Perturb-Seq data. (**B**) The ODC1 gene module expression score in (left) each celland (right) in each subcluster. (**C**) Average expression of genes in the ODC1 gene module (by row) in each perturbation group (by column), scaled by row. (**D**) Effect size of each perturbation on the ODC1 gene module compared with the control group. The perturbation effects of the different genes present a continuous gradient. Error bars represent 95% confidence intervals. (**E**) In situ hybridization for *Cspg4*, a gene in module ODC1 that is a known marker of oligodendrocyte precursor cells (OPC), in the somatosensory cortex of P7 *Chd8*$^{+/-}$ and wild-type littermates. The bottom images are higher magnifications of top images, and the right images are higher magnifications for each cell. (Right) Quantification of *Cspg4* expression in P7 cortex of *Chd8*$^{+/-}$ and wild-type littermates. Each dot represents the gene expression value from one cell; error bars represent standard error of the mean (n = 3 animals per genotype). Scale bars, 1000 μm (bottom left), 100 μm (top left), and 10 μm (right), respectively. (**F**) Immunohistochemistry for PDGFRA and MBP (markers for immature OPC and mature oligodendrocytes, respectively), PDGFRA^{+} cell counts, and distribution of MBP expression, in the somatosensory cortex of P11 *Chd8*$^{+/-}$ animals and wild-type littermates. Scale bars, 1000 μm (left) and 250 μm (right), respectively.

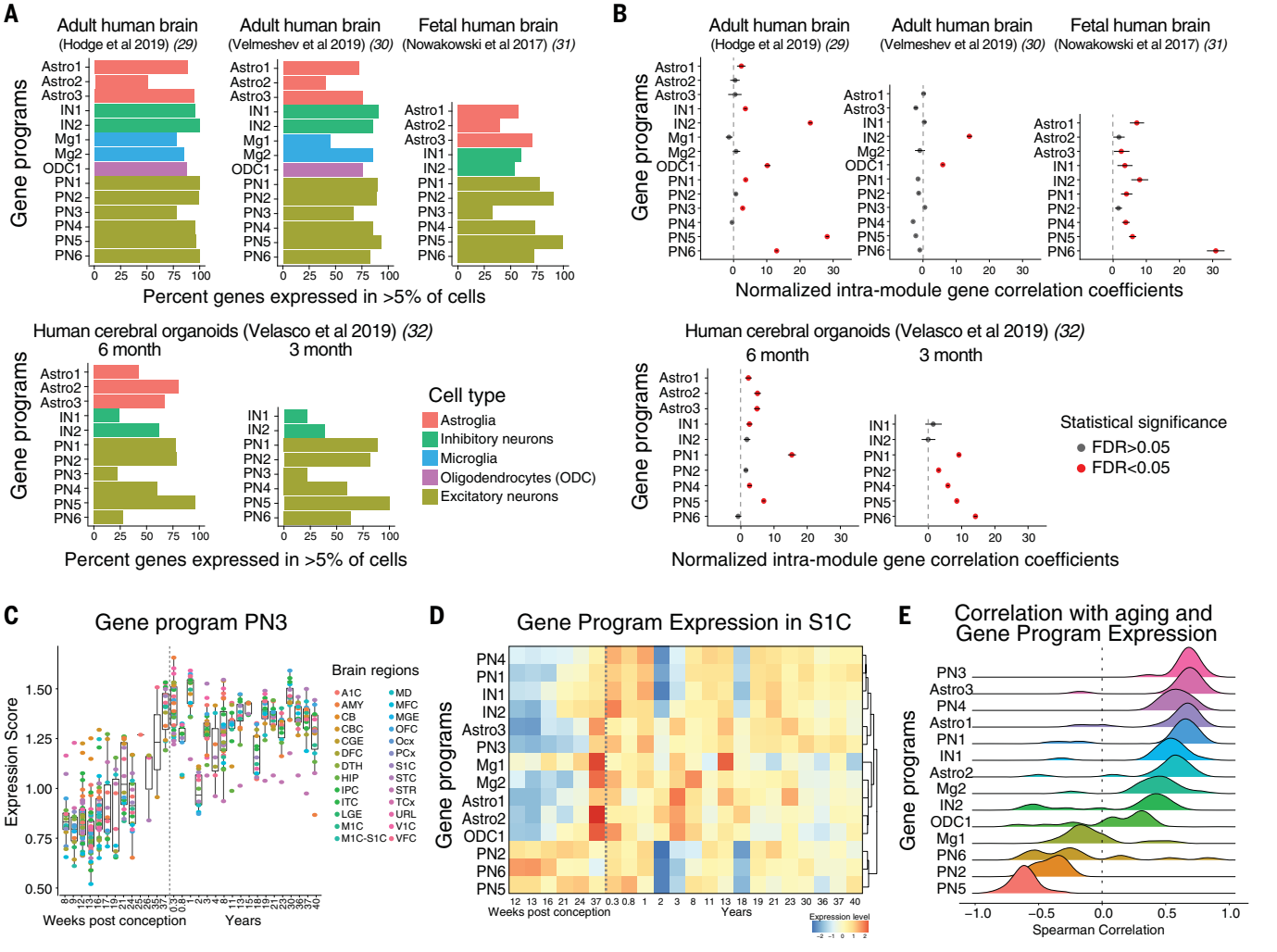

A
Adult human brain (Hodge et al 2019) (29)
Adult human brain (Velmeshev et al 2019) (30)
Fetal human brain (Nowakowski et al 2017) (31)
Gene programs
Astro1
Astro2
Astro3
IN1
IN2
Mg1
Mg2
ODC1
PN1
PN2
PN3
PN4
PN5
PN6
0 25 50 75 100
Percent genes expressed in >5% of cells
B
Adult human brain (Hodge et al 2019) (29)
Adult human brain (Velmeshev et al 2019) (30)
Fetal human brain (Nowakowski et al 2017) (31)
Gene programs
0 10 20 30
Normalized intra-module gene correlation coefficients
Human cerebral organoids (Velasco et al 2019) (32)
6 month
3 month
Gene programs
0 25 50 75 100
Percent genes expressed in >5% of cells
Cell type
Astroglia
Inhibitory neurons
Microglia
Oligodendrocytes (ODC)
Excitatory neurons
Human cerebral organoids (Velasco et al 2019) (32)
6 month
3 month
Gene programs
0 10 20 30
Normalized intra-module gene correlation coefficients
Statistical significance
FDR>0.05
FDR<0.05
C
Gene program PN3
Expression Score
1.50
1.25
1.00
0.75
0.50
Weeks post conception
Years
Brain regions
A1C
AMY
CB
CBC
CGE
DFC
DTH
HIP
IPC
ITC
LGE
M1C
M1C-S1C
MD
MFC
MGE
OFC
Ocx
PCx
S1C
STC
STR
TCx
URL
V1C
VFC
D
Gene Program Expression in S1C
Gene programs
PN4
PN1
IN1
IN2
Astro3
PN3
Mg1
Mg2
Astro1
Astro2
ODC1
PN2
PN6
PN5
12 13 16 21 24 37 0.3 0.8 1 2 3 8 11 13 18 19 21 23 30 36 37 40
Weeks post conception
Years
Expression level
E
Correlation with aging and Gene Program Expression
Gene programs
PN3
Astro3
PN4
Astro1
PN1
IN1
Astro2
Mg2
IN2
ODC1
Mg1
PN6
PN2
PN5
−1.0 −0.5 0.0 0.5 1.0
Spearman Correlation
F
ASD patient cortical single-cell analysis (Velmeshev et al) (30)
Cell type-specific DE gene lists
Test the DE genes in the Perturb-Seq data
Human ASD DE gene expression (Velmeshev et al) (30) in ASD risk gene Perturb-Seq
Velmeshev et al (30) DE gene expression changes across perturbation groups
DE genes in Velmeshev et al
SST
AP1S2
CNR1
RELN
CRIM1
LRRC3B
SAT2
CHCHD2
SPCS1
NRN1
RAB33A
PWP1
HDHD3
STK32A
Patient brain
Adnp Ank2 Ankrd11 Ash1l Asxl3 Chd2 Chd8 Ctnnb1 Cul3 Ddx3x Dscam Dyrk1a Fbxo11 Gatad2b Kdm5b Larp4b Mbd5 Med13l Mll1 Myst4 Pogz Pten Qrich1 Satb2 Scn2a1 Setd2 Setd5 Spen Stard9 Syngap1 Tcf20 Tcf7l2 Tmcb Upf3b Wac
Perturbations
logFC
-1.0 0.0 1.0
-log(Pvalue)
3 6 9 12
Inhibitory neurons
Excitatory neurons
Oligodendrocyte (ODC)
-1.0 -0.5 0.0 0.5 1.0
logFC(ASD-control)
Result from Velshemev et al
Result from Perturb-Seq
Significantly altered across 35 perturbations in Perturb-Seq
FDR < 0.1
FDR > 0.1

Fig. 4. Cell type–specific gene modules from Perturb-Seq are conserved in developing human brains. (**A**) Percent of genes with a human ortholog expressed in >5% of cells of the associated cell type in scRNA-seq datasets from the human brain or human brain organoids. (**B**) Normalized average pairwise correlation of gene expression within each gene module in the human brain or human brain organoids. Correlation values were normalized to the mean correlation from the background distribution and divided by the standard deviation of the background distribution. Correlations are shown for modules with at least four genes after filtering out genes expressed in <5% of cells. Bars represent 95% confidence intervals. Red color represents statistical significance (FDR < 0.05). (**C**) Expression of module PN3 over developmental time in human brain tissues across regions (BrainSpan data). (**D**) Expression of each module over developmental time in human primary somatosensory cortex S1C (BrainSpan). (**E**) Distribution of the Spearman correlation of module expression with age in human brain data over various brain regions (BrainSpan). (**F**) Differential gene expression analysis of human prefrontal cortical samples from ASD donors and controls. (Left) Expression of DE genes across cell types (color bars) from Velmeshev *et al*. (*30*) (rows) in the Perturb-Seq data across a panel of ASD/ND risk genes (columns). (Right) DE gene expression changes in Perturb-Seq data (black dots; each dot indicates an ASD/ND risk gene perturbation) compared with DE values for the 14 genes found to be DE in ASD patients in the Velmeshev *et al*. dataset (*30*) (FDR < 0.2) (red dots). The two highlighted genes, *SST* and *NRN1*, showed decreased expression in the Perturb-Seq data (FDR < 0.1), which is consistent with the ASD patient dataset.

samples (*30*) and bulk RNA-seq of post mortem psychiatric disorder brain samples from the PsychEncode project (*33*).

Using a dataset of snRNA-seq profiles from 15 ASD donors and 16 controls (*30*), we defined differentially expressed (DE) genes in each cell type using a statistically conservative pseudobulk-based analysis with DESeq2 (*34*, *35*), correcting for age, sex, and patient-to-patient variability. We identified genes that were differentially expressed between patients and controls in at least one of three major cell types (inhibitory neurons, excitatory neurons, or oligodendrocytes) with FDR < 0.2 and selected those that have 1:1 orthologs in mice, resulting in 14 genes (Fig. 4F and table S10).

We then compared these 14 genes with our Perturb-Seq data and asked whether these ASD-patient DE genes were also affected by the 35 ASD risk gene perturbations in our dataset. We aggregated the effects of all 35 perturbations and asked whether the aggregated gene expression changes agreed more strongly with the gene expression changes in the ASD patient data than would be expected by chance. For each ASD patient DE gene, we took its mouse ortholog and calculated the median fold change of expression (logFC) over all perturbations in the Perturb-Seq data. We then compared this logFC with the corresponding logFC in the ASD patient data and generated an agreement score for each gene, defined as a high median logFC and a similar direction of change as in the human data. We binned genes by their expression and compared each ASD patient DE gene with others in the same bin to extract *P* values (with FDR correction). From this analysis, we identified two genes, *SST* in interneurons and *NRN1* in excitatory neurons, both of which showed decreased expression in ASD patients and were likewise significantly decreased in expression across our panel of perturbations (FDR < 0.1), albeit with different effect sizes (Fig. 4F and table S10). This indicates that despite the different developmental stages, high clinical heterogeneity in ASD, and patient genetic diversity, genes and cell types can be identified as affected in both our analyses and in studies of human patient tissue.

We also analyzed the 14 gene modules reported in the PsychEncode study of 700 bulk RNA-seq samples of human cortex from a panel of psychiatric disorders (*33*). Of the 14 modules previously reported to be altered in the ASD patients in the PsychEncode analysis, six were also significantly affected across eight of our ASD/ND risk gene perturbations (fig. S14). Although these analyses are limited by the relatively few available datasets of ASD patient brain samples, they suggest that our Perturb-Seq experiments can identify gene program abnormalities seen in human ASD patients.

Discussion

In vivo Perturb-Seq can serve as a scalable tool for systems genetic studies of large gene panels to reveal their cell-intrinsic functions at single-cell resolution in complex tissues. In this work, we demonstrated the application of in vivo Perturb-Seq to ASD/ND risk genes in the developing brain; this method can be applied across diverse diseases and tissues.

ASD/ND affects brain function profoundly, but its cellular and molecular substrates are not yet defined. The large number of highly penetrant de novo risk genes implicated through human genetic studies offers an entry point to identify the cell types, developmental events, and mechanisms underlying ASD/ND. However, this requires scalable methods to define the function of risk-associated genes with cell-type specificity. Using Perturb-Seq to functionally test large gene sets in the developing embryo, we observed gene expression changes linked to ASD/ND genes in different cell types and processes. Within the power of the analysis that can be achieved with the number of cells that can be reasonably sequenced, we found that some recurrent modules are affected across more than one ASD/ND risk gene perturbation. It is likely that this represents an underestimation of the number of convergent modules across perturbations that might be revealed by larger-scale experiments using greater numbers of cells.

We were particularly interested in validating the observed effects of *Ank2* perturbation because of its known roles in the brain. *Ank2* encodes an ankyrin protein and is expressed broadly in excitatory and inhibitory neurons as well as glial cells in the brain (*22*). Ankyrin homologs interact with ion channels in many neuronal types, and *Ankyrin-G* has been shown to stabilize GABAergic synapses (*36*). The roles of *Ank2* in the brain have largely been studied in the context of excitatory neurons. *Ank2* loss of function results in hypoplasia of the corpus callosum and pyramidal tract, and optic nerve degeneration (*23*)—suggesting that it is required in the maintenance of premyelinated axons in excitatory neurons in early neurodevelopment. *Ank2* mutants showed misregulation of intracellular calcium homeostasis and calcium channel expression in excitatory neurons (*24*, *25*), as well as increased axonal branching and ectopic connectivity (*26*). Our Perturb-Seq data suggests an additional role of *Ank2* in interneurons expressing the *Ndnf* gene along with its known roles in excitatory neurons.

In addition to neurons, oligodendrocytes and astrocytes were also affected by several perturbations. Oligodendrocytes modulate and consolidate neural circuit refinement, and abnormal maturation of oligodendrocytes may be linked to long-lasting changes in neural wiring and brain function (*37*). One of the risk genes, *Chd8*, encodes a protein that binds directly to β-catenin to recruit histone proteins and negatively regulates the Wnt signaling pathway, a critical regulator of neural progenitor proliferation and differentiation in the forebrain (*38*–*41*). Our results showed that *Chd8* modulates gene modules for oligodendrocyte differentiation and maturation, which is consistent with previously reported chromatin immunoprecipitation–sequencing (ChIP-Seq) results showing that CHD8 interacts directly with OPC maturation genes at perinatal stages of development (*27*, *42*).

Although we focused on the perinatal neocortex in this study, in vivo Perturb-Seq can be applied to study gene functions systematically across other tissues and developmental ages, to reveal tissue-specific as well as broadly distributed gene functions. This approach can uncover both the impact of individual

disease-associated genes and of combinations of genes and the overall set of processes that they affect. Our findings underscore the importance of using single-cell profiles as a rich, comprehensive, and interpretable phenotypic readout. With advances in other single-cell profiling approaches [such as single-cell assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (*43*), single-cell multi-omics (*44*), and spatial genomics (*45*, *46*)], we expect in vivo Perturb-Seq to be coupled in the near future with diverse readouts to better define the function of disease risk–associated variants, from molecular mechanisms to non–cell autonomous effects in tissues. Spatial transcriptomics in particular should be well suited for in vivo Perturb-Seq and should help uncover non–cell autonomous effects of perturbations. In vivo Perturb-Seq can enable discoveries of pathways and cell types affected in heterogeneous genetic pathologies, directing downstream studies and informing the development of refined models for genetic disorders as we move from genetic variants to function.

Methods summary

In vivo Perturb-Seq experiment

The backbone plasmid contains antiparallel cassettes of two gRNAs (table S5) under mouse U6 and human U6 promoters and the EF1a promoter to express puromycin, BFP, and a polyadenylated barcode specific to each perturbation. Cloning and lentiviral packaging of the 38 vectors were done individually.

All animal experiments were performed according to protocols approved by the Institutional Animal Care and Use Committees (IACUC) of Harvard University and of the Broad Institute of MIT and Harvard. In utero lentiviral injection into the lateral ventricles was performed at E12.5 in Cas9 transgenic mice (*13*) (4 to 6 months old, Jax 026179), and each single-cell library was made by combining the BFP^+ cells from one to three litters (4 to 20 animals) of P7 animals harvested on the same day. Tissue dissociation was performed with the Papain Dissociation kit (Worthington, LK003152). The FACS-purified cells were sorted into cold Hibernate A/B27 medium and subjected to single-cell RNA sequencing library preparation. Our analysis comprises 17 independent libraries of Perturb-Seq cells.

Single-cell RNA-seq libraries were created by using the Chromium Single Cell 3′ Solution v2 kit (10x Genomics) following the manufacturer's protocol. Each library was sequenced with Illumina NextSeq high-output 75-cycle kit with sequencing saturation above 70%. Dial-out polymerase chain reaction (PCR) was performed to extract the perturbation barcode in each cell.

We identified perturbation barcodes by use of two complementary methods. We first used the dial-out sequences to create a cell-by-perturbation UMI (unique molecular identifier) count matrix by means of a modification of from the original Perturb-Seq work (*11*). In addition, we extracted barcode sequences from the 10x Genomics Cell Ranger bam file. Reads were then assigned to the perturbation to which they mapped best. Cell barcodes and UMIs were extracted, and a cell-by-perturbation UMI count matrix was created. We then only kept cells for which either (i) the assigned 10x and dialout perturbations agreed or (ii) the cell was assigned to a perturbation by one method but not assigned to any perturbation in the other.

Perturb-Seq analysis

UMI count data was loaded into R and processed by using the Seurat v 2.2 package (*47*). Clusters were assigned to cell types on the basis of marker genes from the literature, https://mousebrain.org (*15*), and https://DropViz.org (*22*). We focused only on cells of five key types (projection neurons, inhibitory neurons, oligodendrocytes, microglia and macrophages, and astroglia) and removed the rest.

WGCNA and STM were performed for each cell cluster according to the published pipelines (*20*, *21*). Linear regression was used to test the relationship between perturbations and WGCNA gene scores, correcting for batch and number of genes.

RNA in situ hybridization and immunohistochemistry

Multiplexed RNAscope fluorescent in situ hybridization and immunohistochemistry was performed on fixed-frozen tissue. Probes against the following mRNAs were used: *Pdgfra*, *Cspg4*, and *Fezf2* (ACDBio). The antibodies and dilutions were Mouse anti-NeuN antibody (mab377, 1:500; Millipore), Mouse anti-GS antibody (mab302, 1:500; Millipore), Goat anti-Pdgfra antibody (AF1062, 1:200; R&D System), Rabbit Iba1 antibody (019-19741, 1:400; Wako), Chicken anti-GFP antibody (ab16901, 1:500; Millipore), Mouse anti-Satb2 (ab51502, 1:50; Abcam), Rat anti-Ctip2 (ab18465, 1:100, Abcam), Rabbit anti-Sox6 (ab30455, 1:500; Abcam), and Rat anti-Mbp (mab386, 1:100; Millipore). We double-blinded the staining, imaging, and quantifications.

Analysis of human snRNA-seq or scRNA-seq data

For each single cell and nucleus human dataset, the UMI count matrix and metadata were downloaded and processed with Seurat to create Seurat objects. Cell types were extracted from the metadata and combined into more general cell types, namely Microglia, Astroglia (including Radial Glia), Inhibitory neurons, Excitatory neurons, Oligodendrocytes, and other. For differential expression analysis for data from Velmeshev *et al.* (*30*), we removed data from all individuals of <12 years of age and separated prefrontal cortex (PFC) and anterior cingulate cortex (ACC) regions. For each cell type in each region, a pseudobulk profile was constructed, and genes expressed in <5% of cells or with <10 reads were removed. DESeq2 v 1.20.0 (*34*) was then used to perform differential expression analysis between the ASD patients and the controls, correcting for sex and age. We then extracted all genes with 1:1 mouse orthologs (BioMart) and calculated FDR-corrected P values on these genes for both the ACC and PFC. Only analysis on the PFC yielded significant hits, which are presented in Fig. 4F.

To compare these results with the Perturb-Seq data, for each human DE gene, an agreement score was calculated by taking the absolute value of its mouse orthologs' median logFC over all perturbations (calculated with Limma) and giving it a positive sign if its direction agreed with that of the human data, and a negative sign otherwise. Last, genes were binned by expression, and P values were calculated for each gene by comparing the agreement scores with other genes in the same bin.

Detailed procedures for the experiments and data analyses are described in the supplementary materials.

REFERENCE AND NOTES

1. 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
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. L. Jostins *et al.*, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012). doi: 10.1038/nature11582; pmid: 23128233
4. F. K. Satterstrom *et al.*, Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *Cell* **180**, P568 (2018). doi: 10.1016/j.cell.2019.12.036
5. S. J. Sanders *et al.*, De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012). doi: 10.1038/nature10945; pmid: 22495306
6. J. A. Chen, O. Peñagarikano, T. G. Belgard, V. Swarup, D. H. Geschwind, The emerging picture of autism spectrum disorder: Genetics and pathology. *Annu. Rev. Pathol.* **10**, 111–144 (2015). doi: 10.1146/annurev-pathol-012414-040405; pmid: 25621659
7. C. Mullins, G. Fishell, R. W. Tsien, Unifying views of autism spectrum disorders: A consideration of autoregulatory feedback loops. *Neuron* **89**, 1131–1156 (2016). doi: 10.1016/j.neuron.2016.02.017; pmid: 26985722
8. J. A. Miller *et al.*, Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229
9. Data of 9k brain cells from an E18 mouse, 10x Genomics; https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.1.0/neuron_9k.
10. B. Adamson *et al.*, A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882.e21 (2016). doi: 10.1016/j.cell.2016.11.048; pmid: 27984733
11. A. Dixit *et al.*, Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016). doi: 10.1016/j.cell.2016.11.038; pmid: 27984732

12. D. A. Jaitin *et al*., Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896.e15 (2016). doi: 10.1016/j.cell.2016.11.039 pmid: 27984734
13. R. J. Platt *et al*., CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014). doi: 10.1016/j.cell.2014.09.014; pmid: 25263330
14. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008). doi: 10.1088/1742-5468/2008/10/P10008
15. A. Zeisel *et al*., Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018). doi: 10.1016/j.cell.2018.06.021; pmid: 30096314
16. S. Mancinelli, S. Lodato, Decoding neuronal diversity in the developing cerebral cortex: From single cells to functional networks. *Curr. Opin. Neurobiol.* **53**, 146–155 (2018). doi: 10.1016/j.conb.2018.08.001; pmid: 30165269
17. A. L. Haber *et al*., A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017). doi: 10.1038/nature24489; pmid: 29144463
18. B. Duan *et al*., Model-based understanding of single-cell CRISPR screening. *Nat. Commun.* **10**, 2233 (2019). doi: 10.1038/s41467-019-10216-x; pmid: 31110232
19. The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47** (D1), D330–D338 (2019). doi: 10.1093/nar/gky1055; pmid: 30395331
20. M. Roberts, B. Stewart, D. Tingley, stm: R package for structural topic models. *J. Stat. Softw.* **91**, 1–40 (2019). doi: 10.18637/jss.v091.i02
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. A. Saunders *et al*., Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16 (2018). doi: 10.1016/j.cell.2018.07.028; pmid: 30096299
23. P. Scotland, D. Zhou, H. Benveniste, V. Bennett, Nervous system defects of AnkyrinB$^{-/-}$ mice suggest functional overlap between the cell adhesion molecule L1 and 440-kD AnkyrinB in premyelinated axons. *J. Cell Biol.* **143**, 1305–1315 (1998). doi: 10.1083/jcb.143.5.1305; pmid: 9832558
24. S. Tuvia, M. Buhusi, L. Davis, M. Reedy, V. Bennett, Ankyrin-B is required for intracellular sorting of structurally diverse Ca^{2+} homeostasis proteins. *J. Cell Biol.* **147**, 995–1008 (1999). doi: 10.1083/jcb.147.5.995; pmid: 10579720
25. C. F. Kline, J. Scott, J. Curran, T. J. Hund, P. J. Mohler, Ankyrin-B regulates Cav2.1 and Cav2.2 channel expression and targeting. *J. Biol. Chem.* **289**, 5285–5295 (2014). doi: 10.1074/jbc.M113.523639; pmid: 24394417
26. R. Yang *et al*., *ANK2* autism mutation targeting giant ankyrin-B promotes axon branching and ectopic connectivity. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15262–15271 (2019). doi: 10.1073/pnas.1904348116; pmid: 31285321
27. C. Marie *et al*., Oligodendrocyte precursor survival and differentiation requires chromatin remodeling by Chd7 and Chd8. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8246–E8255 (2018). doi: 10.1073/pnas.1802620115; pmid: 30108144
28. M. Nishiyama *et al*., Early embryonic death in mice lacking the beta-catenin-binding protein Duplin. *Mol. Cell. Biol.* **24**, 8386–8394 (2004). doi: 10.1128/MCB.24.19.8386-8394.2004; pmid: 15367660
29. R. D. Hodge *et al*., Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019). doi: 10.1038/s41586-019-1506-7; pmid: 31435019
30. D. Velmeshev *et al*., Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019). doi: 10.1126/science.aav8130; pmid: 31097668
31. 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
32. S. Velasco *et al*., Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019). doi: 10.1038/s41586-019-1289-x; pmid: 31168097
33. 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
34. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
35. A. T. L. Lun, J. C. Marioni, Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. *Biostatistics* **18**, 451–464 (2017). doi: 10.1093/biostatistics/kxw055; pmid: 28334062
36. W. C. Tseng, P. M. Jenkins, M. Tanaka, R. Mooney, V. Bennett, Giant ankyrin-G stabilizes somatodendritic GABAergic synapses through opposing endocytosis of GABAA receptors. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1214–1219 (2015). doi: 10.1073/pnas.1417989112; pmid: 25552561
37. K. K. Bercury, W. B. Macklin, Dynamics and mechanisms of CNS myelination. *Dev. Cell* **32**, 447–458 (2015). doi: 10.1016/j.devcel.2015.01.016; pmid: 25710531
38. R. J. Platt *et al*., Chd8 mutation leads to autistic-like behaviors and impaired striatal circuits. *Cell Rep.* **19**, 335–350 (2017). doi: 10.1016/j.celrep.2017.03.052; pmid: 28402856
39. Y. Katayama *et al*., CHD8 haploinsufficiency results in autistic-like phenotypes in mice. *Nature* **537**, 675–679 (2016). doi: 10.1038/nature19357; pmid: 27602517
40. O. Durak *et al*., Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. *Nat. Neurosci.* **19**, 1477–1488 (2016). doi: 10.1038/nn.4400; pmid: 27694995
41. I. Sakamoto *et al*., A novel beta-catenin-binding protein inhibits β-catenin-dependent Tcf activation and axis formation. *J. Biol. Chem.* **275**, 32871–32878 (2000). doi: 10.1074/jbc.M004089200; pmid: 10921920
42. C. Zhao *et al*., Dual requirement of CHD8 for chromatin landscape establishment and histone methyltransferase recruitment to promote CNS myelination and repair. *Dev. Cell* **45**, 753–768.e8 (2018). doi: 10.1016/j.devcel.2018.05.022; pmid: 29920279
43. A. J. Rubin *et al*., Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361–376.e17 (2019). doi: 10.1016/j.cell.2018.11.022; pmid: 30580963
44. S. Bian *et al*., Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060–1063 (2018). doi: 10.1126/science.aao3791; pmid: 30498128
45. S. G. Rodriques *et al*., Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019). doi: 10.1126/science.aaw1219; pmid: 30923225
46. S. Simmons, seanken/ivPerturbSeq: ForPub. Zenodo (2020). doi: 10.5281/zenodo.4019534

ACKNOWLEDGMENTS

We thank C. Dulac, W. Gilbert, M. Meselson, and C. I. Bargmann for critical reading of our manuscript; L. Gaffney, A. Hupalowska, R. Macrae, J. Brown, S. Smith, as well as members of the Levin laboratory, Regev laboratory, Zhang laboratory, and Arlotta laboratory for technical and intellectual support. **Funding:** This work is supported by the Stanley Center for Psychiatric Research at the Broad Institute, NIH grants (U01MH115727, R01MH096066, and P50MH094271 Conte Center to P.A.); NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation and Harvard William F. Milton Grant (to X.J.); Stanley Center for Psychiatric Research at the Broad Institute (to J.Z.L.); The Klarman Cell Observatory, Howard Hughes Medical Insititue (HHMI) and an NHGRI Center for Cell Circuits CEGS grant (to A.R.); and NIH grants (1R01-HG009761, 1R01-MH110049, and 1DP1-HL141201), HHMI, New York Stem Cell and Mathers Foundations, the Poitras Center for Affective Disorders Research at MIT, the Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT, and J. and P. Poitras (to F.Z.). F.Z. is a New York Stem Cell Foundation–Robertson Investigator. **Author contributions:** X.J., A.R., F.Z., and P.A. conceived the project; X.J., A.G., and M.K. performed the experiments; X.J., S.K.S., and G.D. analyzed the data; A.S.S., V.J., and S.L. helped with cell subtype identification; L.N. assisted with sequencing library preparations; E.R. contributed human genetics interpretations; P.O. and N.C. assisted with molecular cloning; X.J., S.K.S., J.Z.L., A.R., F.Z., and P.A. wrote and revised the manuscript. **Competing interests:** P.A. is a SAB member in System 1 Biosciences and Foresite Labs and is a cofounder of FL60. A.R. is a cofounder of and equity holder in Celsius Therapeutics, Equity holder in Immunitas. Until 31 July 2020, A.R. was an SAB member of ThermoFisher Scientific, Syros Pharmaceuticlas, Asimov, and Neogene Therapeutics. From August 2020, A.R. is an employee of Genetech. F.Z. is a cofounder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences. X.J., S.K.S., A.R., F.Z., and P.A. are co-inventors on in vivo Perturb-Seq and CRISPR inventions filed by the Broad Institute relating to the work in this manuscript. **Data and materials availability:** Data generated for this study are available through the Gene Expression Omnibus (accession no. GSE157977) as well as the Broad single cell portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1184). The analysis pipeline is deposited on GitHub repository (https://github.com/klarman-cell-observatory/ivPerturbSeq) and Zenodo (*46*). All other data are available in the manuscript or the supplementary materials. Material used in this study will be available through Addgene under the Uniform Biological Material Transfer Agreement.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1057/suppl/DC1
Materials and Methods
Figs. S1 to S14
References (*47*–*54*)
Tables S1 to S10
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

26 September 2019; resubmitted 24 May 2020
Accepted 9 October 2020
10.1126/science.aaz6063

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity

Ellen Shrock*, Eric Fujimura*, Tomasz Kula†, Richard T. Timms†, I-Hsiu Lee, Yumei Leng, Matthew L. Robinson, Brandon M. Sie, Mamie Z. Li, Yuezhou Chen, Jennifer Logue, Adam Zuiani, Denise McCulloch, Felipe J. N. Lelis, Stephanie Henson, Daniel R. Monaco, Meghan Travers, Shaghayegh Habibi, William A. Clarke, Patrizio Caturegli, Oliver Laeyendecker, Alicja Piechocka-Trocha, Jonathan Z. Li, Ashok Khatri, Helen Y. Chu, MGH COVID-19 Collection & Processing Team, Alexandra-Chloé Villani, Kyle Kays, Marcia B. Goldberg, Nir Hacohen, Michael R. Filbin, Xu G. Yu, Bruce D. Walker, Duane R. Wesemann, H. Benjamin Larman, James A. Lederer, Stephen J. Elledge‡

INTRODUCTION: A systematic characterization of the humoral response to severe acute respiratory system coronavirus 2 (SARS-CoV-2) epitopes has yet to be performed. This analysis is important for understanding the immunogenicity of the viral proteome and the basis for cross-reactivity with the common-cold coronaviruses.

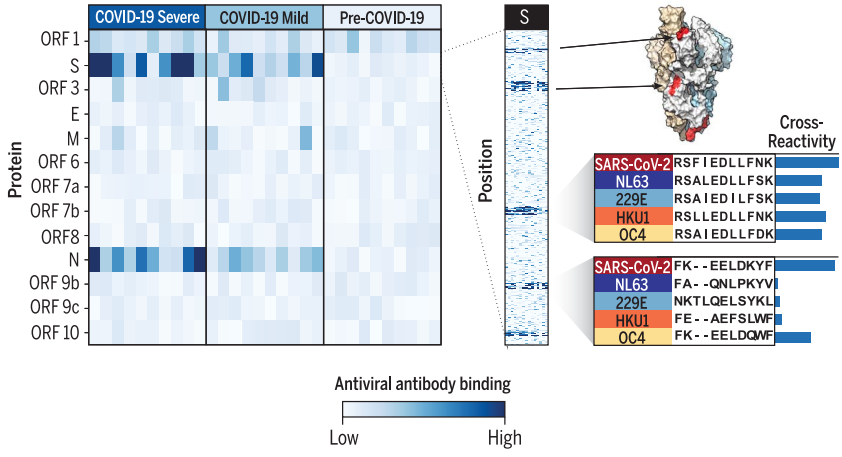
Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, is notable for its variable course, with some individuals remaining asymptomatic whereas others experience fever, respiratory distress, or even death. A comprehensive investigation of the antibody response in individuals with severe versus mild COVID-19—as well as an examination of past viral exposure history—is needed.

RATIONALE: An understanding of humoral responses to SARS-CoV-2 is critical for improving diagnostics and vaccines and gaining insight into variable clinical outcomes. To this end, we used VirScan, a high-throughput method to analyze epitopes of antiviral antibodies in human sera. We supplemented the original VirScan library with additional libraries of peptides spanning the proteomes of SARS-CoV-2 and all other human coronaviruses. These libraries enabled us to precisely map epitope locations and investigate cross-reactivity between SARS-CoV-2 and other coronavirus strains. The original VirScan library allowed us to simultaneously investigate antibody responses to prior infections and viral exposure history.

RESULTS: We screened sera from 232 COVID-19 patients and 190 pre–COVID-19 era controls against the original VirScan and supplemental coronavirus libraries, assaying more than 10^8 antibody repertoire–peptide interactions. We identified epitopes ranging from "private" (recognized by antibodies in only a small number of individuals) to "public" (recognized by antibodies in many individuals) and detected SARS-CoV-2–specific epitopes as well as those that cross-react with common-cold coronaviruses. Several of these cross-reacting antibodies are present in pre–COVID-19 era samples. We developed a machine learning model that predicted SARS-CoV-2 exposure history with 99% sensitivity and 98% specificity from VirScan data. We used the most discriminatory SARS-CoV-2 peptides to produce a Luminex-based serological assay, which performed similarly to gold-standard enzyme-linked immunosorbent assays. We stratified the COVID-19 patient samples by disease severity and found that patients who had required hospitalization exhibited stronger and broader antibody responses to SARS-CoV-2 but weaker overall responses to past infections compared with those who did not need hospitalization. Further, the hospitalized group had higher seroprevalence rates for cytomegalovirus and herpes simplex virus 1. These findings may be influenced by differences in demographic compositions between the two groups, but they raise hypotheses that may be tested in future studies. Using alanine scanning mutagenesis, we precisely mapped 823 distinct epitopes across the entire SARS-CoV-2 proteome, 10 of which are likely targets of neutralizing antibodies. One cross-reactive antibody epitope in S2 has been previously suggested to be neutralizing and, as it exists in pre–COVID-19 era samples, could affect the severity of COVID-19.

CONCLUSION: We present a highly detailed view of the epitope landscape within the SARS-CoV-2 proteome. This knowledge may be used to produce diagnostics with improved specificity and can provide a stepping stone to the isolation and functional dissection of both neutralizing antibodies and antibodies that might exacerbate patient outcomes through antibody-dependent enhancement or immune distraction.

Our study reveals notable correlations between COVID-19 severity and both viral exposure history and overall strength of the antibody response to past infections. These findings are likely influenced by demographic covariates, but they generate hypotheses that may be tested with larger patient cohorts matched for age, gender, race, and other demographic variables. ■



SARS-CoV-2 epitope mapping. VirScan detects antibodies against SARS-CoV-2 in COVID-19 patients with severe and mild disease. Heatmap color represents the strength of the antibody response in each sample (columns) to each protein (rows, left) or peptide (rows, right). VirScan reveals the precise positions of epitopes, which can be mapped onto the structure of the spike protein (S). Examination of SARS-CoV-2 and seasonal coronavirus sequence conservation explains epitope cross-reactivity. A, Ala; D, Asp; E, Glu; F, Phe; I, Ile; K, Lys; L, Leu; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Author affiliations and MGH COVID-19 Collection & Processing Team participants and affiliations are available in the full article online.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: selledge@genetics.med.harvard.edu

Cite this article as E. Shrock *et al.*, *Science* **370**, eabd4250 (2020). DOI: 10.1126/science.abd4250

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd4250

RESEARCH ARTICLE

CORONAVIRUS

Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity

Ellen Shrock[1,2]*, Eric Fujimura[1,2,3]*, Tomasz Kula[1,2]†, Richard T. Timms[1,2]†, I-Hsiu Lee[4], Yumei Leng[1,2], Matthew L. Robinson[5], Brandon M. Sie[1,2], Mamie Z. Li[1,2], Yuezhou Chen[6,7], Jennifer Logue[8], Adam Zuiani[6,7], Denise McCulloch[8], Felipe J. N. Lelis[6,7], Stephanie Henson[9], Daniel R. Monaco[9], Meghan Travers[6,7], Shaghayegh Habibi[6,7], William A. Clarke[10], Patrizio Caturegli[11], Oliver Laeyendecker[5,12], Alicja Piechocka-Trocha[7,13], Jonathan Z. Li[7,14], Ashok Khatri[15], Helen Y. Chu[8], MGH COVID-19 Collection & Processing Team[16]‡, Alexandra-Chloé Villani[17], Kyle Kays[18], Marcia B. Goldberg[19], Nir Hacohen[20], Michael R. Filbin[18], Xu G. Yu[7,14,16,21], Bruce D. Walker[7,13,22], Duane R. Wesemann[6,7], H. Benjamin Larman[9], James A. Lederer[23], Stephen J. Elledge[1,2,7]§

Understanding humoral responses to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is critical for improving diagnostics, therapeutics, and vaccines. Deep serological profiling of 232 coronavirus disease 2019 (COVID-19) patients and 190 pre–COVID-19 era controls using VirScan revealed more than 800 epitopes in the SARS-CoV-2 proteome, including 10 epitopes likely recognized by neutralizing antibodies. Preexisting antibodies in controls recognized SARS-CoV-2 ORF1, whereas only COVID-19 patient antibodies primarily recognized spike protein and nucleoprotein. A machine learning model trained on VirScan data predicted SARS-CoV-2 exposure history with 99% sensitivity and 98% specificity; a rapid Luminex-based diagnostic was developed from the most discriminatory SARS-CoV-2 peptides. Individuals with more severe COVID-19 exhibited stronger and broader SARS-CoV-2 responses, weaker antibody responses to prior infections, and higher incidence of cytomegalovirus and herpes simplex virus 1, possibly influenced by demographic covariates. Among hospitalized patients, males produce stronger SARS-CoV-2 antibody responses than females.

Coronaviruses constitute a large family of enveloped, positive-sense single-stranded RNA viruses that cause diseases in birds and mammals (*1*). Among the strains that infect humans are the alphacoronaviruses HCoV-229E and HCoV-NL63 and the betacoronaviruses HCoV-OC43 and HCoV-HKU1, which cause common colds (Fig. 1A). Three additional betacoronavirus species result in severe infections in humans: severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel coronavirus that emerged in late 2019 in Asia and quickly spread throughout the world (*2*). As of November 2020, SARS-CoV-2 has caused more than 50 million confirmed infections and nearly 1.3 million deaths (*3*).

The clinical course of coronavirus disease 19 (COVID-19)—the disease resulting from SARS-CoV-2 infection—is notable for its extreme variability: Some individuals remain entirely asymptomatic, whereas others experience fever, anosmia, diarrhea, severe respiratory distress, pneumonia, cardiac arrhythmia, blood clotting disorders, liver and kidney distress, enhanced cytokine release and, in a small percentage of cases, death (*4*). Therefore, understanding the factors that influence this spectrum of outcomes is an intense area of research. Disease severity is correlated with advanced age, sex, ethnicity, socioeconomic status, and comorbidities such as diabetes, cardiovascular disease, chronic lung disease, obesity, and reduced immune function (*4*). Additional relevant factors likely include the inoculum of virus at infection and individual genetic background and viral exposure history. The complex interplay of these elements also determines how individuals respond to therapies aimed at mitigating disease severity. Detailed knowledge of the immune response to SARS-CoV-2 could improve our understanding of diverse outcomes and inform the development of improved diagnostics, vaccines, and antibody-based therapies.

In this study, we used VirScan, a programmable phage-display immunoprecipitation and sequencing (PhIP-Seq) technology that we developed previously, to explore antiviral antibody responses across the human virome (*5*–*8*). Here we describe a detailed analysis of the humoral response in COVID-19 patients.

Results

Development of a VirScan library targeting human coronaviruses

Our existing VirScan phage-display platform is based on an oligonucleotide library encoding 56–amino acid (56-mer) peptides tiling every 28 amino acids across the proteomes of all known pathogenic human viruses (~400 species and strains) plus many bacterial proteins (*8*). To investigate the serological response to SARS-CoV-2 and other human coronaviruses (HCoVs), we supplemented this library with three additional sublibraries: Sublibrary 1 encodes a 56-mer peptide library tiling every 28 amino acids through each of the open reading frames (ORFs) expressed by the six HCoVs and three bat coronaviruses closely related to SARS-CoV-2; sublibrary 2 encodes 20-mer peptides tiling every 5 amino acids across the SARS-CoV-2 proteome, enabling more precise localization of epitopes; and sublibrary 3 encodes triple-alanine scanning mutants of the 56-mer peptides tiling across the SARS-CoV-2 proteome, enabling the mapping of epitope boundaries at amino acid resolution (Fig. 1, A to C, and table S1) (*9*, *10*).

We used VirScan (Fig. 1C) to profile the antibody repertoires of nine cohorts of individuals from multiple locations, including Baltimore, MD, Boston, MA, and Seattle, WA (tables S2 to S8). These cohorts comprised longitudinal samples from individuals enrolled in prospective studies of COVID-19 infection, cross-sectional samples from patients with active

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA. [2]Howard Hughes Medical Institute, Division of Genetics, Brigham and Women's Hospital, Program in Virology, Harvard Medical School, Boston, MA, USA. [3]Chemical Biology Program, Harvard University, Cambridge, MA, USA. [4]Center for Systems Biology, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [5]Division of Infectious Diseases, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [6]Division of Allergy and Immunology and Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [7]Massachusetts Consortium on Pathogen Readiness, Boston, MA, USA. [8]Department of Medicine, University of Washington, Seattle, WA, USA. [9]Institute for Cell Engineering, Immunology Division, Department of Pathology, Johns Hopkins University, Baltimore, MD, USA. [10]Division of Clinical Chemistry, Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, USA. [11]Division of Immunology, Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, USA. [12]Division of Intramural Research, NIAID, NIH, Baltimore, MD, USA. [13]Howard Hughes Medical Institute, Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [14]Infectious Disease Division, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [15]Endocrine Unit and Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [16]Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA. [17]Massachusetts General Hospital Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital Cancer Center, Department of Medicine, Harvard Medical School, Boston, MA, USA. [18]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, USA. [19]Center for Bacterial Pathogenesis, Division of Infectious Diseases, Department of Medicine and Microbiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [20]Massachusetts General Hospital Cancer Center, Department of Medicine, Harvard Medical School, Boston, MA, USA. [21]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [22]Centre for the AIDS Programme of Research in South Africa, Congella, South Africa. [23]Department of Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
*These authors contributed equally to this work. †These authors contributed equally to this work. ‡MGH COVID-19 Collection & Processing Team participants and affiliations appear at the end of this paper.
§Corresponding author. Email: selledge@genetics.med.harvard.edu

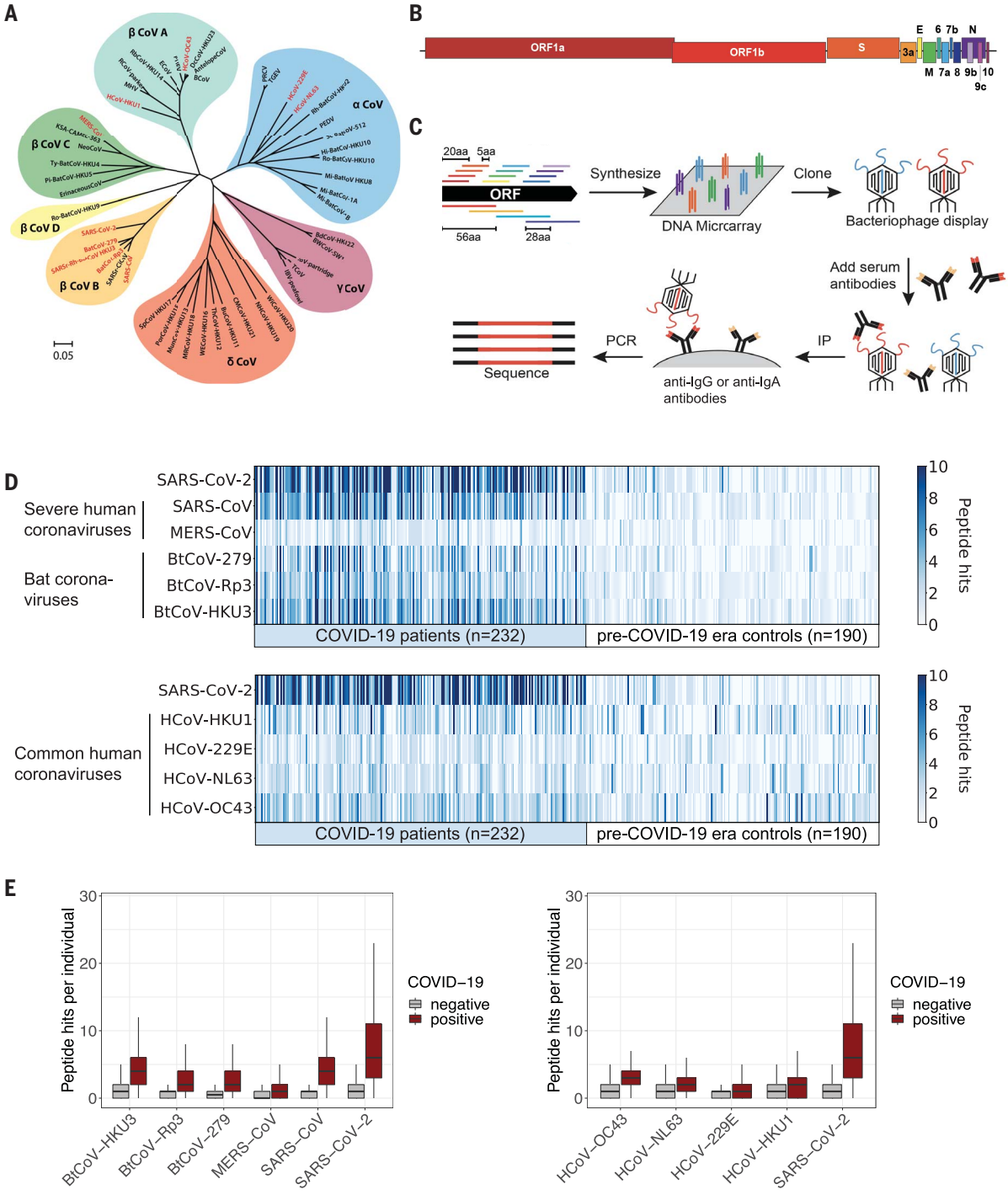


Fig. 1. VirScan detects the humoral response to SARS-CoV-2 in sera from COVID-19 patients. (**A**) Phylogeny tree of 50 coronavirus sequences (*32*) constructed using MEGA X (*33*, *34*). The scale bar indicates the estimated number of base substitutions per site (*35*). Coronaviruses included in the updated VirScan library are indicated in red. (**B**) Schematic representation of the ORFs encoded by the SARS-CoV-2 genome (*10*, *36*). (**C**) Overview of the VirScan procedure (*5*–*8*). The coronavirus oligonucleotide library includes 56-mer peptides tiling every 28 amino acids (aa) across the proteomes of 10 coronavirus strains and 20-mer peptides tiling every 5 amino acids across the SARS-CoV-2 proteome. Oligonucleotides were cloned into a T7 bacteriophage display vector and packaged into phage particles displaying the encoded peptides on their surface. The phage library was mixed with sera containing antibodies that bind to their cognate epitopes on the phage surface; bound phage were isolated by IP with either anti-IgG– or anti-IgA–coated magnetic beads. Lastly, PCR amplification and Illumina sequencing from the DNA of the bound phage revealed the peptides targeted by the serum antibodies. (**D**) Detection of antibodies targeting coronavirus epitopes by VirScan. Heatmaps depict the humoral response from COVID-19 patients (n = 232) and pre–COVID-19 era control samples (n = 190). Each column represents a sample from a distinct individual. The color intensity indicates the number of 56-mer peptides from the indicated coronaviruses significantly enriched by IgG antibodies in the serum sample. (**E**) Box plots illustrate the number of peptide hits from the indicated coronaviruses in COVID-19 patients and pre–COVID-19 era controls. The box indicates the interquartile range, with a line at the median. The whiskers span 1.5 times the interquartile range.

COVID-19 who were receiving treatment in either hospital or outpatient settings, and cross-sectional samples from convalescent individuals with a past history of COVID-19. Our cohorts also included a diverse set of control sera collected before the COVID-19 outbreak. We profiled the targets of IgG and IgA (immunoglobulins G and A) antibodies separately: IgG and IgA are the most abundant isotypes in blood, whereas IgA is the principal isotype secreted on mucosal surfaces, including the respiratory tract. Collectively, we analyzed ~550 samples in duplicate, in total assessing ~100 million potential antibody repertoire–peptide interactions.

Detection of SARS-CoV-2 seropositivity with VirScan

To measure immune responses to SARS-CoV-2, we compared VirScan profiles of serum samples from COVID-19 patients to those of controls obtained before the emergence of SARS-CoV-2 in 2019. These pre–COVID-19 era controls facilitate identification of (i) SARS-CoV-2 peptides encoding epitopes specific to COVID-19 patients and (ii) SARS-CoV-2 peptides encoding epitopes that are cross-reactive with antibodies developed in response to the ubiquitous common-cold HCoVs. Sera from COVID-19 patients exhibited much more SARS-CoV-2 reactivity than did sera from pre–COVID-19 era controls (Fig. 1, D and E). Some cross-reactivity toward SARS-CoV-2 peptides was observed in the pre–COVID-19 era samples, but this was expected because nearly all people have been exposed to one or more HCoVs (*11*).

COVID-19 patient sera also showed significant levels of cross-reactivity with the other highly pathogenic HCoVs, SARS-CoV and MERS-CoV, although less was observed against the more distantly related MERS-CoV. Extensive cross-reactivity was also observed against peptides derived from the three bat coronaviruses that share the greatest proportion of sequence identity with SARS-CoV-2 (Fig. 1, A, D, and E) (*9*). We know that these represent cross-reactivities because, given the low prevalence and circumscribed geographical location of SARS-CoV and MERS-CoV, none of the individuals in this study are likely to have encountered these viruses.

COVID-19 patient sera also exhibited a significantly higher level of reactivity to seasonal HCoV peptides than did sera from pre–COVID-19 era controls (Fig. 1, D and E). This could be due to the elicitation of novel antibodies that cross-react or to an anamnestic response that boosts B cell memory against HCoVs. The converse is not always true: Many pre–COVID-19 era samples exhibit strong recognition of seasonal HCoV peptides but little or no recognition of SARS-CoV-2 peptides (Fig. 1D). In some cases, the concentrations of antibodies against seasonal HCoVs may be below the threshold of detection in the pre–COVID-19 era samples.

Coronavirus proteins targeted by antibodies in COVID-19 patients

Analysis of SARS-CoV-2 proteins targeted by COVID-19 patient antibodies revealed that the primary responses to SARS-CoV-2 are reactive with peptides derived from spike protein (S) and nucleoprotein (N) (Fig. 2, A and B). Compared with sera from pre–COVID-19 era controls, COVID-19 patient sera exhibit significant differential recognition of these two proteins, indicating that this recognition is a result of antibody responses to SARS-CoV-2. Third-most frequently recognized is the replicase polyprotein ORF1, but unlike S and N, ORF1 is recognized to a similar extent by sera from COVID-19 patients and pre–COVID-19 era controls. This suggests that recognition of SARS-CoV-2 ORF1 is a result of cross-reactions from antibodies elicited by exposure to other pathogens, possibly HCoVs. Antibody responses to peptides from membrane glycoprotein (M), ORF3, and ORF9b were occasionally detected in COVID-19 patients.

We also analyzed longitudinal samples from 23 COVID-19 patients. Most patients displayed an antibody response to peptides derived from S or N in the second week after symptom onset, with many displaying an antibody response by the end of the first week (Fig. 2C). The relative strength and onset of the antibody response to S and N differed markedly between individuals, and the initial immune response showed no preference for S or N. The signal intensity of antibodies recognizing SARS-CoV-2 ORF1 epitopes did not increase over time, further suggesting that ORF1 antibodies likely represent a preexisting cross-reactive response.

Identification of immunogenic regions of SARS-CoV-2 proteins

To more precisely define the immunogenic regions of the SARS-CoV-2 proteome, we examined the specific 56-mer and 20-mer peptides detected by VirScan in COVID-19 patients compared with those in pre–COVID-19 era controls. An example IgG response from a single patient to SARS-CoV-2 S and N is shown in Fig. 3A. We observed strong concordance between the viral regions enriched by the 56-mer and 20-mer fragments, demonstrating the robustness of VirScan. In many cases, we observed recognition of overlapping 56-mer peptides, indicating an epitope in the common region.

Next, we compared the protein regions recognized by IgG and IgA across COVID-19 patients (Fig. 3B). We identified four regions each in S and N that are recurrently targeted by antibodies from >15% of COVID-19 patients, with additional regions recognized less frequently. Overall, IgG and IgA recognize the same protein regions with similar frequencies across the population. However, when IgG and IgA responses were compared within individuals, we observed considerable divergence (Fig. 3C): Many epitopes were recognized by only IgG, only IgA, or both IgG and IgA within an individual patient. Together, these data suggest that patients generate distinct IgG and IgA antibody responses to SARS-CoV-2, but the targeted regions are largely shared at the population level.

Machine learning guides the design of a Luminex assay for rapid COVID-19 diagnosis

To predict SARS-CoV-2 exposure history from VirScan data, we developed a gradient-boosting algorithm (XGBoost) that integrates both IgG and IgA data and predicts current or past COVID-19 with 99.1% sensitivity and 98.4% specificity (Fig. 4, A and B). We used Shapley Additive exPlanations (SHAP)—a method to compute the contribution of each feature of the data to the predictive model (*12*)—to identify peptides from SARS-CoV-2 S and N plus homologous peptides from SARS-CoV and BatCoV-HKU-3 and BatCoV-279 that were highly predictive of SARS-CoV-2 exposure (Fig. 4, C and D).

We leveraged these insights to develop a simple, rapid Luminex-based diagnostic for COVID-19. We chose 12 SARS-CoV-2 peptides predicted by VirScan data and the machine learning model to be highly indicative of SARS-CoV-2 exposure history (table S9). These SARS-CoV-2 peptides, two positive control peptides from rhinovirus A and Epstein-Barr virus (EBV) that are recognized in >80% of seropositive individuals by VirScan (*7*), and a negative control peptide from HIV-1 were coupled to Luminex beads (*13*). We tested 163 COVID-19 patient samples and 165 pre–COVID-19 era controls for IgG reactivity to the Luminex panel. We detected clear responses to SARS-CoV-2 peptides in COVID-19 patient samples but rarely in the pre–COVID-19 era controls (Fig. 4E). Using the Luminex data, we developed a logistic regression model that predicts COVID-19 infection history with 89.6% sensitivity and 95.2% specificity [area under the curve (AUC) = 0.97] (Fig. 4, F and G). A subset of COVID-19–positive samples (n = 107) was also examined with an in-house enzyme-linked immunosorbent assay (ELISA) using three SARS-CoV-2 antigens: N, S, and the S receptor-binding domain (RBD). Considering a sample to be positive if it scored above the 99% specificity threshold on any one of the three ELISA antigens, we determined that the sensitivity of the Luminex assay for this subset (88.8%) was similar to that of the ELISA (90.7%) (fig. S1). Among samples run on all three assays, VirScan significantly outperformed both the Luminex and ELISAs (fig. S1, A and C). Notably, our optimal model integrated only three SARS-CoV-2 peptides—residues 386 to 406 of N (N 386-406), residues 810 to 830 of S (S 810-830), and residues 1146 to 1166 of S (S 1146-1166)—which were also the

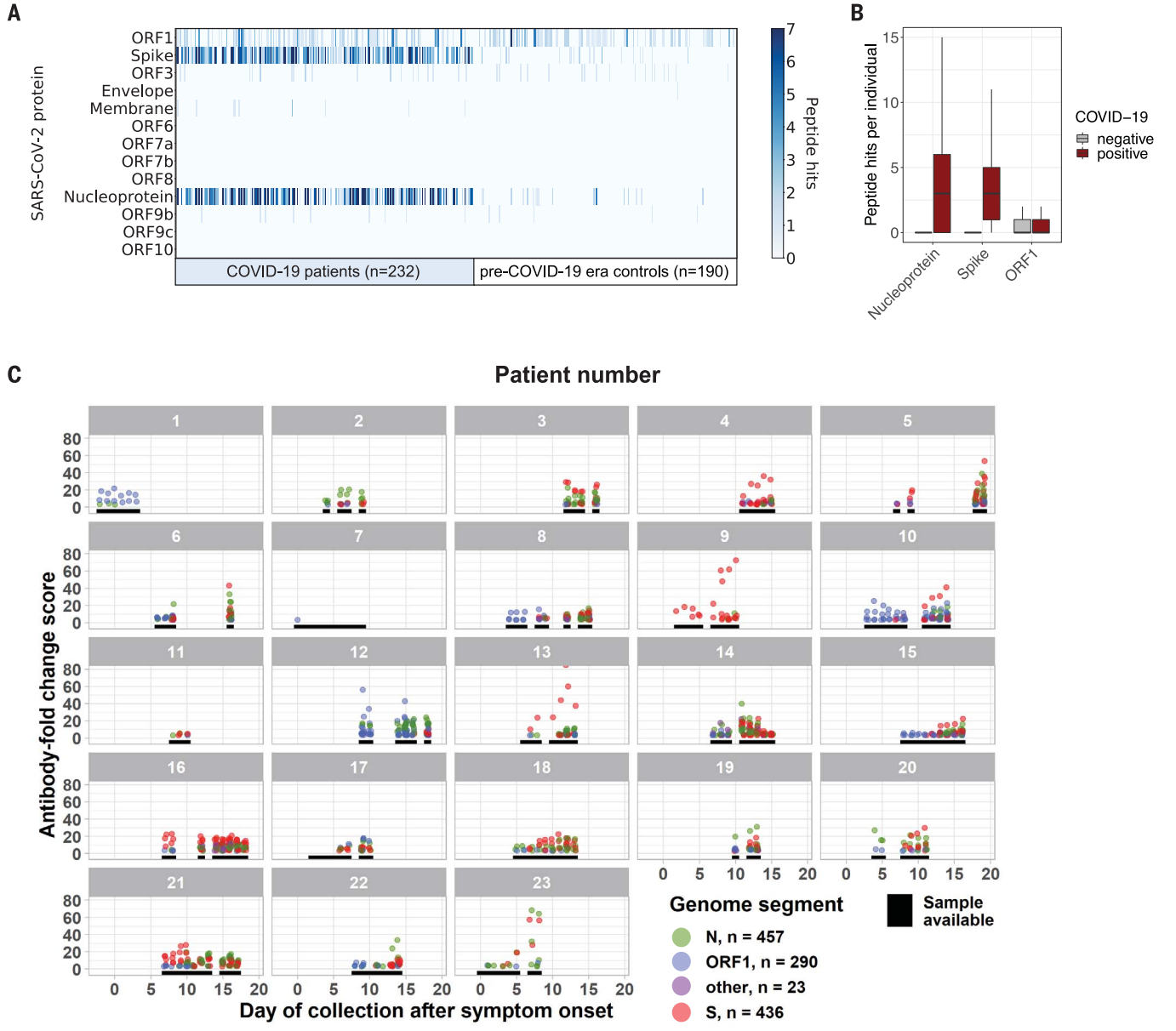


Fig. 2. SARS-CoV-2 protein recognition in COVID-19 patient versus control sera. (**A**) Antibodies targeting SARS-CoV-2 proteins. Each column represents a distinct patient sample, and each row represents a SARS-CoV-2 protein. The color intensity in each cell of the heatmap indicates the number of 56-mer peptides, as in Fig. 1D. (**B**) Box plots (as in Fig. 1E) illustrate the number of peptide hits from each of the indicated SARS-CoV-2 proteins detected in the IgG antibody response of COVID-19 patients and controls. (**C**) Longitudinal analysis of the antibody response to SARS-CoV-2 for 23 patients with confirmed COVID-19. Black lines indicate days when a sample was available for analysis. Each point represents the maximum antibody fold-change score per SARS-CoV-2 peptide in each sample, colored by protein target.

most discriminatory 20-mers in the VirScan data. IgG responses in COVID-19 patients were highly correlated between the Luminex and VirScan assays, providing orthogonal validation of the VirScan data and supporting the prevalence of SARS-CoV-2–induced humoral responses to these regions of S and N (fig. S1D).

Differential antibody responses to common viruses in hospitalized versus nonhospitalized COVID-19 patients

We next considered whether differences in the antibody response to SARS-CoV-2 or to other viruses might be associated with the severity of COVID-19. We grouped the COVID-19 patients into two subsets: those who required hospitalization (n = 101) and those who did not (n = 131). We compared the responses to peptides derived from the SARS-CoV-2 S and N proteins between the hospitalized (H) and nonhospitalized (NH) groups and found that the H group exhibited stronger and broader antibody responses to S and N peptides that might be due to epitope spreading (Fig. 5A). We then analyzed 32 NH COVID-19 samples, 32 H COVID-19 samples, and 32 pre–COVID-19 era negative controls with the Luminex assay and similarly observed that the H group had stronger and broader antibody responses to SARS-CoV-2–specific peptides than did the NH group (Fig. 5B).

VirScan also offers the opportunity to examine the history of previous viral infections and to determine correlates of COVID-19 outcomes. For example, prior viral exposure could provide some protection if cross-reactive neutralizing antibodies or T cell responses are stimulated upon exposure to SARS-CoV-2 (*14*, *15*). Alternatively, cross-reactive antibodies to viral

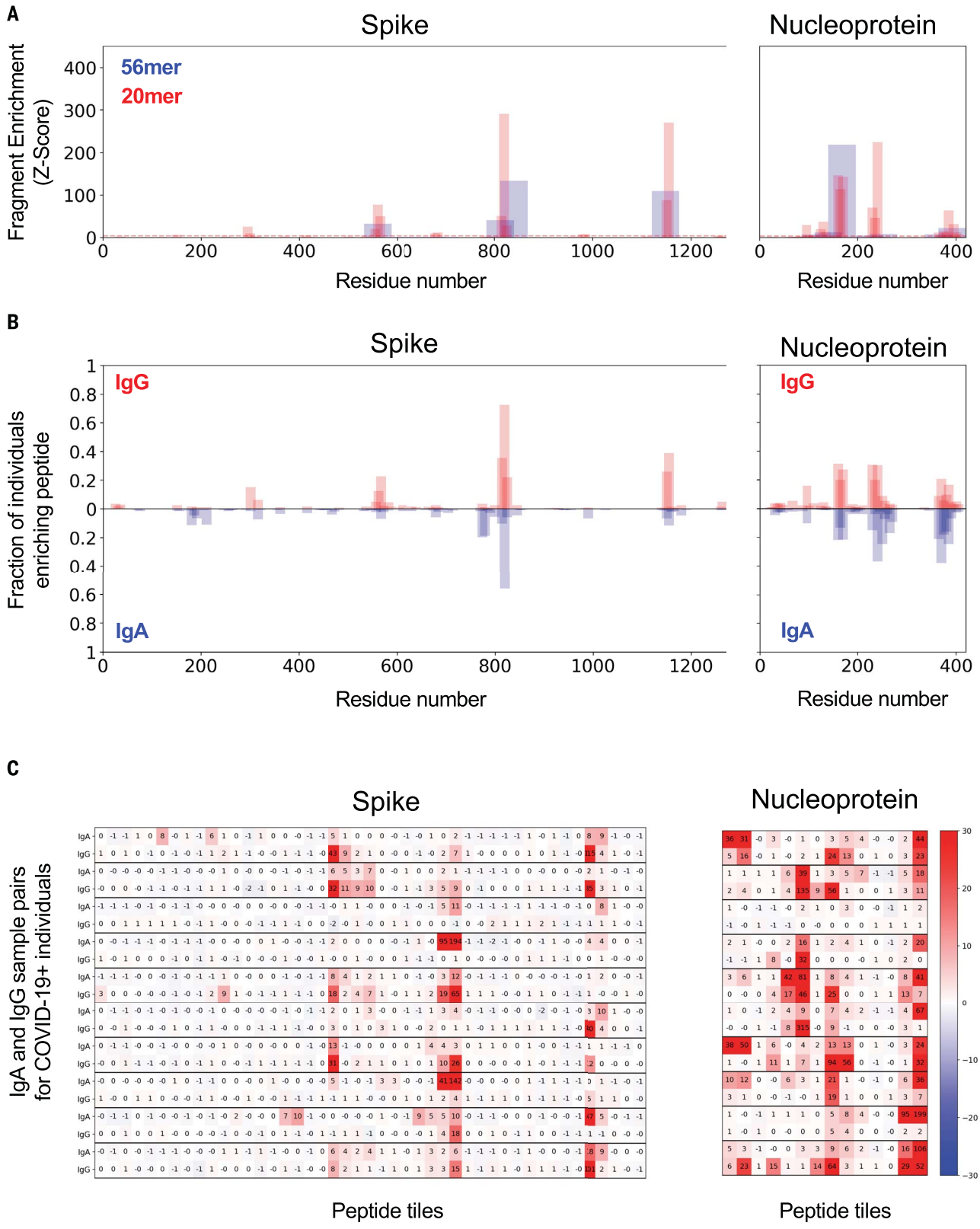


Fig. 3. IgG and IgA recognition of immunodominant regions in SARS-CoV-2 spike and nucleoprotein. (**A**) Example response to S and N proteins from a single COVID-19 patient. The *y* axis indicates the strength of enrichment (*z*-score; see Materials and methods) of each 56-mer (blue) or 20-mer (red) peptide recognized by the IgG antibodies present in the serum sample. (**B**) Common responses to S and N proteins across COVID-19 patients. The *y* axis indicates the fraction of COVID-19 patient samples (n = 348) enriching each 20-mer peptide with either IgG (top) or IgA (bottom) antibodies. (**C**) Comparison of the IgA and IgG responses in individual COVID-19 patients. Each set of two rows represents the IgG and IgA antibody specificities of a single patient, with data displayed for 10 representative COVID-19 patients. Numeric values indicate the degree of enrichment (*z*-score) of each peptide tiling across the S and N proteins.

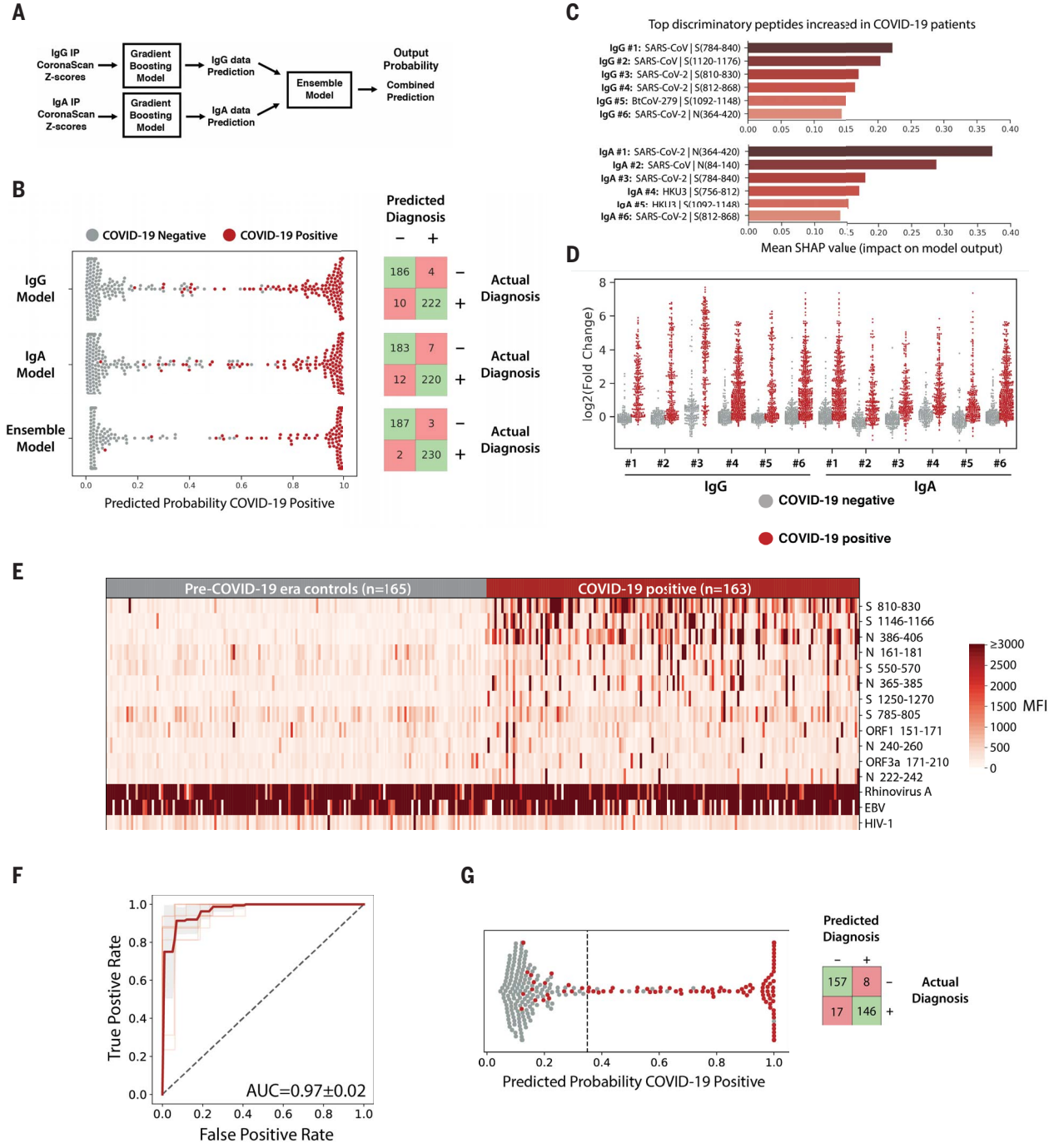


Fig. 4. Machine learning models trained on VirScan data discriminate COVID-19–positive and –negative individuals with very high sensitivity and specificity. (**A**) Gradient-boosting machine learning models were trained on IgG and IgA VirScan data from 232 COVID-19 patients and 190 pre–COVID-19 era controls. Separate models were created for the IgG and IgA data, and then a third model (Ensemble) was trained to combine the outputs of the first two. (**B**) The plot shows the predicted probability that each sample is positive for COVID-19. True COVID-19–positive samples are shown as red dots; true COVID-19–negative samples are shown as gray dots. The corresponding confusion matrix for each model is shown on the right. (**C** and **D**) SHAP analysis to identify the most discriminatory peptides informing the models in (B). The chart in (C) summarizes the relative importance of the most discriminatory peptides increased among COVID-19 patients identified by the IgG and IgA gradient-boosting models. The enrichment [$\log_2$(fold change) of the normalized read counts in the sample IP versus in no-serum control reactions] of each of these peptides across all samples is shown in (D). (**E**) Luminex assay using highly discriminatory SARS-CoV-2 peptides identifies IgG antibody responses in COVID-19 patients but rarely in pre–COVID-19 era controls. Each column represents a COVID-19 patient (n = 163) or pre–COVID-19 era control (n = 165); each row is a SARS-CoV-2–specific peptide. Peptides containing public epitopes from rhinovirus A, EBV, and HIV-1 served as positive and negative controls. The color scale indicates the median fluorescence intensity (MFI) signals after background subtraction. (**F**) Receiver operating characteristic (ROC) curve for the Luminex assay predicting SARS-CoV-2 infection history, evaluated by 10× cross-validation. The light red lines indicate the ROC curve for each test set, the dark line indicates the average, and the gray region represents ±1 SD. The average area under the curve (AUC) is shown. (**G**) (Left) Predicted probability that each sample is positive for COVID-19, using the Luminex model, as in (B). The dashed line indicates the model threshold. (Right) Confusion matrix for the Luminex model.

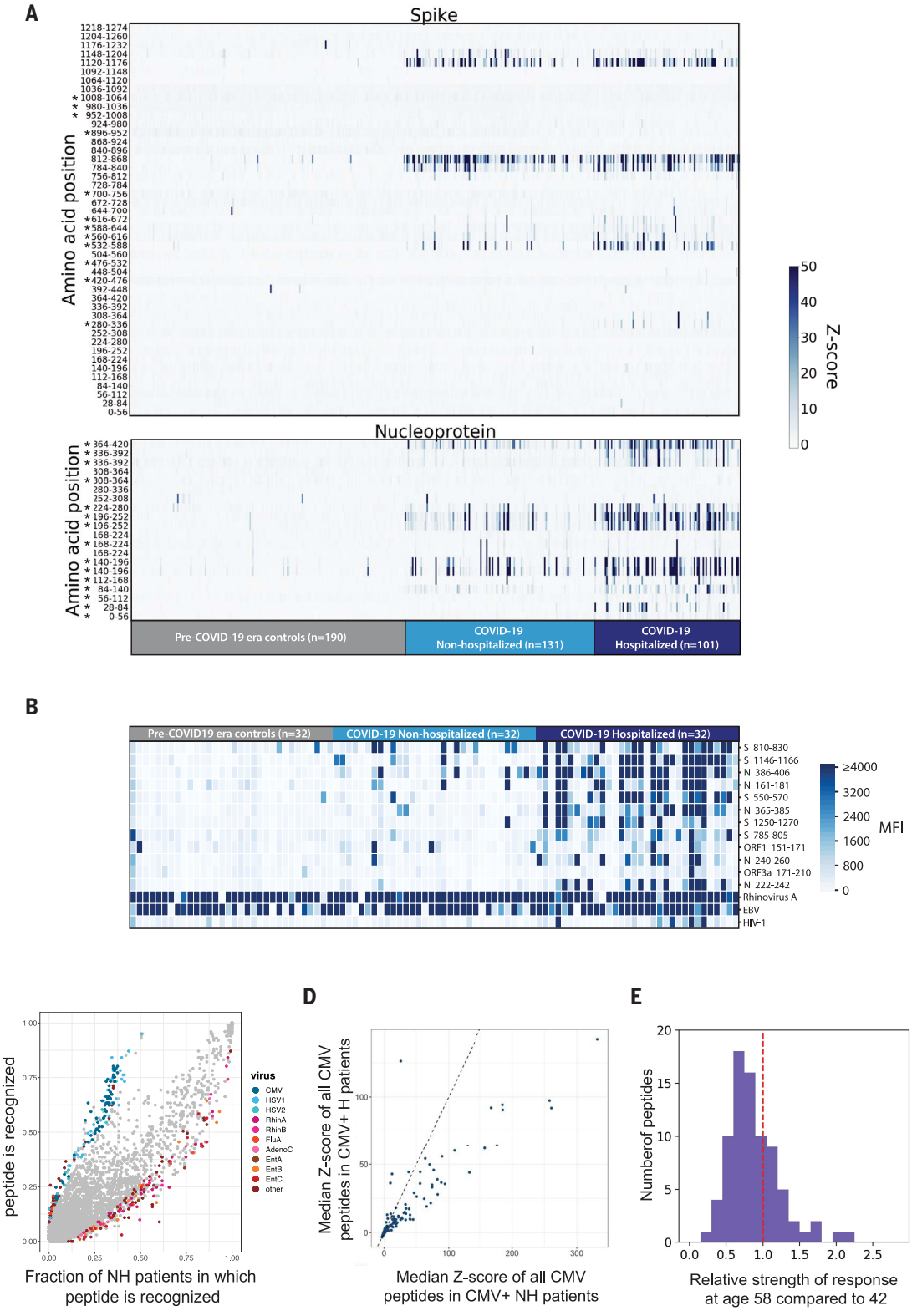


Fig. 5. Correlates of COVID-19 severity. (**A**) Differential recognition of peptides from SARS-CoV-2 N and S between COVID-19 nonhospitalized patients (n = 131), hospitalized patients (n = 101), and pre–COVID-19 era negative controls (n = 190). Each column represents a specific patient and each row represents a peptide tile; tiles are labeled by amino acid start and end position and may be duplicated for intervals for which amino acid sequence diversity is represented in the library. Color intensity represents the degree of enrichment (z-score) of each peptide in IgG samples. Asterisks indicate peptides that exhibit a significant increase in recognition by sera from hospitalized versus nonhospitalized patients (Kolmogorov-Smirnov test, Bonferroni-corrected P value thresholds of 0.001 for S and 0.0025 for N). (**B**) SARS-CoV-2 Luminex assay identifies stronger IgG responses in hospitalized COVID-19 patients than in nonhospitalized COVID-19 patients. Each column represents either a nonhospitalized (n = 32) or hospitalized (n = 32) COVID-19+ patient or a pre–COVID-19 era control (n = 32); each row represents a peptide in the Luminex assay. The color scale indicates the MFI signals after background subtraction. (**C**) All peptides in the VirScan library are plotted by the fraction of nonhospitalized (x axis) and hospitalized COVID-19 patient IgG samples (y axis) in which they are recognized. A z-score threshold of 3.5 was used as an enrichment cutoff to count a peptide as positive. Peptides that exhibit statistically significant associations with hospitalization status are colored by virus of origin (Fisher's exact test, Bonferroni-corrected P value threshold of 8.52×10^{-7}). All peptides that do not exhibit significant association with hospitalization status are shown in gray. The significant peptides shown are collapsed for high sequence identity. (**D**) All peptides derived from CMV that are present in the VirScan library are plotted by median z-score for the nonhospitalized (x axis) and hospitalized COVID-19 patients (y axis). The line $y = x$ is shown as a dashed line. (**E**) Reduced recognition of mild disease–associated antigens with age. The histogram shows the relative recognition in healthy donors at age 58 compared with age 42 for each distinct antigen that was more strongly recognized by antibodies in nonhospitalized than hospitalized COVID-19 patients.

surface proteins could increase the risk of severe disease due to antibody-dependent enhancement (ADE), as has been observed for SARS-CoV (*16*, *17*). Furthermore, exposure to certain viruses could affect the response to SARS-CoV-2 by altering the immune system. To examine these possibilities, we analyzed the virome-wide VirScan data and found that overall, the NH patients exhibited greater responses to individual peptides from common viruses such as rhinoviruses, influenza viruses, and enteroviruses, whereas the H patients displayed more robust responses to peptides from cytomegalovirus (CMV) and herpes simplex virus 1 (HSV-1) (Fig. 5C). These observations may be influenced by demographic differences in the NH and H cohorts, as described below.

We sought to understand whether the differential reactivity to CMV and HSV-1 between the H and NH patients was due to differences in the strength of antibody responses or the prevalence of infection (these viruses are common, but not ubiquitous like rhinoviruses, enteroviruses, and influenza viruses). Using VirScan data, we found that the H group had a higher incidence of both CMV and HSV-1 infection: 82.2% (83 of 101) of the H group were positive for CMV versus 37.4% (49 of 131) of the NH group, whereas 92.1% (93 of 101) of the H group were positive for HSV-1 versus 45.8% (60/131) of the NH group. To examine the relative strength of the antibody responses, we considered only CMV- or HSV-1–seropositive individuals from the NH and H groups: The antibody response to both CMV (Fig. 5D) and HSV-1 (fig. S2) was stronger among the NH individuals. Thus, the differing seroprevalence of CMV and HSV-1 in the NH versus H groups likely explains the results shown in Fig. 5C. We conclude that antibody responses to nearly all viruses, except SARS-CoV-2, were weaker in the H patients than in the NH patients.

These notable differences led us to examine potential demographic covariates between the NH and H groups. We found that age, sex, and race were all significantly associated with COVID-19 severity (fig. S3), as has been reported (*18*, *19*). Older age, male sex, and nonwhite racial groups were significantly overrepresented in the H group compared with the NH group (fig. S3 and table S3). Furthermore, hospitalized males exhibited stronger responses to N than hospitalized females, whereas nonhospitalized males and females did not exhibit differential responses to any SARS-CoV-2 proteins (fig. S3E). Advanced age is a dominant risk factor for severe COVID-19 and is correlated with reduced immune function (*20*). In light of the age difference between the H (median age: 58) and NH (median age: 42) patients in our cohort, we reasoned that the antigens recognized more strongly in the NH group might reflect more general age-associated changes in humoral immunity. To test this hypothesis, we examined VirScan data for a cohort of 648 healthy, pre-pandemic donors. We characterized the recognition of each NH-associated peptide in subsets of healthy donors representing different age groups and observed a general decline in recognition with age, including a median 19% reduction in recognition from age 42 to 58 (Fig. 5E). These data suggest that age-related changes to the immune system may partially explain the observation of weaker antibody responses to most viruses in the H group. Although it is correlative and potentially influenced by other demographic differences between the NH and H cohorts, the broad age-related diminution in immune system activity that we observed could be an important aspect of the increased severity in the H group.

Cross reactivity of SARS-CoV-2 epitopes

We returned to the question of epitope cross-reactivity, this time examining antibody responses to the triple-alanine scanning library. For each 56-mer peptide spanning the SARS-CoV-2 proteome, this library contained a collection of scanning mutants: The first mutant peptide encoded three alanines instead of the first three residues, the second mutant peptide contained the three alanines moved one residue downstream, and so on (fig. S4). Antibodies that recognize the wild-type 56-mer peptide will not recognize mutant versions of the peptide containing alanine substitutions at critical residues. Thus, the location of the linear epitope can be deduced by looking for "antibody footprints," indicated by stretches of alanine mutants missing from the pool of immunoprecipitated phage. The first and last triple-alanine mutations to interfere with binding are expected to start two amino acids before the first residue that is essential for antibody binding and end two amino acids after the last.

With respect to cross-reactivity, IgG from COVID-19 patients recognized more 56-mer peptides from the common HCoVs HKU1, OC43, 299E, and NL63 than IgG from pre–COVID-19 era controls. This difference is primarily driven by a pronounced increase in recognition of S peptides from the HCoVs and is likely a result of cross-reactivity of antibodies developed during SARS-CoV-2 infection (Fig. 6A).

We mapped the position of all HCoV S peptides that display increased recognition in COVID-19 patient samples onto the SARS-CoV-2 S protein. This revealed four immunodominant regions recognized by >25% of COVID-19 patients (Fig. 6B). Comparing the frequency of peptide recognition between the COVID-19 patients and pre–COVID-19 era controls showed that two of these immunogenic regions in SARS-CoV-2 S are likely to cross-react strongly with homologous regions of other HCoVs, as the frequency of recognition of the HCoV peptides at these regions rises in COVID-19 patients. For instance, peptides from all four seasonal HCoVs that span the region corresponding to residues 811 to 830 of SARS-CoV-2 S are frequently recognized by COVID-19 patients but much less so by pre–COVID-19 era controls, suggesting that this recognition is a result of antibodies developed or boosted in response to SARS-CoV-2 infection. Using triple-alanine scanning mutagenesis (fig. S4), we mapped the antibody footprints in this region to an 11–amino acid stretch that is highly conserved between SARS-CoV-2 and all four common HCoVs, which explains the cross-reactivity (Fig. 6, C and D). Similarly, both SARS-CoV-2 and HCoV-OC43 peptides corresponding to S 1144-1163 are recognized much more frequently by COVID-19 patients than pre–COVID-19 era controls, and triple-alanine-scanning mutagenesis confirmed that the antibody footprints are located within a 10–amino acid stretch conserved between SARS-CoV-2 and HCoV-OC43 but not the other HCoVs. By contrast, the epitope sequences around S 551-570 and S 766-785 are not conserved between SARS-CoV-2 and the seasonal HCoVs, and indeed these epitopes are not cross-reactive. One HCoV-HKU1 peptide spanning S 551-570 scores in both COVID-19 patients and pre–COVID-19 era control samples; however, its frequency of detection is not further boosted in COVID-19 patients, suggesting that the antibody that recognizes the SARS-CoV-2 S 551-570 peptide is distinct from the antibody recognizing the HCoV-HKU1 peptide, consistent with sequence differences at this location (Fig. 6C).

Notably, we detect antibody responses to SARS-CoV-2 S 811-830 in 79.9% of COVID-19 patients. However, we also see responses to the corresponding peptides from OC43 and 229E in ~20% of the pre–COVID-19 era controls, and these responses seem to cross-react with SARS-CoV-2. It is possible that some patients have preexisting antibodies to this region that cross-react and are expanded during SARS-CoV-2 infection. This might explain the very high prevalence of antibody responses to this epitope and suggests that anamnestic responses to seasonal coronaviruses may influence the antibody response to SARS-CoV-2. Of note, this region is located directly after the predicted S2′ cleavage site for SARS-CoV-2 and overlaps the fusion peptide. A recent study showed that adding an excess of the fusion peptide reduced neutralization, implying that an antibody that binds the fusion peptide might contribute to neutralization by interfering with membrane fusion (*21*, *22*). Given the frequency of seroreactivity toward this epitope in COVID-19 patients, it will be important to determine whether the antibodies that recognize this epitope are neutralizing in future studies. If so, the prior presence of antibodies recognizing this

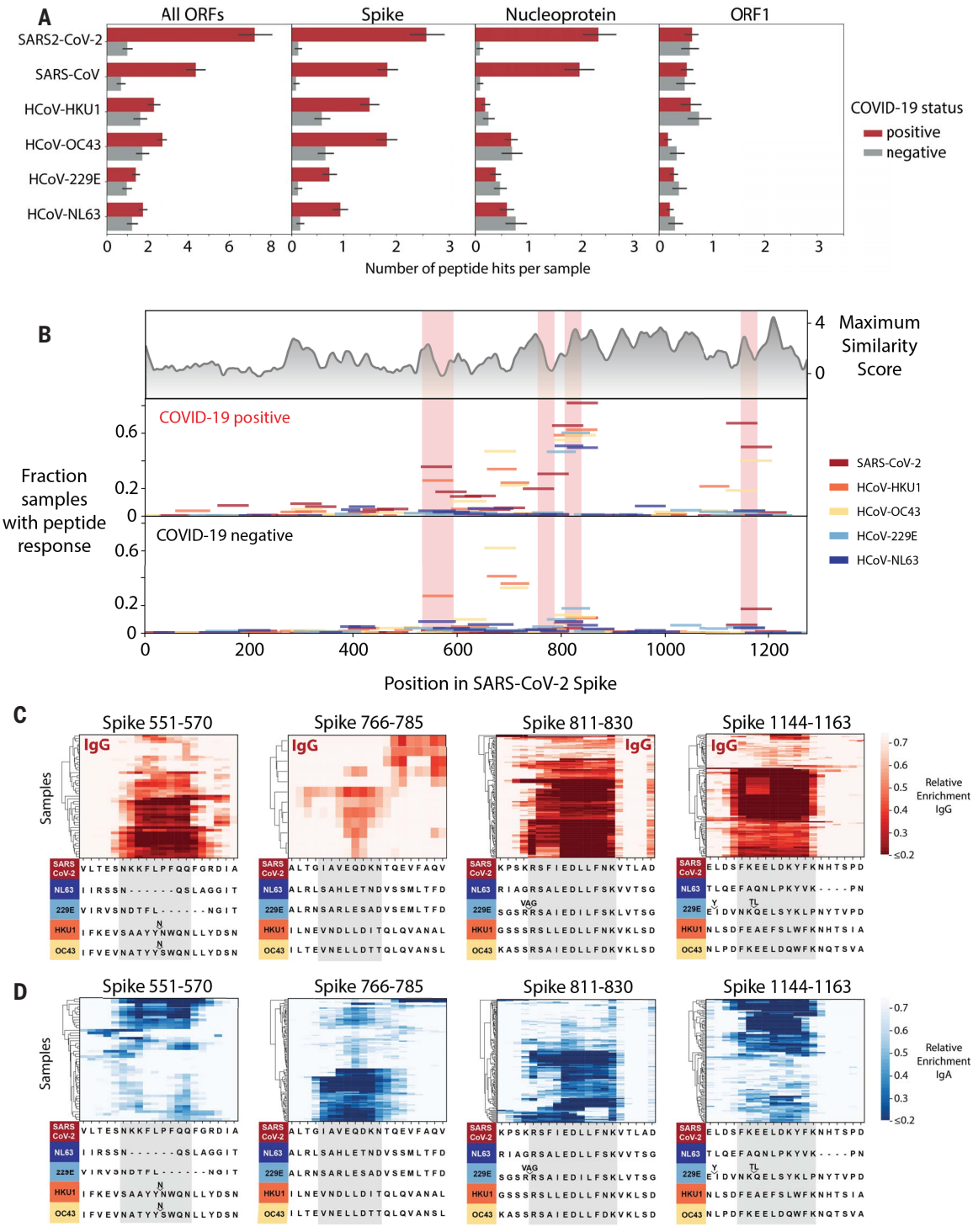


Fig. 6. Cross-reactive epitopes among human coronaviruses. (**A**) Bar graphs depict the average number of 56-mer peptides derived from SARS-CoV-2, SARS-CoV, and each of the four common HCoVs that are significantly enriched per sample (IgG IP). Error bars represent 95% confidence intervals. (**B**) Analysis of cross-reactive epitopes for HCoV S proteins. The upper plot shows the similarity of each region of the SARS-CoV-2 S protein to the corresponding region in the four common HCoVs (see Materials and methods). The frequency of peptide recognition is shown in the bottom two plots. Peptides from each virus are indicated by colored lines: The length of each line along the *x* axis indicates the corresponding region of the SARS-CoV-2 S protein covered by each peptide according to a pairwise protein alignment; the height of each line corresponds to the fraction of samples in which that peptide scored in either the IgG or IgA IPs. The epitopes mapped in (C) and (D) are highlighted in pink. (**C** and **D**) Mapping of recurrently recognized SARS-CoV-2 S IgG (C) and IgA (D) epitopes by triple-alanine scanning mutagenesis. Each plot represents a 20–amino acid region of the SARS-CoV-2 S protein within the regions highlighted in (B). Each column of the heatmap corresponds to an amino acid position, and each row represents a sample. The color intensity indicates the average enrichment of 56-mer peptides containing an alanine mutation at that site relative to the median enrichment of all mutants of that 56-mer in each sample. COVID-19 patients with a minimum relative enrichment below 0.6 in the specified window are shown. The amino acid sequence across each region of SARS-CoV-2 S, as well as an alignment of the corresponding sequences in the common HCoVs, is shown below each heatmap. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

epitope may affect the course of COVID-19 and mitigate severity.

Epitope mapping reveals hundreds of distinct SARS-CoV-2 epitopes, including likely epitopes of neutralizing antibodies

We also used the triple-alanine scanning mutagenesis library to map antibody footprints across the entire SARS-CoV-2 proteome (Fig. 7, fig. S5, and tables S10 to S19). We used a hidden Markov model (HMM) to analyze the mutagenesis data and detect antibody footprints. By integrating signals across stretches of consecutive residues, the HMM successfully distinguished antibody footprints from random noise and thus detected regions containing epitopes with improved sensitivity and far greater resolution than was possible with the 56-mer peptide data alone (see Materials and methods) (figs. S6 and S7 and tables S15 to S18). We performed hierarchical clustering on the antibody footprints identified by the HMM to determine the number of distinct epitopes (here defined as distinct antibody footprints) that we detected across the SARS-CoV-2 proteome (fig. S8 and table S10). Overall, we identified 3103 antibody footprints across 169 COVID-19 patient samples and mapped 823 distinct epitopes (table S19). These epitopes are not evenly distributed along the proteins but rather fall into 303 epitope clusters, each of which contains multiple overlapping epitopes (fig. S8). For example, across the 89 IgA samples that recognized the epitope cluster from S 1135-1165, we identified nine epitopes that overlap but have distinct triple-alanine scanning profiles that indicate distinct antibody-epitope interactions (fig. S8C). Individual epitopes are recognized at a wide range of frequencies in the COVID-19 patients. The average COVID-19 patient sample contained antibodies to ~18 distinct linear epitopes (fig. S9), although this is likely an underestimate of the total epitope count per person, as VirScan does not efficiently detect antibodies recognizing discontinuous (conformational) epitopes (although such antibodies may retain some affinity to linear peptides that constitute the epitope).

The SARS-CoV-2 epitope landscape includes regions recognized by antibodies in a large fraction of COVID-19 patients ("public" epitopes) and regions recognized by antibodies in only one or a few individuals ("private" epitopes). For example, we mapped six distinct epitopes in the region spanning N 151-175 (fig. S5C). One of these epitopes was recognized by nearly one-third of the COVID-19 patients, whereas the rest were detected by <2% of the COVID-19 patients. Similarly, the region spanning S 766-835 contained more than 20 distinct epitopes, including the highly public epitope cluster near S815 and the public epitope cluster near S770 that is preferentially recognized by IgA (Fig. 7B). The public epitope cluster near S770 was recognized in 43% of COVID-19 patient IgA samples but only 4% of COVID-19 patient IgG samples. In another example, we detected at least 20 distinct epitopes within a stretch of just 46 residues in N 363-408, 10 of which were specific to IgA and 2 of which were specific to IgG (fig. S5D). The positions of several public epitope clusters are shown mapped onto the structure of SARS-CoV-2 in fig. S10.

We also mapped at least 12 distinct epitopes in the SARS-CoV-2 RBD, including 5 in the receptor binding motif that binds ACE2, the human receptor for SARS-CoV-2, and 6 that overlap ACE2 binding sites (Fig. 7, C and D, and fig. S6A). For example, S 414-427 (labeled E2 in Fig. 7) spans residue K417 in the RBD; K417 makes a direct contact with the human ACE2 protein in structures of ACE2 bound to the RBD. Thus, antibodies that recognize E2 are likely to block ACE2 binding and have neutralizing activity (Fig. 7E). Epitope S 454-463 (labeled E6 in Fig. 7) also overlaps ACE2 contact residues and partially overlaps the binding site of the neutralizing antibody CB6, which suggests that antibodies recognizing this epitope also have neutralizing potential (*23*–*25*) (Fig. 7G). Several other epitopes also span or are adjacent to critical residues contacted by ACE2 (Fig. 7, F and H). Thus, our data reveal some of the likely binding sites for neutralizing antibodies.

Discussion

In this study, we used VirScan to analyze sera from COVID-19 patients and pre–COVID-19 era controls to provide an in-depth serological description of antibody responses to SARS-CoV-2. We mapped the landscape of linear epitopes in the SARS-CoV-2 proteome, characterized their specificity or cross-reactivity, and investigated serological and viral exposure history correlates of COVID-19 severity.

Identification of SARS-CoV-2 epitopes recognized by COVID-19 patients

VirScan detected robust antibody responses to SARS-CoV-2 in COVID-19 patients. These were primarily directed against the S and N proteins, with substantial cross-reactivity to SARS-CoV and milder cross-reactivity with the distantly related MERS-CoV and seasonal HCoVs. Cross-reactive responses to SARS-CoV-2 ORF1 were frequently detected in pre–COVID-19 era controls, suggesting that these result from antibodies induced by other pathogens.

At the population level, most SARS-CoV-2 epitopes were recognized by both IgA and IgG antibodies. We found that individuals often exhibited a "checkerboard" pattern, using either IgG or IgA antibodies against a given epitope. This suggests that a given IgM clone often evolves into either an IgG or IgA antibody, potentially influenced by local signals, and that, within an individual, there may often be a largely monoclonal response to a given epitope.

Examination of the humoral response to SARS-CoV-2 at the epitope level using the triple-alanine scanning mutagenesis library revealed 145 epitopes in S, 116 in N, and 562 across the remainder of the SARS-CoV-2 proteome (table S10). Most S epitopes were located on the surface of the protein or within unstructured regions that often abut, but seldom overlap, glycosylation sites (fig. S11). These epitopes ranged from private to highly public, with one region of S (S 811-830) being recognized by 79.9% of COVID-19 patients. Triple-alanine scanning mutagenesis showed highly conserved antibody footprints for some epitope clusters and diverse antibody footprints for others, indicating varying levels of conservation at the antibody-epitope interface among individuals (fig. S8). Peptides containing public epitopes could be used to isolate and clone antibodies from B cells bearing antigen-specific receptors. If these antibodies are found to lack protective effects or have deleterious effects, these regions could be mutated in future vaccines to divert the immunological response to other regions of S that might have more protective effects. Epitopes also varied in cross-reactivity, which can be explained by the presence or absence of sequence conservation between seasonal HCoVs and SARS-CoV-2 at these regions. Antibodies against several conserved epitopes in HCoVs seemed to be anamnestically boosted in COVID-19 patients. Antibodies recognizing one of these epitopes in the fusion peptide of S2 have been implicated in neutralization, and their presence prior to SARS-CoV-2 infection could mitigate the severity of COVID-19. Collectively, these data help explain why many serological assays for SARS-CoV-2 produce false positives due to preexisting cross-reactive antibodies, some of which may potentially affect the consequences of future SARS-CoV-2 infections.

Development of SARS-CoV-2 signature peptides for detecting seroconversion by Luminex

Using machine learning models trained on VirScan data, we developed a classifier that predicts SARS-CoV-2 exposure history with 99% sensitivity and 98% specificity. We identified peptides frequently and specifically recognized by COVID-19 patients and used these to create a Luminex assay that predicts SARS-CoV-2 exposure with 90% sensitivity and 95% specificity. Notably, the Luminex assay requires only three peptides to perform comparably to full-antigen ELISAs and could be further optimized in the future. This highlights the utility of VirScan-based serological profiling in the development of rapid and efficient diagnostic assays based on public epitopes.

Correlates of severity in COVID-19 patients

An important goal is to uncover serological correlates of COVID-19 severity. To this end, we compared cohorts of COVID-19 patients

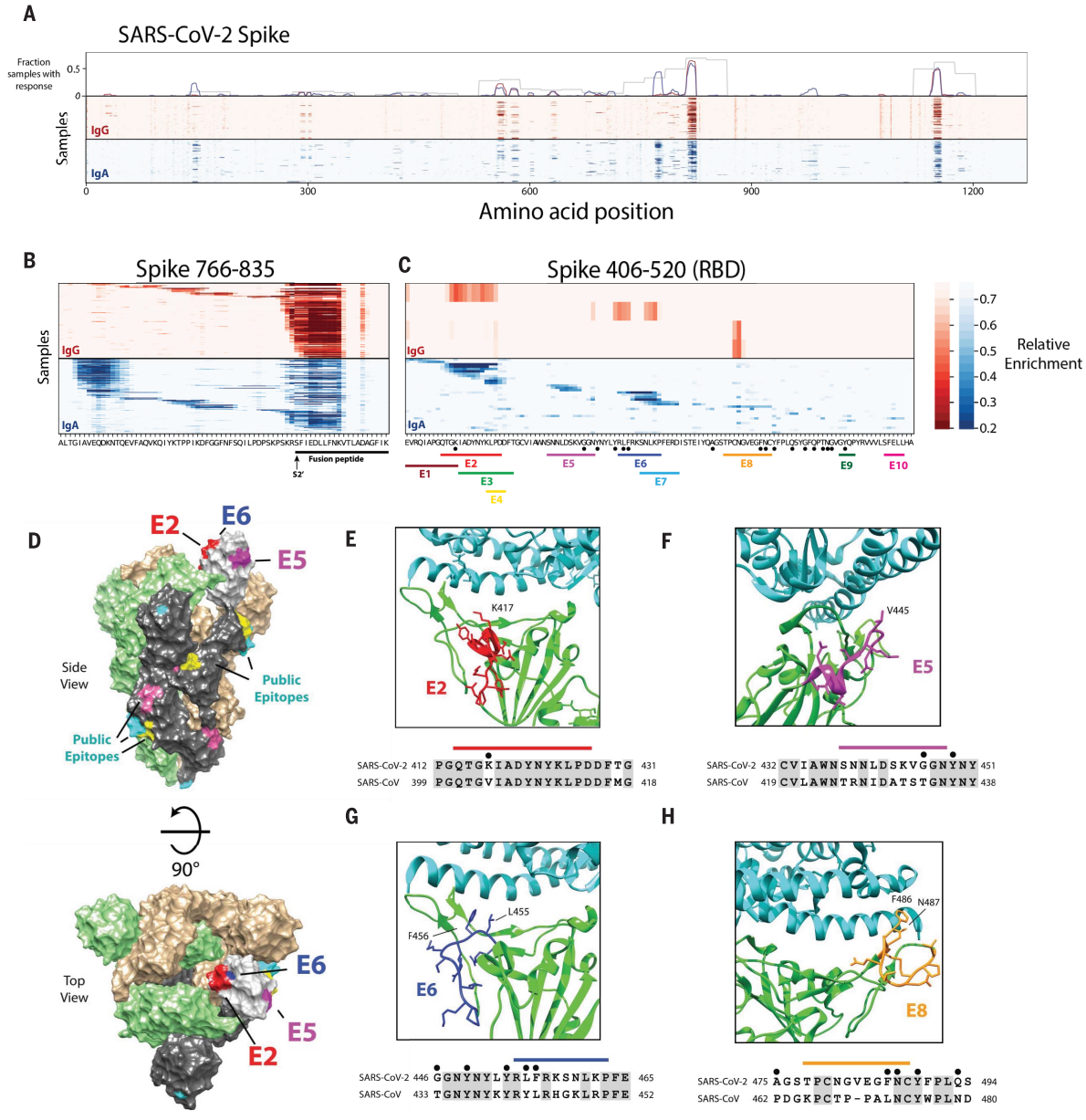


Fig. 7. High-resolution mapping of SARS-CoV-2 epitopes. (**A**) Mapping of antibody epitopes in the SARS-CoV-2 S protein using triple-alanine scanning mutagenesis. Each column of the heatmap corresponds to an amino acid position, and each row represents a COVID-19 patient. The color intensity indicates the average enrichment of three triple-alanine mutant 56-mer peptides containing an alanine mutation at that site, relative to the median enrichment of all mutants of that 56-mer. The upper panel shows the fraction of samples that recognized each region of S as mapped by the IgA 56-mer (gray) versus the IgA and IgG triple-alanine scanning data (blue and red, respectively). (**B** and **C**) Detailed plot of the triple-alanine scanning mutagenesis in (A) to show the epitope complexity within two regions: S 766-835 (B) and S 406-520 (C). The amino acid sequence at each position is shown on the *x* axis. In (B), the fusion peptide and predicted S2′ cleavage site are indicated below the sequence (*21*, *22*). In (C) the distinct IgA epitopes identified by the HMM and clustering algorithms are depicted by colored bars. Black dots correspond to ACE2 contact residues in the crystal structure of the RBD receptor complex (6M0J) (*23*). Epitopes in regions E9 and E10 were not picked up by the HMM classifier because of their short length; however, these regions score in multiple samples and correspond to accessible regions in the crystal structure, which suggests that they may represent true epitopes. (**D**) Cryo–electron microscopy (cryo-EM) structure of the partially open SARS-CoV-2 S trimer (6VSB) (*24*), highlighting the locations of the antibody epitopes mapped by triple-alanine scanning mutagenesis. The three S monomers are depicted in tan, green, and gray for the two closed and single open-conformation monomers, respectively. The RBD of the open monomer is show in light gray. Three of the RBD epitopes from (C) that overlap ACE2 contact residues and are resolved in the cryo-EM structure (E2, E5, and E6) are highlighted in red, purple, and blue, respectively. The locations of additional public epitopes that were mapped in at least 10 samples across the IgG and IgA experiments are depicted in yellow, pink, and cyan. (**E** to **H**) The locations of four of the epitope footprints mapped in (C) are shown in relation to the RBD-ACE2 binding interface. The upper image for each panel shows the structure (6M0J) of SARS-CoV-2 RBD (green) in complex with ACE2 (cyan). The E2, E5, E6, and E8 epitopes are highlighted in red, purple, blue, and orange, respectively. Below each structure image is the sequence alignment of the regions of the SARS-CoV-2 and the SARS-CoV S proteins encompassing each epitope. The colored bars indicate each epitope, the black dots indicate residues that directly interact with ACE2 in the crystal structure, and the shaded residues indicate conservation between SARS-CoV-2 and SARS-CoV.

who did (H) or did not (NH) require hospitalization. Using both VirScan and the COVID-19 Luminex assay, we noticed a pronounced and somewhat counterintuitive increase in recognition of peptides derived from the SARS-CoV-2 S and N proteins among the H group, with more extensive epitope spreading. Whether this is a cause or a consequence of severe disease is not clear. Individuals whose innate and adaptive immune responses are not able to quell the infection early may experience a higher viral antigen load, a prolonged period of antibody evolution, and epitope spreading. Consequently, these patients might develop stronger and broader antibody responses to SARS-CoV-2 and could be more likely to have hyperinflammatory reactions such as cytokine storms that increase the probability of hospitalization. We noticed that hospitalized males had more robust antibody responses to SARS-CoV-2 than hospitalized females. This finding may indicate that males in this group are less able to control the virus soon after infection, and it is consistent with reported differences in disease outcomes for males and females (*18*, *19*).

VirScan allowed us to examine viral exposure history, which revealed two notable correlations. First, the seroprevalence of CMV and HSV-1 was much greater in the H group than the NH group. The demographic differences in our relatively small cohort of H versus NH COVID-19 patients make it impossible for us to conclusively determine whether CMV or HSV-1 infection affects disease outcome or is simply associated with other covariates such as age, race, and socioeconomic status. Although CMV prevalence slightly increases with age after 40, it also differs greatly among ethnic and socioeconomic groups (*26*, *27*). CMV is a chronic herpesvirus that is known to have a profound impact on the immune system: It can skew the naïve T-cell repertoire (*28*) and decrease T and B cell function (*29*) and is associated with higher systemic levels of inflammatory mediators (*30*) and increased mortality of people >65 years of age (*31*). The effects of CMV on the immune system could potentially influence COVID-19 outcomes.

The second notable correlation we observed was a substantial decrease in the levels of antibodies that target ubiquitous viruses such as rhinoviruses, enteroviruses, and influenza viruses in COVID-19 H patients compared with NH patients. When we examined only the CMV+ or HSV-1+ individuals in the two groups, we found that the strength of the antibody response to CMV and HSV-1 peptides was also reduced in the H group. We examined the effects of age on viral antibody levels in a pre–COVID-19 era cohort and found a diminution with age in the antibody response against viral peptides differentially recognized between the H and NH groups, consistent with previous studies on the effects of aging on the immune system (*20*). This inferred reduced immunity during aging could affect the severity of COVID-19 outcomes.

In correlative analyses such as these, it is difficult to draw strong conclusions about causality, given the demographic differences in the NH versus H groups. The NH group is younger and has a higher percentage of white and female individuals (average age 42, 66% female) than the H group (average age: 58; 42% female) (fig. S2), consistent with well-documented demographic skews in severely affected COVID-19 patients (*18*, *19*). However, even if age and other demographic factors are covariates, CMV seropositivity and age-related reduction in antibody titers against viral antigens, as described here, could still influence the severity of infection. To test these hypotheses, a much larger cohort of COVID-19 patients with severe and mild disease that could be matched for age, race, and sex is required. Such future studies have the potential to enhance our understanding of the biological mechanisms underlying variable outcomes of COVID-19.

Deep serological profiling can provide a window into the breadth of viral responses, how they differ in patients with diverse outcomes, and how past infections may influence present responses to viral infections. Understanding the epitope landscape of SARS-CoV-2, particularly within S, provides a stepping stone to the isolation and functional dissection of both neutralizing antibodies and antibodies that might exacerbate patient outcomes through ADE and could inform the production of improved diagnostics and vaccines for SARS-CoV-2.

Materials and methods

Sources of serum used in this study

Cohort 1

Plasma samples were from volunteers recruited at Brigham and Women's Hospital who had recovered from a confirmed case of COVID-19. All volunteers had a polymerase chain reaction (PCR)–confirmed diagnosis of COVID-19 before being admitted to the study. Volunteers were invited to donate specimens after recovering from their illness and were required to be symptom free for a minimum of 7 days. Participants provided verbal and/or written informed consent and provided blood specimens for analysis. Clinical data, including date of initial symptom onset, symptom type, date of diagnosis, date of symptom cessation, and severity of symptoms, were recorded for all participants, as were results of COVID-19 molecular testing. Participation in these studies was voluntary, and the study protocols have been approved by the respective institutional review boards (IRBs).

Cohort 2

Serum samples from patients with PCR-confirmed COVID-19 cases while admitted to the hospital and from patients who were actively enrolled into a prospective study of COVID-19 infection were provided by collaborators from the University of Washington. Residual clinical blood specimens were used. Clinical data, including symptom duration and comorbidities, were extracted from medical records and participant-completed questionnaires. All study procedures have been approved by the University of Washington Institutional Review Board.

Cohort 3

Plasma samples were provided by collaborators from Ragon Institute of MGH, MIT and Harvard and Massachusetts General Hospital from study participants in three categories: (i) PCR-confirmed COVID-19 cases while admitted to the hospital; (ii) PCR-confirmed SARS-CoV-2–infected cases seen in an ambulatory setting; and (iii) PCR-confirmed COVID-19 cases in their convalescent stage. All study participants provided verbal and/or written informed consent. Basic data on days since symptom onset were recorded for all participants, as were results of COVID-19 molecular testing. Participation in these studies was voluntary, and the study protocols have been approved by the Partners Institutional Review Board.

Cohort 4

Patients were enrolled in the emergency department (ED) at Massachusetts General Hospital in Boston from 15 March to 15 April 2020 during the peak of the COVID-19 surge, with an IRB-approved waiver of informed consent. These included patients 18 years or older with a clinical concern for COVID-19 upon ED arrival and acute respiratory distress with at least one of the following: (i) tachypnea ≥22 breaths per minute, (ii) oxygen saturation ≤92% on room air, (iii) a requirement for supplemental oxygen, or (iv) positive-pressure ventilation. A blood sample was obtained in a 10-ml EDTA tube concurrent with the initial clinical blood draw in the ED. Blood was also drawn on days 3 and 7 if the patient was still hospitalized on those dates. Clinical course was followed to 28 days post-enrollment or until hospital discharge if that occurred after 28 days.

Enrolled individuals who were positive for SARS-CoV-2 were categorized into four outcome groups: (i) requiring mechanical ventilation, with subsequent death; (ii) requiring mechanical ventilation and subsequently recovered; (iii) requiring hospitalization on supplemental oxygen but not mechanical ventilation; and (iv) discharged from ED and not subsequently readmitted with supplemental oxygen. Demographic, past medical, and clinical data were collected and summarized for each outcome group, using medians with interquartile ranges and proportions with 95% confidence intervals, where appropriate.

Cohorts 5 and 6

Longitudinal Hopkins cohort: Remnant serum specimens were collected longitudinally from PCR-confirmed COVID-19 patients seen at Johns Hopkins Hospital. Samples were de-identified before analysis, with linked time since onset of symptom information. Specimens were obtained and used in accordance with an approved IRB protocol.

Cohorts 7 and 8

Cohorts 7 and 8 were from previous studies (*7*, *8*).

Cohort 9

Plasma samples were collected from consenting participants (37 female and 51 male individuals; 18 to 85 years old) of the Partner's Biobank program at Brigham and Women's Hospital during the period from July to August 2016. Plasma was harvested after a 10-min 1200xg ficoll density centrifugation from blood that was diluted 1:1 in phosphate buffered saline. Samples were frozen at −30°C in 1-ml aliquots. All samples were collected with Partners Institutional Review Board approval.

Blood sample collection methods

For cohorts 1 to 3: Blood samples were collected into EDTA (ethylenediamine tetraacetic acid) tubes and spun for 15 min at 2600 rpm according to standard protocol. Plasma was aliquoted into 1.5-ml cryovials and stored at −80°C until analyzed. Only de-identified plasma aliquots including metadata (e.g., days since symptom onset, severity of illness, hospitalization, ICU status, survival) were shared for this study. When appropriate for nonconvalescent samples, plasma or serum was also heat-inactivated at 56°C for 60 min and stored at ≤20°C until analyzed.

For cohort 4: Blood samples were collected in EDTA tubes and processed no more than 3 hours post–blood draw in a biosafety level 2+ laboratory on site. Whole blood was diluted with room temperature RPMI medium in a 1:2 ratio to facilitate cell separation for other analyses using the SepMate PBMC isolation tubes (STEMCELL) containing 16 ml of Ficoll (GE Healthcare). Diluted whole blood was centrifuged at 1200 rcf for 20 min at 20°C. After centrifugation, plasma (5 ml) was pipetted into 15-ml conical tubes and placed on ice during PBMC separation procedures. Plasma was then centrifuged at 1000 rcf for 5 min at 4°C, pipetted in 1.5-ml aliquots into three cryovials (4.5 ml total), and stored at −80°C. For the current study, samples (200 μl) were first randomly allocated onto a 96-well plate on the basis of disease outcome grouping.

Design and cloning of the SARS-CoV-2 tiling and triple-alanine scanning library

Multiple VirScan libraries were constructed as described below. We created ~200-nt oligos encoding peptide sequences 56 amino acids in length, tiled with 28–amino acid overlap through the proteomes of all coronaviruses known to infect humans, including HCoV-NL63, HCoV-229E, HCoV-OC43, HCoV- HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2, as well as three closely related bat viruses (BatCoV-Rp3, BatCoV-HKU3, and BatCoV-279). For SARS-CoV-2, we included a number of coding variants available in early sequencing of the viruses. For SARS-CoV-2, we additionally made a 20–amino acid peptide library tiling every five amino acids. Additionally, for SARS-CoV-2 we made triple-alanine mutant sequences scanning through all 56-mer peptides. Non-alanine amino acids were mutated to alanine, and alanines were mutated to glycine. Each peptide in all three libraries was encoded in two distinct ways such that there were duplicate peptides that could be distinguished by DNA sequencing. We reverse-translated the peptide sequences into DNA sequences that were codon-optimized for expression in *Escherichia coli*, that lacked restriction sites used in downstream cloning steps (EcoRI and XhoI), and that were distinct in the 50 nt at the 5′ end to allow for unambiguous mapping of the sequencing results. Then we added adapter sequences to the 5′ and 3′ ends to form the final oligonucleotide sequences (table S1): These adapter sequences facilitated downstream PCR and cloning steps. Different adapters were added to each sublibrary so that they could be amplified separately. The resulting sequences were synthesized on a releasable DNA microarray (Agilent). We PCR-amplified the DNA oligo library with the primers shown below, digested the product with EcoRI and XhoI, and cloned it into the EcoRI/SalI site of the T7FNS2 vector (*5*). We packaged the resultant library into T7 bacteriophage using the T7 Select Packaging Kit (EMD Millipore) and amplified the library according to the manufacturer's protocol.

Primers used for analysis of the different libraries employed:

CoV 56-mer library:

5′ adapter: 5′-GAATTCGGAGCGGT-3′

3′ adapter: 5′-CACTGCACTCGAGA-3′

Forward primer: 5′-AATGATACGGCGTGAATTCGGAGCGGT-3′

Reverse primer: 5′-CAAGCAGAAGACGTCTCGAGTGCAGTG-3′

SARS CoV-2 triple-alanine scanning library:

5′ adapter: 5′-GAATTCCGCTGCGT-3′

3′ adapter: 5′-CAGGGAAGAGCTCG-3′

Forward primer: 5′-AATGATACGGCGGGAATTCCGCTGCGT-3′

Reverse primer: 5′-CAAGCAGAAGACTCGAGCTCTTCCCTG-3′

SARS-CoV-2 20-mer library:

5′ adapter: 5′-GAATTCCGCTGCGT-3′

3′ adapter: 5′-GTACTATACCTACGGAAGGCTCG-3′

Forward primer: 5′-AATGATACGGCGGGAATTCCGCTGCGT-3′

Reverse primer: 5′-TATCTCGCATAGCGCATATACTCGAGCCTTCCGTAGGTATAGTAC-3′

Phage immunoprecipitation and sequencing

We performed phage IP and sequencing as described previously or with slight modifications (*5*–*8*). For the IgA and IgG chain isotype-specific IPs, we substituted magnetic protein A and protein G Dynabeads (Invitrogen) with 6 μg of Mouse Anti-Human IgG Fc-BIOT (Southern Biotech) or 4 μg of Goat Anti-Human IgA-BIOT (Southern Biotech) antibodies. We added these antibodies to the phage and serum mixture and incubated the reactions overnight a 4°C. Next, we added 25 or 20 μl of Pierce Streptavidin Magnetic Beads (Thermo-Fisher) to the IgG or IgA reactions, respectively, and incubated the reactions for 4 hours at room temperature, then continued with the washing steps and the remainder of the protocol, as previously described (*5*–*8*).

Machine learning classifiers

Gradient-boosting classifier models for the VirScan data were generated using the XGBoost algorithm (version 1.0.2). Classifier models were trained to discriminate either COVID-19+ and COVID-19− patients (n = 232 and 190, respectively) or severe disease and mild disease (n = 101 hospitalized patients and n = 131 nonhospitalized patients). Two models were generated in each case, one using the z-scores for each VirScan peptide from the IgG IP as input features, and the other using the z-scores for each VirScan peptide from the IgA IP as input features. Additionally, a third logistic regression classifier was trained on the output probabilities from the IgG and IgA models to generate a combined prediction. The performance of each of the three model was assessed using a 20-fold cross-validation procedure, whereby predictions for each 5% of the data points were generated from a model trained on the remaining 95%. The SHAP package was used to identify the top discriminatory peptide features from each of the XGBoost models. The logistic regression models for the Luminex data were generated using the scikit-learn python package. The raw median fluorescence intensity (MFI) values were preprocessed using the RobustScalar function, then a logistic regression model was trained using the three most discriminate SARS-CoV-2 peptides. The model performance was quantified by 10-fold cross-validation.

High-resolution epitope identification and clustering

For each position in the 56-mer, the relative enrichment for each amino acid was calculated as the mean fold change of the three mutant peptides containing an alanine mutation

at that location relative to the median fold change of all alanine mutants for the 56-mer. Overlapping 56-mers were combined by taking the minimum value at each shared position to account for the possibility that an epitope is interrupted in one of the tiles by the peptide junction. To map the boundaries of antibody footprints from the triple-alanine scanning data for each sample we used the hmmlearn python package to develop a three-state HMM assuming a Gaussian distribution of relative enrichment emissions for each state. Mapped antibody footprints smaller than five amino acids in length were removed from the subsequent analysis. Next, we performed a two-step hierarchical clustering procedure to identify the number of distinct epitopes. First, for each protein all antibody footprints identified across the 169 COVID-19+ patient samples were clustered based on the start and stop locations predicted by the HMM classifier to generate epitope clusters. Next, to identify distinct epitopes, we performed an additional step of hierarchical clustering on the samples with epitopes within each epitope cluster based on the relative enrichment values of the triple-alanine mutants spanning the epitope (fig. S8).

Similarity-score calculation

Pairwise alignments were generated for the S proteins of SARS-CoV-2 and each of the four common HCoVs. Similarity scores were calculated separately for a 21–amino acid window centered at each position of the SARS-CoV-2 S protein. The mean similarity score between SARS-CoV-2 and the corresponding sequence of the other HCoV was calculated for each window using the BLOSUM62 substitution matrix with a gap opening and extending penalty of –10 and –1, respectively. The maximum similarity was score was calculated as the maximum value among the pairwise similarity scores between SARS-CoV-2 and each of the four common HCoVs for the sliding window centered at each position.

Luminex multiplex peptide epitope serology assays

Multiplexed SARS-CoV-2 peptide epitope assays were built using the peptides listed in table S9. Peptides were synthesized by the Ragon/MGH Peptide Core Facility with a Propargylglycine (Pra, X) moiety in the N terminus to facilitate cross-linking to Luminex beads using a "click" chemistry strategy as described previously (*13*). In brief, Luminex beads were first functionalized with amine-PEG4-azide and then reacted with the peptides to generate 20 different Luminex beads with attached peptides. Luminex bead–based serology assays were performed in 96-well U-bottom polypropylene plates using PBS + 0.1% bovine serum albumin as the assay buffer. Bead washes were done using PBS + 0.05% Triton X-100 by incubation for 1 min on a strong magnetic plate (Millipore-Sigma, Burlington, MA). All assay incubation times were 20 min. In the first step, beads were incubated with 20 µl of plasma samples. Samples used for the classifier were diluted 1:100, samples used to compare disease severity were diluted 1:300. After a wash step, bound IgA or IgG was detected by adding 40 µl of biotin-labeled anti-IgA or IgG antibodies at 0.1 µg/ml (Southern Biotechnology, Birmingham, AL). Next, 40 µl of phycoerythrin (PE)–labeled streptavidin (0.2 µg/ml) (Biolegend, San Diego, CA) and assay plates were analyzed on a Luminex FLEXMAP 3D instrument (Luminex Corporation, Austin, Texas) to generate MFI values to quantify peptide-specific IgA or IgG levels.

ELISA serology assays

ELISAs were performed separately using the SARS-CoV-2 N protein, S protein, or the S receptor-binding domain (RBD). 96-well plates were coated with antigen overnight. The plates were then blocked in PBS + 3% BSA. After washing with PBS + 0.05% Tween-20, the plasma sample were diluted 1:100, added to the plates and incubated overnight at 4°C. After incubation, the plates were washed three times with PBS + 0.05% Tween-20. The bound IgG was detected by adding anti-human IgG-alkaline phosphatase (Southern Biotech, Birmingham, AL) and incubating for 90 min at room temperature. The plates were washed an additional three times, after which p-nitrophenyl phosphate solution (1.6 mg/ml in 0.1 M glycine, 1 mM $ZnCl_2$, 1 mM $MgCl_2$, pH 10.4) was added to each well and allowed to develop for 2 hours. Bound IgG was quantified by measuring the OD405, and the reported values were calculated as the fold change over the pre–COVID-19 controls.

REFERENCES AND NOTES

1. J. Cui, F. Li, Z.-L. Shi, Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019). doi: 10.1038/s41579-018-0118-9; pmid: 30531947
2. D.-G. Ahn *et al.*, Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). *J. Microbiol. Biotechnol.* **30**, 313–324 (2020). doi: 10.4014/jmb.2003.03011; pmid: 32238757
3. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU); https://coronavirus.jhu.edu/map.html.
4. K. Yuki, M. Fujiogi, S. Koutsogiannaki, COVID-19 pathophysiology: A review. *Clin. Immunol.* **215**, 108427 (2020). doi: 10.1016/j.clim.2020.108427; pmid: 32325252
5. H. B. Larman *et al.*, Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* **29**, 535–541 (2011). doi: 10.1038/nbt.1856; pmid: 21602805
6. D. Mohan *et al.*, PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat. Protoc.* **13**, 1958–1978 (2018). doi: 10.1038/s41596-018-0025-6; pmid: 30190553
7. G. J. Xu *et al.*, Comprehensive serological profiling of human populations using a synthetic human virome. *Science* **348**, aaa0698 (2015). doi: 10.1126/science.aaa0698; pmid: 26045439
8. M. J. Mina *et al.*, Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* **366**, 599–606 (2019). doi: 10.1126/science.aay6485; pmid: 31672891
9. National Center for Biotechnology Information, National Library of Medicine, Protein database (2004) [cited 29 February 2020]; www.ncbi.nlm.nih.gov/protein/.
10. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020). doi: 10.1038/s41586-020-2012-7; pmid: 32015507
11. G. J. Gorse, G. B. Patel, J. N. Vitale, T. Z. O'Connor, Prevalence of antibodies to four human coronaviruses is lower in nasal secretions than in serum. *Clin. Vaccine Immunol.* **17**, 1875–1880 (2010). doi: 10.1128/CVI.00278-10; pmid: 20943876
12. S. M. Lundberg *et al.*, From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020). doi: 10.1038/s42256-019-0138-9; pmid: 32607472
13. M. B. Coppock, D. N. Stratis-Cullum, A universal method for the functionalization of dyed magnetic microspheres with peptides. *Methods* **158**, 12–16 (2019). doi: 10.1016/j.ymeth.2019.01.014; pmid: 30707950
14. A. Grifoni *et al.*, Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* **181**, 1489–1501.e15 (2020). doi: 10.1016/j.cell.2020.05.015; pmid: 32473127
15. N. Le Bert *et al.*, SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020). doi: 10.1038/s41586-020-2550-z; pmid: 32668444
16. Y. Wan *et al.*, Molecular Mechanism for Antibody-Dependent Enhancement of Coronavirus Entry. *J. Virol.* **94**, e02015-19 (2020). doi: 10.1128/JVI.02015-19; pmid: 31826992
17. S.-F. Wang *et al.*, Antibody-dependent SARS coronavirus infection is mediated by antibodies against spike proteins. *Biochem. Biophys. Res. Commun.* **451**, 208–214 (2014). doi: 10.1016/j.bbrc.2014.07.090; pmid: 25073113
18. M. Webb Hooper, A. M. Nápoles, E. J. Pérez-Stable, COVID-19 and Racial/Ethnic Disparities. *JAMA* **323**, 2466–2467 (2020). doi: 10.1001/jama.2020.8598; pmid: 32391864
19. S. Garg *et al.*, Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *Morb. Mortal. Wkly. Rep.* **69**, 458–464 (2020). doi: 10.15585/mmwr.mm6915e3; pmid: 32298251
20. E. Montecino-Rodriguez, B. Berent-Maoz, K. Dorshkind, Causes, consequences, and reversal of immune system aging. *J. Clin. Invest.* **123**, 958–965 (2013). doi: 10.1172/JCI64096; pmid: 23454758
21. C. M. Poh *et al.*, Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat. Commun.* **11**, 2806 (2020). doi: 10.1038/s41467-020-16638-2; pmid: 32483236
22. A. C. Walls *et al.*, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281–292.e6 (2020). doi: 10.1016/j.cell.2020.02.058; pmid: 32155444
23. J. Lan *et al.*, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020). doi: 10.1038/s41586-020-2180-5; pmid: 32225176
24. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020). doi: 10.1126/science.abb2507; pmid: 32075877
25. R. Shi *et al.*, A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020). doi: 10.1038/s41586-020-2381-y; pmid: 32454512
26. R. Lachmann *et al.*, Cytomegalovirus (CMV) seroprevalence in the adult population of Germany. *PLOS ONE* **13**, e0200267 (2018). doi: 10.1371/journal.pone.0200267; pmid: 30044826
27. S. L. Bate, S. C. Dollard, M. J. Cannon, Cytomegalovirus seroprevalence in the United States: The national health and nutrition examination surveys, 1988-2004. *Clin. Infect. Dis.* **50**, 1439–1447 (2010). doi: 10.1086/652438; pmid: 20426575
28. P. Klenerman, P. R. Dunbar, CMV and the art of memory maintenance. *Immunity* **29**, 520–522 (2008). doi: 10.1016/j.immuni.2008.09.008; pmid: 18957264
29. G. Pawelec *et al.*, Immunosenescence, suppression and tumour progression. *Cancer Immunol. Immunother.* **55**, 981–986 (2006). doi: 10.1007/s00262-005-0109-3; pmid: 16333622
30. J. L. Craigen *et al.*, Human cytomegalovirus infection up-regulates interleukin-8 gene expression and stimulates neutrophil transendothelial migration. *Immunology* **92**, 138–145 (1997). doi: 10.1046/j.1365-2567.1997.00310.x; pmid: 9370936

31. G. M. Savva *et al*., Cytomegalovirus infection is associated with increased mortality in the older population. *Aging Cell* **12**, 381–387 (2013). doi: 10.1111/acel.12059; pmid: 23442093
32. J. F. W. Chan *et al*., Middle East respiratory syndrome coronavirus: Another zoonotic betacoronavirus causing SARS-like disease. *Clin. Microbiol. Rev.* **28**, 465–522 (2015). doi: 10.1128/CMR.00102-14; pmid: 25810418
33. N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987). pmid: 3447015
34. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018). doi: 10.1093/molbev/msy096; pmid: 29722887
35. K. Tamura, M. Nei, S. Kumar, Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11030–11035 (2004). doi: 10.1073/pnas.0404206101; pmid: 15258291
36. D. E. Gordon *et al*., A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020). doi: 10.1038/s41586-020-2286-9; pmid: 32353859

ACKNOWLEDGMENTS

We thank J. Bensko, D. Gakpo, G. DeGregorio, and S. Fiscsh (Brigham and Women's Hospital) for assistance organizing the human cohorts; G. Alter (MGH) for facilitating samples from the Seattle cohort and advice; H. Chen for help on the VirScan assays; and the BPF Next-Gen Sequencing Core Facility at Harvard Medical School for their expertise and instrument availability that supported this work. **Funding:** This work was also supported by grants from the National Institutes of Health (NIH AI121394 and AI139538) and the Burroughs Wellcome Fund and the Fast Grant award to D.R.W.; from the Division of Intramural Research, NIAID, NIH, to O.L.; from the VoVRN to S.J.E.; from the Executive Committee on Research, MGH, to M.B.G. and M.R.F.; from the MassCPR to B.D.W., S.J.E., and D.R.W.; from the National Institutes of Health and Department of Defense (AI138318, AR070253, and W81XWH-16-1-0464) to J.A.L.; and from the NIH/NIAID (U24 grant) to H.B.L. and S.J.E. R.T.T. is supported by the Pemberton-Trinity Fellowship and a Sir Henry Wellcome Fellowship (201387/Z/16/Z). E.S. is funded by the NSF Graduate Research Fellowship Program. The COVID-19 sample biorepository was supported by a gift from E. Schwartz and by the Mark and Lisa Schwartz Foundation; the Massachusetts Consortium for Pathogen Readiness; and the Ragon Institute of MGH, MIT and Harvard. S.J.E. and B.D.W. are investigators with the Howard Hughes Medical Institute. **Author contributions:** Experimental design, E.S., E.F., T.K. R.T.T., J.A.L., H.B.L., and S.J.E. Investigation, E.S. E.F., X.Y., A.P.-T., T.K. J.A.L., I.-H.L., M.L.R., B.M.S., M.Q.L., Y.L., and R.T.T. Reagents and samples, Y.C., A.Z., D.M., Y.C., J.L., A.Z., D.R.M., F.J.N.L., M.T., S.H., J.L., MGH COVID-19 Collection & Processing Team, P.C., O.L., A.K., A.C.V., K.K., X.Y., A.P.-T., B.D.W., and D.R.W. Writing, E.S., E.F., R.T.T., T.K., and S.J.E. Supervision, S.J.E., D.R.W., A.C.V., K.K., M.R.F., B.W., H.Y.C., N.H., M.B.G., B.D.W., and H.B.L. **Competing interests:** S.J.E. and T.K. are founders of TSCAN Therapeutics. SJ.E. is a founder of MAZE Therapeutics and Mirimus. S.J.E. serves on the scientific advisory board of Homology Medicines, TSCAN Therapeutics, MAZE, XChem, and is an advisor for MPM, none of which affect this work. S.J.E., T.K., and H.B.L. are inventors on a patent application (US20160320406A) filed by Brigham and Women's Hospital that covers the use of the VirScan library to identify pathogen antibodies in blood. All other authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. All reasonable requests for materials will be fulfilled. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

MGH COVID-19 Collection & Processing Team participants
Collection Team: Kendall Lavin-Parsons[1], Blair Parry[1], Brendan Lilley[1], Carl Lodenstein[1], Brenna McKaig[1], Nicole Charland[1], Hargun Khanna[1], Justin Margolin[1]
Processing Team: Anna Gonye[2], Irena Gushterova[2], Tom Lasalle[2], Nihaarika Sharma[2], Brian C. Russo[3], Maricarmen Rojas-Lopez[3], Moshe Sade-Feldman[4], Kasidet Manakongtreecheep[4], Jessica Tantivit[4], Molly Fisher Thomas[4]
Massachusetts Consortium on Pathogen Readiness: Betelihem A. Abayneh[5], Patrick Allen[5], Diane Antille[5], Katrina Armstrong[5], Siobhan Boyce[5], Joan Braley[5], Karen Branch[5], Katherine Broderick[5], Julia Carney[5], Andrew Chan[5], Susan Davidson[5], Michael Dougan[5], David Drew[5], Ashley Elliman[5], Keith Flaherty[5], Jeanne Flannery[5], Pamela Forde[5], Elise Gettings[5], Amanda Griffin[5], Sheila Grimmel[5], Kathleen Grinke[5], Kathryn Hall[5], Meg Healy[5], Deborah Henault[5], Grace Holland[5], Chantal Kayitesi[5], Vlasta LaValle[5], Yuting Lu[5], Sarah Luthern[5], Jordan Marchewka (Schneider)[5], Brittani Martino[5], Roseann McNamara[5], Christian Nambu[5], Susan Nelson[5], Marjorie Noone[5], Christine Ommerborn[5], Lois Chris Pacheco[5], Nicole Phan[5], Falisha A. Porto[5], Edward Ryan[5], Kathleen Selleck[5], Sue Slaughenhaupt[5], Kimberly Smith Sheppard[5], Elizabeth Suschana[5], Vivine Wilson[5], Galit Alter[6], Alejandro Balazs[6], Julia Bals[6], Max Barbash[6], Yannic Bartsch[6], Julie Boucau[6], Josh Chevalier[6], Fatema Chowdhury[6], Kevin Einkauf[6], Jon Fallon[6], Liz Fedirko[6], Kelsey Finn[6], Pilar Garcia-Broncano[6], Ciputra Hartana[6], Chenyang Jiang[6], Paulina Kaplonek[6], Marshall Karpell[6], Evan C. Lam[6], Kristina Lefteri[6], Xiaodong Lian[6], Mathias Lichterfeld[6], Daniel Lingwood[6], Hang Liu[6], Jinqing Liu[6], Natasha Ly[6], Ashlin Michell[6], Ilan Millstrom[6], Noah Miranda[6], Claire O'Callaghan[6], Matthew Osborn[6], Shiv Pillai[6], Yelizaveta Rassadkina[6], Alexandra Reissis[6], Francis Ruzicka[6], Kyra Seiger[6], Libera Sessa[6], Christianne Sharr[6], Sally Shin[6], Nishant Singh[6], Weiwei Sun[6], Xiaoming Sun[6], Hannah Tichell[6], Alicja Trocha-Piechocka[6], Daniel Worrall[6], Alex Zhu[6], George Daley[7], David Golan[7], Howard Heller[7], Arlene Sharpe[7], Nikolaus Jilg[8], Alex Rosenthal[8], Colline Wong[8]

[1]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, USA. [2]Massachusetts General Hospital Cancer Center, Boston, MA, USA. [3]Division of Infectious Diseases, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [4]Massachusetts General Hospital Center for Immunology and Inflammatory Diseases, Boston, MA, USA. [5]Massachusetts General Hospital, Boston, MA, USA. [6]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [7]Harvard Medical School, Boston, MA, USA. [8]Brigham and Women's Hospital, Boston, MA, USA.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/eabd4250/suppl/DC1
Figs. S1 to S11
Tables S1 to S19
References (*37*–*39*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

22 June 2020; accepted 25 September 2020
Published online 29 September 2020
10.1126/science.abd4250

RESEARCH ARTICLE SUMMARY

TRANSCRIPTION

Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase

Hai Zheng*, Yilun Qi*, Shibin Hu*, Xuan Cao*, Congling Xu*, Zhinang Yin, Xizi Chen, Yan Li, Weida Liu, Jie Li, Jiawei Wang, Gang Wei, Kaiwei Liang, Fei Xavier Chen†, Yanhui Xu†

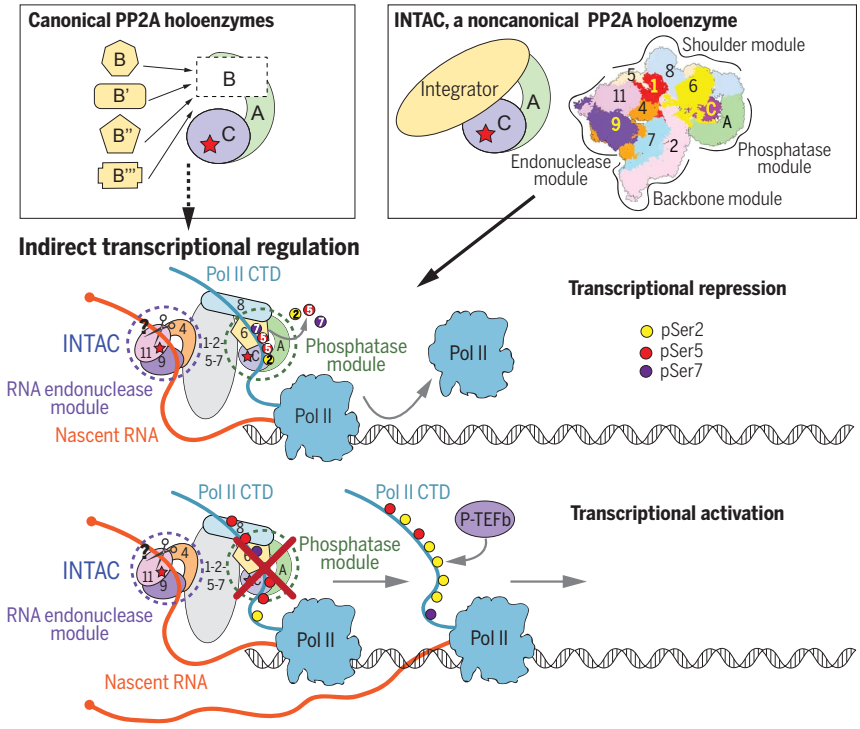
INTRODUCTION: RNA cleavage and phosphorylation-dephosphorylation of the RNA polymerase II (Pol II) C-terminal domain (CTD) are two regulatory mechanisms of transcription. The metazoan-specific Integrator is an RNA endonuclease that regulates Pol II–mediated transcription by cleaving a variety of RNAs. Integrator is composed of at least 14 subunits, INTS1 to INTS14, with a molecular weight greater than 1.4 MDa; mechanistic studies are limited to a few structures of isolated domains. Phosphorylation of 52 heptad repeats of $Y_1S_2P_3T_4S_5P_6S_7$ of human Pol II CTD at Ser^2, Ser^5, and Ser^7 involves multistep transcriptional regulation, and a few phosphatases have been identified (Y, Tyr; S, Ser; P, Pro; T, Thr). The serine-threonine protein phosphatase PP2A represents the majority of phosphatase activities in many human tissues and plays a crucial role in a variety of cellular processes. PP2A indirectly regulates gene expression by targeting various signaling pathways. However, it remains elusive whether PP2A directly regulates transcription.

RATIONALE: In a study of PP2A functions in the nucleus, we found that Integrator associates with the PP2A core enzyme formed by PP2A-A (scaffold subunit) and PP2A-C (catalytic subunit). We reconstituted the human Integrator-PP2A complex and determined the cryo–electron microscopy structure. The chromatin localization of the complex was detected by chromatin immunoprecipitation sequencing. Biochemical and cellular analyses were performed to investigate the binding of Integrator-PP2A to Pol II and dephosphorylation of Pol II CTD. The effect of the phosphatase and endonuclease of INTAC on transcription was tested by structure-guided functional analyses.

RESULTS: We identified a stable Integrator-containing PP2A-AC complex, which we term INTAC. The 3.5-Å-resolution structure shows that nine human Integrator subunits and PP2A core enzyme assemble into a rigid four-module complex. A cruciform-shaped central scaffold is formed by the backbone (INTS1-INTS2-INTS7) and shoulder (INTS5-INTS8) modules. The phosphatase (INTS6 and PP2A-AC) and endonuclease (INTS4-INTS9-INTS11) modules flank the opposite sides, with the phosphatase (PP2A-C) and endonuclease (INTS11) subunits being apart from each other by as much as ~150 Å.

Canonical PP2A holoenzyme consists of the PP2A core enzyme and a single regulatory subunit (PP2A-B). By contrast, INTAC functions as a noncanonical PP2A holoenzyme and Integrator serves as a multicomponent regulatory subunit, by which the PP2A core enzyme is recruited to chromatin-associated Pol II and dephosphorylates Pol II CTD at Ser^2, Ser^5, and Ser^7. The phosphatase of INTAC suppresses transcription of coding genes and noncoding elements and affects transcription initiation, pausing, and elongation. By contrast, the endonuclease of INTAC mainly modulates the pausing-elongation transition and plays a more evident role in regulating the transcription of small nuclear RNAs. Thus, consistent with their architectural separation in the INTAC structure, the RNA endonuclease and phosphatase play distinct roles in the regulation of Pol II–mediated transcription.

CONCLUSION: We present the identification, structural visualization, and functional characterization of the Integrator-PP2A complex, a noncanonical PP2A holoenzyme. Our study provides a direct connection between PP2A-mediated dephosphorylation and transcriptional regulation, two of the most fundamental cellular processes, and reveals how dual enzymatic activities are structurally and functionally integrated into the INTAC complex. ■



Modular organization and function of INTAC complex. Schematic model of canonical PP2A holoenzymes (top left) and the INTAC complex (top right), which indirectly and directly regulate transcription, respectively. A, B, and C represent the scaffold (PP2A-A), regulatory (PP2A-B), and catalytic subunit (PP2A-C) of a canonical PP2A holoenzyme, respectively, and the red star indicates the phosphatase catalytic cavity; on the right, numbers indicate Integrator subunits. (Bottom) INTAC suppresses transcription through binding Pol II and dephosphorylating Pol II CTD at Ser^2, Ser^5, and Ser^7. The loss of phosphatase activity increases the levels of Pol II phosphorylation and leads to dysregulated transcription of INTAC target genes. pSer2, pSer5, and pSer7 represent the phosphorylated forms of Ser^2, Ser^5, and Ser^7; the numbers indicate Integrator subunits. P-TEFb, the positive transcription elongation factor b.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: feixchen@fudan.edu.cn (F.X.C.); xuyh@fudan.edu.cn (Y.X.)
Cite this article as H. Zheng *et al.*, *Science* **370**, eabb5872 (2020). DOI: 10.1126/science.abb5872

RESEARCH ARTICLE

TRANSCRIPTION

Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase

Hai Zheng[1,2]*, Yilun Qi[1]*, Shibin Hu[1]*, Xuan Cao[3]*, Congling Xu[1]*, Zhinang Yin[4], Xizi Chen[1], Yan Li[1], Weida Liu[1], Jie Li[1], Jiawei Wang[5], Gang Wei[3], Kaiwei Liang[4], Fei Xavier Chen[1]†, Yanhui Xu[1,2,6]†

The 14-subunit metazoan-specific Integrator contains an endonuclease that cleaves nascent RNA transcripts. Here, we identified a complex containing Integrator and protein phosphatase 2A core enzyme (PP2A-AC), termed INTAC. The 3.5-angstrom-resolution structure reveals that nine human Integrator subunits and PP2A-AC assemble into a cruciform-shaped central scaffold formed by the backbone and shoulder modules, with the phosphatase and endonuclease modules flanking the opposite sides. As a noncanonical PP2A holoenzyme, the INTAC complex dephosphorylates the carboxy-terminal repeat domain of RNA polymerase II at serine-2, -5, and -7 and thus regulates transcription. Our study extends the function of PP2A to transcriptional regulation and reveals how dual enzymatic activities—RNA cleavage and RNA polymerase II dephosphorylation—are structurally and functionally integrated into the INTAC complex.

The metazoan-specific Integrator binds the C-terminal domain (CTD) of RNA polymerase II (Pol II) and functions as an RNA endonuclease to cleave different classes of RNAs, thereby regulating the transcription of both protein-coding genes and noncoding elements (*1*–*14*). Dysregulation of Integrator subunits has been found in severe neurodevelopmental syndromes and cancer (*1*, *15*–*17*). Integrator is composed of at least 14 subunits, INTS1 to INTS14, with a molecular weight greater than 1.4 MDa (*9*, *18*). The RNA endonuclease subunit INTS11 binds a catalytically inactive homolog INTS9 and a scaffold subunit INTS4 to form a cleavage module (*19*). INTS3 and INTS6 form another complex called the sensor of single-stranded DNA complex (SOSS) that is involved in the DNA damage response (*20*, *21*). It is well known that Integrator primarily functions as an endonuclease. However, endonuclease-independent functions of Integrator remain elusive. Despite its functional importance, the mechanistic studies of Integrator are limited to a few structures of isolated modules (*21*, *22*) owing to its compositional complexity and sample scarcity.

The dynamic regulation of phosphorylation and dephosphorylation of Pol II CTD is critical for multiple steps of eukaryotic transcription (*23*–*26*). In mammals, phosphorylation of 52 heptad repeats of $Y_1S_2P_3T_4S_5P_6S_7$ of Pol II CTD at Ser^5 and Ser^2 plays a regulatory role in controlling transcription initiation and elongation, respectively (Y, Tyr; S, Ser; P, Pro; T, Thr) (*23*, *26*, *27*). Unlike the extensive studies of phosphorylation, our understanding of Pol II CTD dephosphorylation is limited to fewer studies of several identified phosphatases, such as FCP1 (*28*), SSU72 (*29*), and RPAP2 (*30*, *31*). Here, we provide structural and biochemical evidence showing that protein phosphatase 2A (PP2A) core enzyme stably associates with Integrator and that the Integrator-PP2A complex regulates transcription through dephosphorylation of Pol II CTD.

PP2A accounts for as much as 1% of the total cellular protein in the brain and represents the majority of serine/threonine phosphatase activities in many tissues (*32*–*35*). PP2A is highly conserved across species and plays a crucial role in a variety of cellular processes—such as signal transduction, the cell cycle, cell proliferation, and neuronal physiology—and is dysregulated in some cancer and neurodegenerative diseases. The physiologically relevant PP2A core enzyme (PP2A-AC) consists of a catalytic (PP2A-C) subunit and a scaffold (PP2A-A) subunit (*32*, *36*). The PP2A holoenzyme is formed by PP2A-AC and one of at least 26 substrate-binding regulatory subunits, which are grouped into four subfamilies: B (B55/PR55), B′ (B56/PR61), B″ (PR48/PR72/PR130), and B‴ (PR93/PR110) (*34*, *37*). Previous studies have shown that PP2A could influence gene expression by targeting various signaling pathways (*38*–*42*). However, it remains elusive whether PP2A directly regulates transcription.

Here, we identified the Integrator-containing PP2A-AC (INTAC) complex and determined its structure at near-atomic resolution. As a noncanonical PP2A holoenzyme, INTAC dephosphorylates Pol II CTD and thus regulates transcription. Our study connects PP2A-mediated dephosphorylation and transcriptional regulation, two of the most fundamental cellular processes.

PP2A core enzyme forms a stable complex with Integrator

To explore the nuclear function of PP2A, we sought to identify PP2A-binding proteins in the nucleus. The Protein A–tagged PP2A-A and PP2A-C were individually overexpressed in human embryonic kidney (HEK) Expi293 cells, which was followed by nuclear extraction, purification, and mass spectrometry analyses (fig. S1A). Notably, most of the Integrator subunits were copurified with PP2A-A and PP2A-C (Fig. 1A). Ectopic expression and affinity purification of two representative Integrator subunits confirmed the binding of Integrator to the PP2A core enzyme but not to any known PP2A regulatory subunits (Fig. 1A and fig. S1A). The Integrator-PP2A interaction was further validated by reciprocal coimmunoprecipitation with antibodies against endogenous PP2A core enzyme and representative Integrator subunits (INTS1, INTS3, INTS5, and INTS11) followed by Western blotting (Fig. 1B).

To obtain a purified complex for further biochemical and structural analyses, the human Integrator subunits were cotransfected and overexpressed in HEK Expi293 cells. The Integrator-PP2A complex was purified to homogeneity by immunoprecipitation followed by density gradient centrifugation. Considerable amounts of endogenous PP2A-A and PP2A-C and all the Integrator subunits comigrated to the peak fractions (Fig. 1C and fig. S1B, lanes 13 to 15), whereas Integrator subcomplexes lacking PP2A-AC migrated to lower–molecular weight fractions (fig. S1B, lanes 7 to 11). The above analyses suggest the existence of a stable complex, which we term INTAC.

Chromatin localization of the INTAC complex

We next examined whether PP2A and Integrator form a complex on chromatin. The cellular fractionations confirmed evident chromatin localization of representative Integrator subunits (*43*) and showed noticeable chromatin localization of PP2A (Fig. 1D). The relative proportion of PP2A on chromatin is less than that of Integrator subunits, consistent with the major cytoplasmic distribution of PP2A (*44*, *45*)

[1]Fudan University Shanghai Cancer Center, Institutes of Biomedical Sciences, State Key Laboratory of Genetic Engineering and Shanghai Key Laboratory of Medical Epigenetics, Shanghai Medical College of Fudan University, Shanghai 200032, China. [2]The International Co-laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, China, Department of Systems Biology for Medicine, School of Basic Medical Sciences, Shanghai Medical College of Fudan University, Shanghai 200032, China. [3]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [4]Department of Pathophysiology, School of Basic Medical Sciences, Wuhan University, Wuhan 430071, China. [5]State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, 100084 Beijing, China. [6]Human Phenome Institute, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200433, China.
*These authors contributed equally to this work.
†Corresponding author. Email: feixchen@fudan.edu.cn (F.X.C.); xuyh@fudan.edu.cn (Y.X.)

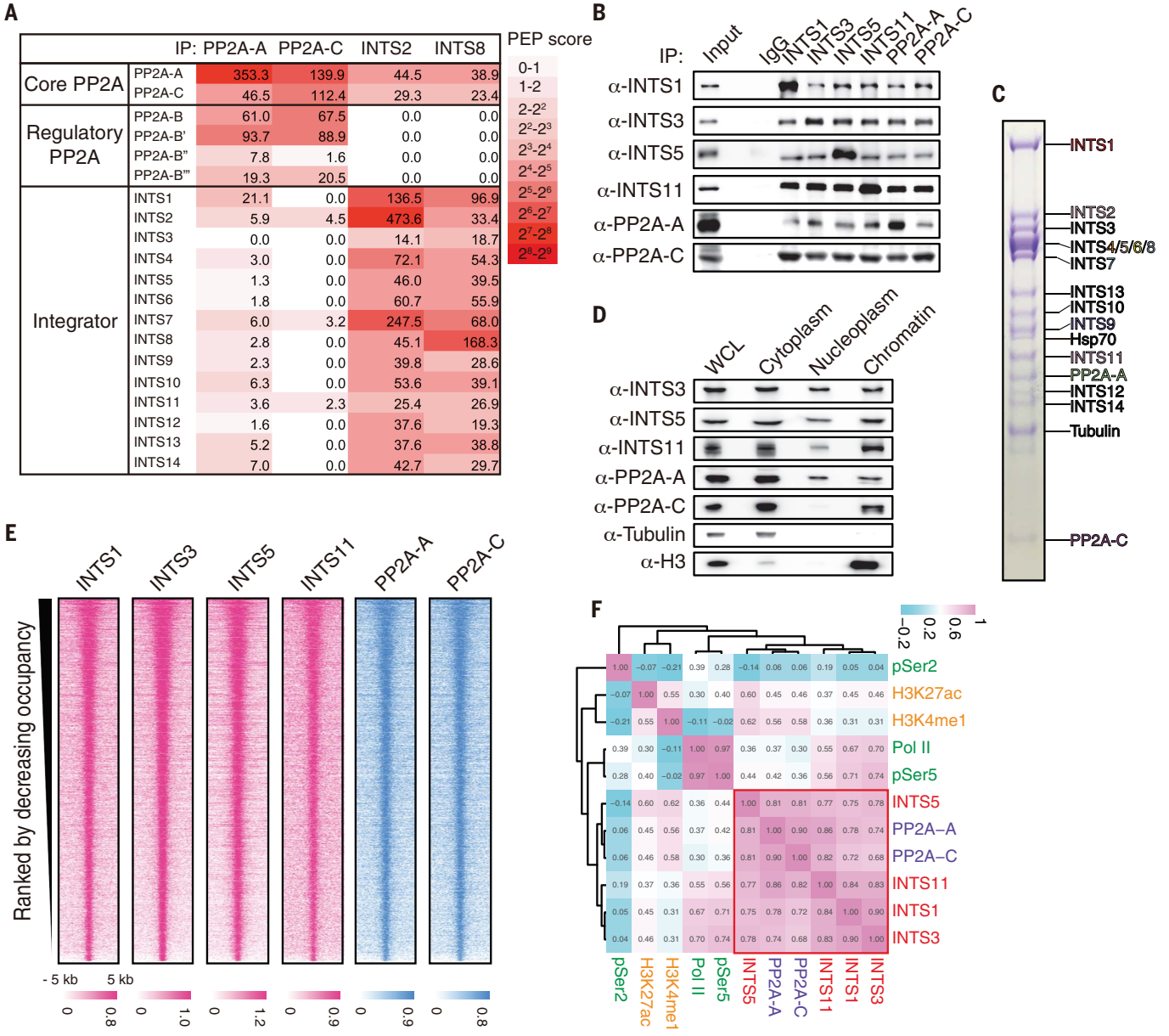


Fig. 1. Identification, characterization, and genome-wide profiling of the INTAC complex. (**A**) Mass spectrometry analyses of Protein A–tagged PP2A-A, PP2A-C, INTS2, and INTS8. The PEP score represents the sum of posterior error probability (PEP). IP, immunoprecipitation. (**B**) Coimmunoprecipitation and Western blotting of endogenous Integrator subunits and PP2A core enzyme confirming the interaction between Integrator and PP2A-AC. (**C**) The purified INTAC complex was subjected to SDS-PAGE (SDS–polyacrylamide gel electrophoresis) followed by Coomassie blue staining. (**D**) Subcellular fractionation and Western blotting of INTAC subunits showing its localization on chromatin. WCL, whole-cell lysate. (**E**) Heatmaps of genome-wide occupancy of INTS1, INTS3, INTS5, INTS11, PP2A-A, and PP2A-C. Color-scaled intensities are in units of log2 reads per million. (**F**) Hierarchical clustering of the correlation between genomic occupancy (peaks) for INTAC subunits, total and phosphorylated Pol II, H3K4me1, and H3K27ac. Numbers indicate Pearson correlation coefficients. The red box highlights the correlation coefficient values between INTAC subunits.

and the primary chromatin localization of Integrator (*43*). We conducted chromatin immunoprecipitation sequencing (ChIP-seq) of PP2A-A, PP2A-C, and representative Integrator subunits and compared their genomic distributions with that of Pol II and histone modifications. Heatmaps ranked by decreasing Integrator occupancy show a comparable genomic occupancy of Integrator subunits and PP2A-AC (Fig. 1E), which bind both promoter and enhancer regions (fig. S1C). Notably, the Pearson correlation coefficient (PCC) values are the highest between PP2A and Integrator subunits compared with that between INTAC and Pol II or histone modifications (Fig. 1F), consistent with the presence of a stable Integrator-PP2A complex.

Structure determination and overall structure of the INTAC complex

We next determined the cryo–electron microscopy (cryo-EM) structure of INTAC using single-particle reconstruction. The cryo-EM map was refined to an overall resolution of 3.5 Å, and locally refined maps were obtained to support model building (Fig. 2 and figs. S2 to S4). The structural model was mainly built ab initio with available structural templates (*22*, *46*, *47*) as references (fig. S3, table S1, and movies S1 to S7). The INTS3, INTS10, INTS12, INTS13, and INTS14 subunits were observed in the peak fractions in the density gradient centrifugation (fig. S1B, lanes 13 to 15) but

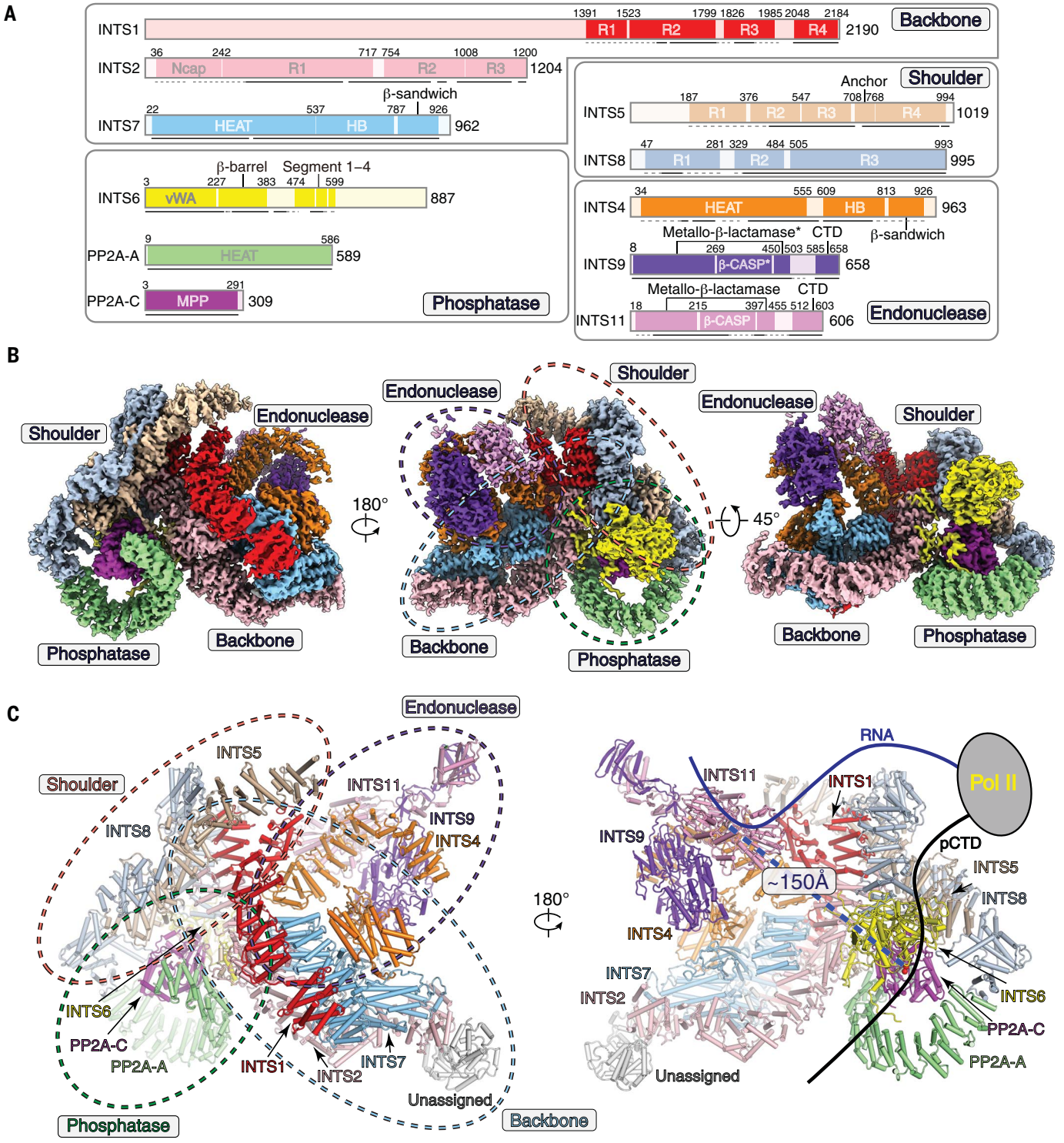


Fig. 2. Overall structure of the INTAC complex. (**A**) Schematic modular organization and domain structures. Residues at domain boundaries are indicated. Solid and dashed lines indicate regions that were modeled with residues and poly-alanine, respectively. The color scheme is indicated and used throughout all figures. R1, R2, R3, and R4, repetitive helix hairpin domains; HEAT, HEAT repeat domain; HB, helix bundle domain; MPP, metallophosphatase domain. The * represents an inactive catalytic domain of INTS9. (**B**) The 3.5-Å-resolution cryo-EM map of the INTAC complex in three different views, with subunit surfaces colored as in (A). The four modules are indicated with dashed circles. (**C**) Cartoon model of the INTAC structure shown in two different views. The putative binding sites of RNA transcript on INTS11 and phosphorylated CTD (pCTD) of Pol II on PP2A-C are indicated. The zinc and magnesium cations at the catalytic centers of INTS11 and PP2A-C are shown as gray and red spheres, respectively, and their distance is indicated by the blue dashed line.

were not assigned in the structural model, suggesting that these subunits were invisible in the cryo-EM map because of flexibility. The unassigned model (Fig. 2C) and the weak peripheral cryo-EM density (fig. S3E) may be derived from these missing subunits and/or unmodeled regions.

The cryo-EM map reveals a compact fold with approximate dimensions of ~280 Å by 260 Å by 150 Å and can be divided into four modules, including backbone (INTS1-INTS2-INTS7), shoulder (INTS5-INTS8), phosphatase (INTS6 and PP2A-AC), and endonuclease (INTS4-INTS9-INTS11) modules (Fig. 2, B and C). The shoulder and backbone modules form a central cruciform scaffold, and the phosphatase and endonuclease modules flank the opposite sides. The phosphatase (PP2A-C) and endonuclease (INTS11) subunits are apart from each other by as much as ~150 Å (Fig. 2C). INTS2 is a scaffold subunit of the backbone module and mediates intermodular contacts through binding the shoulder and phosphatase modules. INTS2, INTS5, and INTS8 together grasp the phosphatase module. INTS7 and INTS4 mediate contacts between the backbone and endonuclease modules. Consistent with the structural rigidity, the majority of the INTAC complex consists of tandemly repeated α helices, huntingtin-elongation-A subunit-TOR (HEAT) repeats, and β-sandwich and mixed α/β domains (Fig. 2A).

The backbone and shoulder modules form a cruciform scaffold

In the backbone module, INTS1, INTS2, and INTS7 closely associate with each other to assemble a rigid scaffold (Fig. 3, A and B). INTS2 adopts an elongated conformation and consists of an N-terminal helical cap (Ncap) and three tandemly repeated α-helix hairpins (denoted INTS2R1 to INTS2R3) (Fig. 3A, top left). The repeated helical regions of INTS1, INTS5, and INTS8 are similarly termed for simplicity (Fig. 2A). INTS2R1 exhibits an extended arch-shaped conformation, with the Ncap further capped by an unassigned helical repeat region (Fig. 3C). An INTS2R2-INTS2R3-INTS1R4 "bridge" extends to and binds the shoulder module (Fig. 3D). The rest of INTS1 (INTS1R1 to INTS1R3) adopts an arch-shaped conformation. INTS7 consists of an N-terminal superhelical α solenoid, an eight-helix bundle, and a C-terminal β-sandwich domain (Fig. 3E). INTS1, INTS2, and INTS7 merge at a three-way junction and make extensive contacts with each other, nucleating the assembly of the whole INTAC complex (Fig. 3, B, D, and F).

The shoulder (INTS5-INTS8) and the backbone modules form a cruciform scaffold (Fig. 3A). INTS5 consists of four helical repeats arrayed in a nearly straight conformation (Fig. 3G). INTS5R2 to INTS5R4 are wrapped by INTS8, creating a rigid shoulder module (Fig. 3, H and I). INTS8R3 adopts a superhelical conformation and wraps around INTS1R4, INTS2R3, and INTS5R2 and therefore enlaces the shoulder and backbone modules at the intersection (Fig. 3, A and I).

Assembly of the phosphatase module into the INTAC complex

The Integrator-PP2A contacts generate a buried surface area of ~2500 Å^2 and allow the PP2A core enzyme to be stably assembled into the Integrator complex (Fig. 4, A to C). INTS2 of the backbone and INTS5 and INTS8 of the shoulder together grasp the phosphatase module with an INTS5 loop (residues 708 to 768, denoted INTS5-anchor) (Fig. 3G) anchored into a groove between PP2A and INTS2R2 (Figs. 2 and 4, A and B). INTS8R1 merges with the first HEAT repeat (HEAT1) of PP2A-A and stabilizes this otherwise flexible end (*48*, *49*) (Fig. 4A and fig. S5). PP2A-C is stabilized within the phosphatase module through a network of intermolecular interactions between INTS5R4, INTS8R3, INTS6, and PP2A-C, which merge at a four-way junction (Fig. 4A). Patient-derived deletion (residues E^{973}VL^{975}; E, Glu; V, Val; L, Leu) of INTS8 around this site impaired the function of Integrator (*15*), suggesting a critical role of this contact.

Immunoprecipitation assays confirmed the essential role of the shoulder module in incorporating PP2A into INTAC (Fig. 4D). Endogenous PP2A associated with Protein A–tagged INTS2, the scaffold subunit of the INTAC complex, when all Integrator subunits were overexpressed (lane 11). The Integrator-PP2A association was impaired by excluding INTS5 (lane 12) or INTS8 (lane 14), supporting the critical roles of these shoulder subunits in the assembly of the phosphatase module into INTAC. By contrast, the deletion of INTS11 (lane 15) did not affect the Integrator-PP2A interaction, consistent with the architectural separation of the two catalytic modules (Fig. 2, B and C).

INTS6 facilitates assembly of PP2A into the INTAC complex

INTS6 binds the PP2A core enzyme and facilitates its assembly into INTAC (Fig. 4, A to C). Consistent with its structural role as a scaffold subunit of the phosphatase module, the deletion of INTS6 also led to the dissociation of PP2A from INTAC (Fig. 4D, lane 13). The von Willebrand factor type A domain (vWA) domain of INTS6 reveals a mixed α/β fold and binds INTS8R3 and the small helix domain of PP2A-C (Fig. 4, B and C, and fig. S6). Four segments (residues 474 to 599) of INTS6 sequentially wind through the surfaces of INTS2 and PP2A and glue the backbone module and PP2A together (Fig. 4B, left). Segment 1 packs against the surface of INTS2R3. Segment 2 binds the INTS5-anchor and is sandwiched by INTS2R2 and PP2A-AC. Segment 3 binds the β-sandwich domain of PP2A-C and merges with the edge of the catalytic cavity. Segment 4 packs against the ridge of HEAT repeats of PP2A-A.

The vWA and β-barrel domains of INTS6 adopt a rigid bilobular fold, which sits on top of PP2A-C and stands out of the phosphatase module (Fig. 4B and fig. S6). The β-barrel domain adopts a fold similar to that of DNA-binding protein Ku70 (*50*) and a phosphopeptide-binding SPOC domain (*51*), suggesting a role in mediating intermolecular interaction.

The INTAC complex is a noncanonical PP2A holoenzyme

The structure suggests that INTAC is a catalytically active and noncanonical PP2A holoenzyme (Fig. 4, A to C, and fig. S5). The PP2A core enzyme adopts a conformation similar to that in canonical PP2A holoenzymes (*33*, *34*), as exemplified by the AC-B/PR55 complex (*52*) (Fig. 4B, right). In the INTAC complex, PP2A-A exhibits a characteristic horseshoe-shaped conformation and PP2A-C packs against PP2A-A on the ridge of HEAT11-HEAT15 (Fig. 4A and fig. S5B). PP2A-C is cradled by INTS2, INTS5, INTS8, INTS6, and PP2A-A and is positioned in the center of the phosphatase module. The edge of the PP2A-C catalytic cavity merges with an INTS8R1 insert (INTS8-insert) (Fig. 4C), and the catalytic cavity faces toward a V-shaped open groove formed by PP2A-A and INTS6, showing a potential substrate entry groove distinct from that of the canonical PP2A holoenzymes (Fig. 4B and fig. S5).

In all the structures of PP2A holoenzymes solved to date, regulatory subunits associate with the ridge of N-terminal HEAT repeats of PP2A-A (*46*, *52*–*55*) and recruit substrates to PP2A-C (*52*, *55*) (Fig. 4B and fig. S5). By contrast, Integrator makes multiple contacts with PP2A-C but not PP2A-A. Structural comparison shows that regulatory subunits (B, B′, and B″) would clash with INTS6 or INTS8 in the context of INTAC, suggesting a mutually exclusive assembly of INTAC and the known PP2A holoenzymes. The structural analyses are consistent with the absence of regulatory subunits in the purified INTAC (Fig. 1, A and C). As described below, INTAC binds Pol II and dephosphorylates Pol II CTD (Fig. 5). Thus, INTAC represents a noncanonical PP2A holoenzyme in which Integrator binds the PP2A core enzyme and serves as a regulatory subunit to recruit substrates, such as Pol II, for dephosphorylation.

Integrator recruits PP2A to chromatin

We have shown chromatin localization of the INTAC complex (Fig. 1, D to F). The known chromatin and Pol II association of Integrator (*9*, *10*) suggests that Integrator recruits PP2A-AC to chromatin. To test this, we performed ChIP-seq of PP2A subunits, along with Pol II, in DLD-1 colorectal adenocarcinoma cells with the depletion of INTAC subunits using

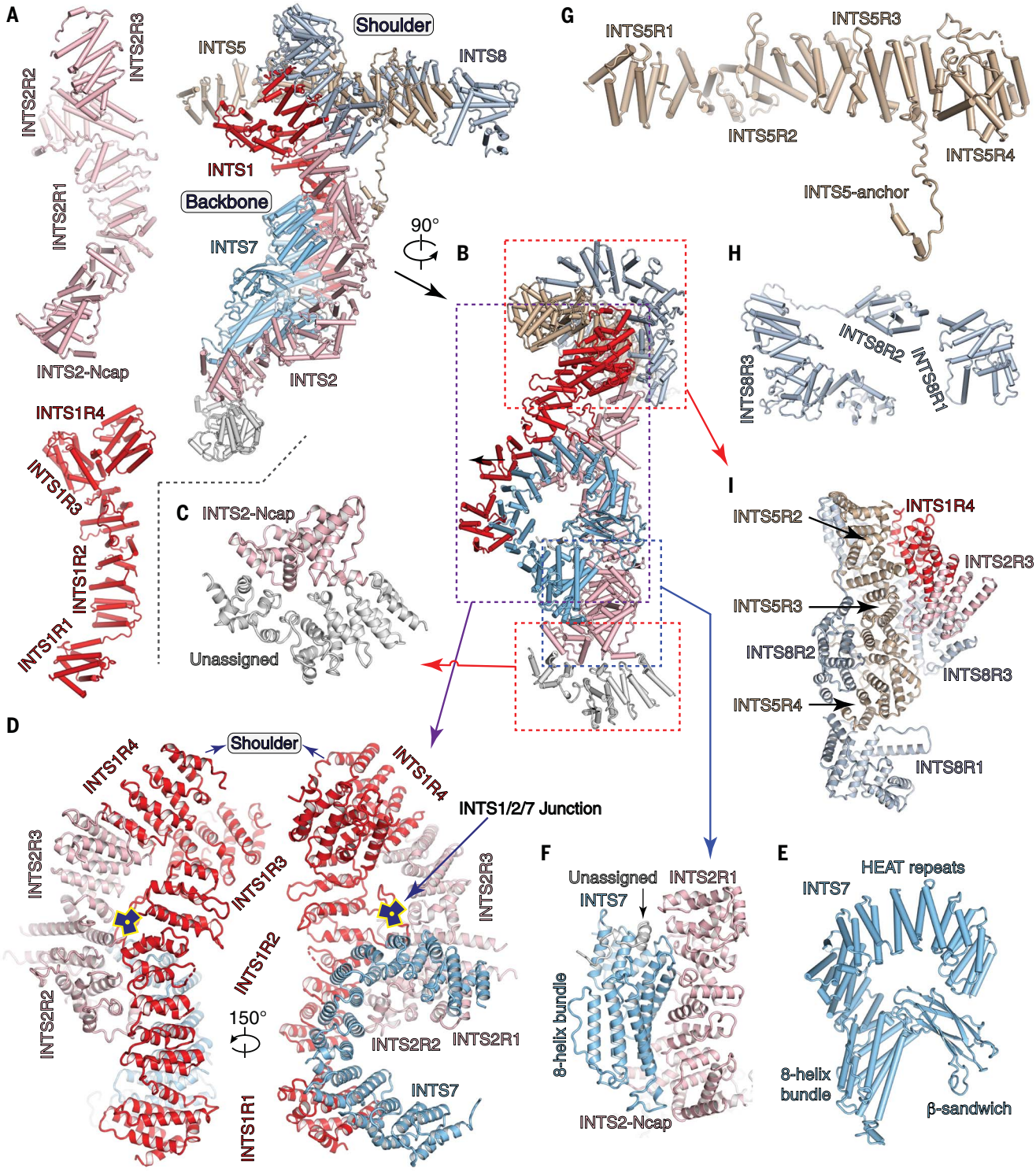


Fig. 3. Structure of the backbone and shoulder modules. (**A** and **B**) Cartoon model of the backbone and shoulder modules shown in two different views. The isolated INTS1 and INTS2 subunits are shown in (A). Locations of intermolecular contacts are highlighted in (B). (**C**, **D**, **F**, and **I**) Close-up views of intermolecular contacts. (**E**, **G**, and **H**) The isolated INTS7 (E), INTS5 (G), and INTS8 (H) subunits.

independent short hairpin RNAs (shRNAs) (fig. S7, A to C). Because Integrator-PP2A association requires INTS8 (the shoulder subunit), but not INTS11 (endonuclease subunit) (Fig. 4D), INTS8, INTS11, and PP2A-Cα/Cβ (double deletion) were depleted for analyses. Compared to that of Pol II, the relative occupancy of both PP2A-A (Fig. 4E) and PP2A-C (Fig. 4F) exhibited a dramatic decrease upon depleting PP2A-Cα/Cβ or INTS8 but not INTS11. However, attenuated expression of PP2A-Cα/Cβ had a subtle effect on the relative level of Integrator, as represented by INTS3 (fig. S7D). These data indicate that Integrator mediates the recruitment of PP2A to chromatin-associated Pol II but not vice versa.

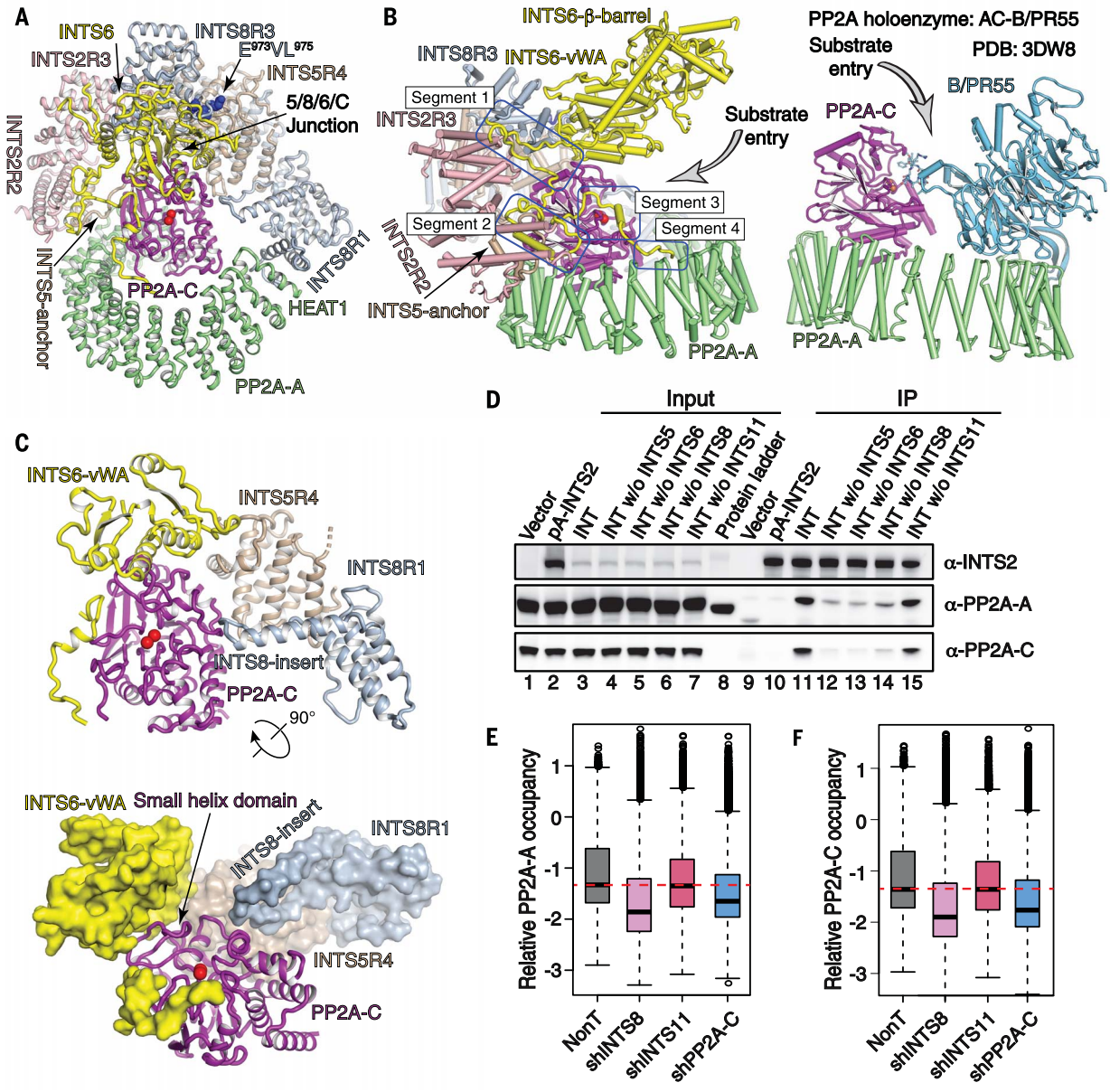


Fig. 4. The phosphatase module and its assembly into INTAC. (**A**) Cartoon model of the phosphatase module and the associated domains of INTS2, INTS5, and INTS8. The INTS6 β-barrel domain was omitted for clarity. Residues $E^{973}VL^{975}$ of INTS8 that are deleted in patients with a severe neurodevelopmental syndrome (*15*) are shown as blue spheres, which are near the four-way junction of INTS5, INTS8, INTS6, and PP2A-C. (**B**) Side view of the phosphatase module with four PP2A-binding segments of INTS6 indicated with blue boxes (left). The structure of PP2A holoenzyme AC-B/PR55 (PDB ID: 3DW8) (*52*) is shown for comparison (right). The putative substrate entry grooves are indicated with arrows. (**C**) Two views of intermolecular contacts around PP2A-C, with surrounding regions omitted for clarity. Integrator subunits are shown in surface representations (bottom). (**D**) Immunoprecipitation of Protein A–tagged INTS2 and PP2A-AC. The Protein A–tagged INTS2 alone (lane 10), Protein A–tagged INTS2 with all other untagged Integrator subunits (lane 11), or Protein A–tagged INTS2 with all but one indicated subunit (lanes 12 to 15) were overexpressed, immunoprecipitated, and subjected to Western blotting. PA, Protein A. (**E** and **F**) Boxplots showing the relative occupancy of PP2A-A (E) and PP2A-C (F) compared with Pol II in cells transduced with NonT, shINTS8, shINTS11, or shPP2A-C. Boxplot whiskers represent the minimal and maximal values of each sample; red dashed lines indicate the median value of NonT.

INTAC dephosphorylates Pol II CTD

Integrator has been known to bind Pol II on chromatin (*9*, *10*). To test whether Integrator within the INTAC complex could still bind Pol II, we performed an in vitro pulldown assay using purified INTAC, PP2A core enzyme, and unphosphorylated Pol II (Fig. 5A and fig. S8, A to D). The Integrator subunits and PP2A in the context of INTAC, but not PP2A core enzyme alone, were pulled out by the immobilized Pol II, indicating that Integrator in INTAC directly binds Pol II and bridges an indirect interaction between PP2A and Pol II. Interestingly, hyperphosphorylated Pol II was converted into the hypophosphorylated form after incubating with INTAC in the pull-down assay (Fig. 5B), suggesting a previously unknown phosphatase activity of PP2A, in the context of INTAC, against phosphorylated CTD of Pol II.

We next conducted an in vitro phosphatase assay using purified phosphatases and hyperphosphorylated Pol II as the substrate (Fig. 5C). The basal activities of these phosphatases

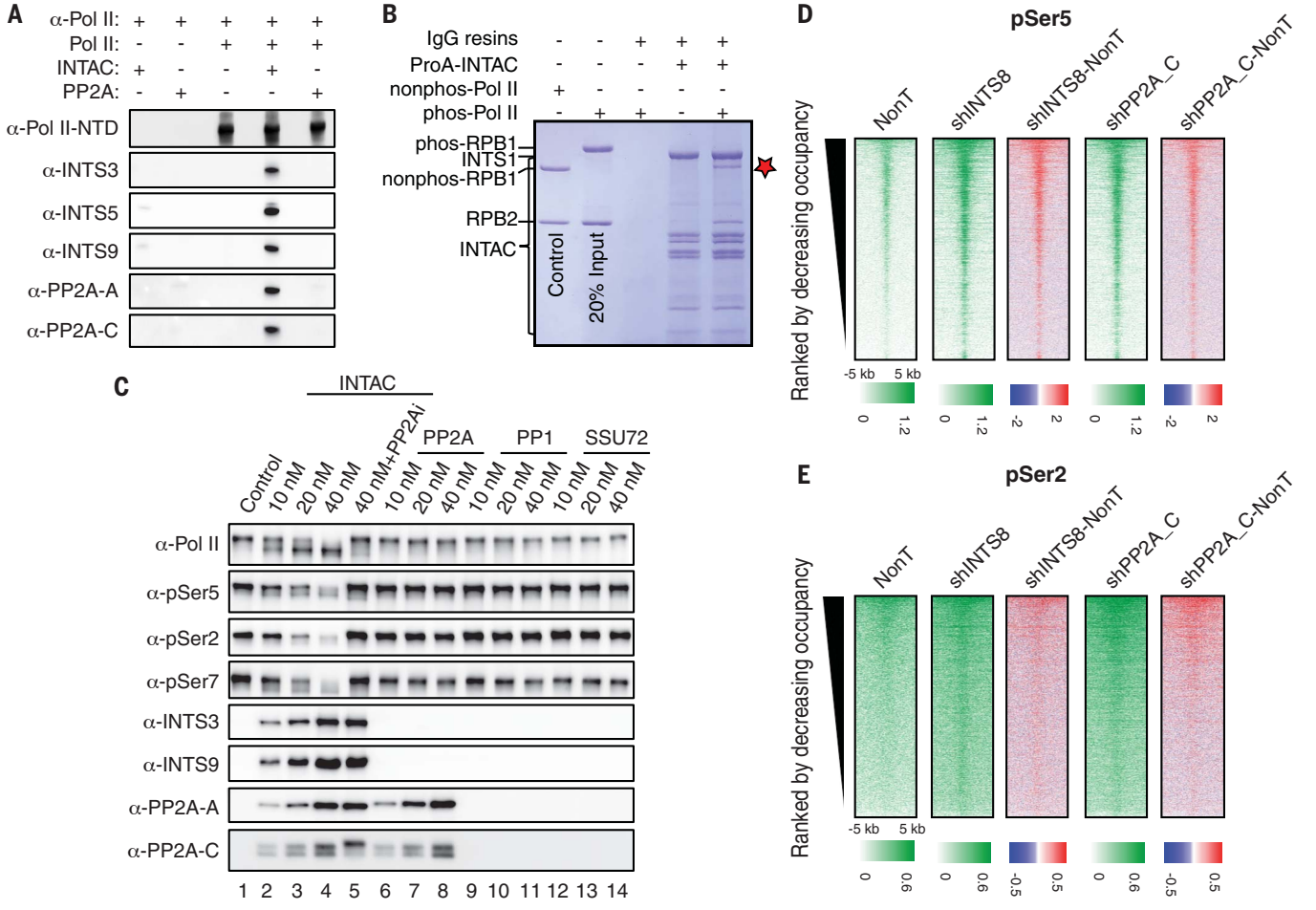


Fig. 5. Dephosphorylation of Pol II CTD by INTAC. (**A**) In vitro pulldown assay using Pol II antibody, purified unphosphorylated Pol II, INTAC holoenzyme, and PP2A core enzyme followed by Western blotting. (**B**) In vitro pulldown assay using purified Pol II and immobilized INTAC. The bound proteins were subjected to SDS-PAGE followed by Coomassie blue staining. Phosphorylated Pol II was pulled out and converted to the unphosphorylated form, as indicated with a red star. (**C**) In vitro phosphatase assay using purified INTAC, PP2A core enzyme, PP1, and SSU72 as enzymes and hyperphosphorylated Pol II treated by TFIIH as substrate. The reactions were subjected to Western blotting using the indicated antibodies. The enzyme concentrations are indicated above the blot. (**D** and **E**) Heatmaps of pSer^5 (D) and pSer^2 (E) genome-wide occupancy in cells transduced with NonT, shINTS8, or shPP2A-C. Rows are sorted by decreasing INTAC occupancy in the NonT condition. Color-scaled intensities are in units of reads per million mapped reads (RPM). For differential heatmaps, the color bars depict log2 values.

were confirmed by the reactions using a nonspecific phosphatase substrate, *para*-nitrophenyl phosphate (pNPP) (fig. S8E). Of note, INTAC efficiently converted Pol II from the hyperphosphorylated to the hypophosphorylated forms by dephosphorylating Ser^2, Ser^5, and Ser^7 of Pol II CTD in a dose-dependent manner (Fig. 5C, lanes 1 to 4). By contrast, PP2A core enzyme, PP1, and SSU72 showed undetectable phosphatase activity against Pol II CTD (lanes 6 to 14). The phosphatase activity of INTAC was blocked by the incorporation of catalytically inactive (H59K/H241K; H, His; K, Lys) PP2A-C (*56*) (fig. S8F) or cotreatment with microcystin-LR (MC-LR), a selective and potent inhibitor of PP2A (Fig. 5C, compare lanes 4 and 5). Thus, INTAC binds Pol II and dephosphorylates Pol II CTD at Ser^2, Ser^5, and Ser^7. It is known that PP2A holoenzymes dephosphorylate more than 300 substrates (*33*). The substrate of INTAC may not necessarily be restricted to Pol II CTD. Other transcriptional regulators, especially Integrator-interacting proteins, are potential substrates of the INTAC phosphatase.

To validate INTAC-mediated CTD dephosphorylation in cells, we conducted ChIP-seq and measured genome-wide levels of phosphorylated Pol II upon depleting PP2A-C or INTS8. To better annotate the phosphatase functions, we focused on phosphorylated Ser^5 and Ser^2 (pSer^5 and pSer^2), the two representative marks of transcription initiation and elongation, respectively (*23*, *26*, *27*). Notably, as exemplified by *CCND1*, *S100A10*, *EPDR1*, and *GDA* (fig. S9, A to D), attenuated expression of either INTS8 or PP2A-C led to increased levels of both pSer^5 and pSer^2 of Pol II at a genome-wide level (Fig. 5, D and E). These in vitro and in vivo results together reveal that phosphorylated Pol II is a bona fide substrate of INTAC, in which Integrator recruits Pol II to PP2A for dephosphorylation of the CTD.

INTAC regulates transcription of protein-coding genes and enhancers

The dynamic regulation of Pol II phosphorylation is coupled with and regulates the transitions between steps in the transcription cycle (*23*, *24*, *27*). To elucidate the impact of INTAC on transcription, we examined the levels of Pol II phosphorylation at promoters and enhancers and observed a substantial increase of pSer^5 and pSer^2 in both regions upon depletion of INTS8 or PP2A-C (Fig. 6A). As mentioned above, pSer^5 peaks at promoter regions and primarily represents initiating and paused Pol II, whereas pSer^2 is more enriched in gene bodies and mainly reflects elongating polymerase (*23*, *26*, *27*). Therefore, our results imply that depleting INTS8 or PP2A-C enhances

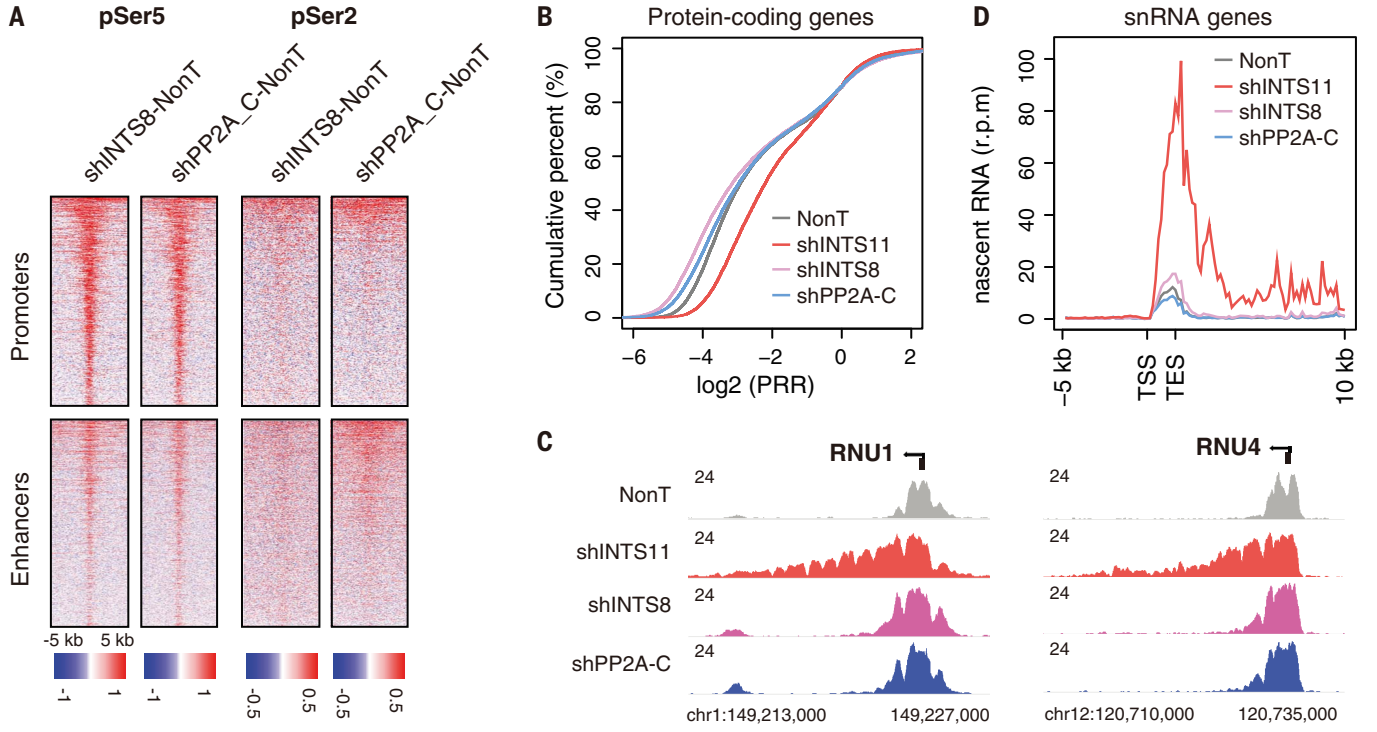


Fig. 6. Independent functions of the phosphatase and endonuclease modules in regulating transcription. (**A**) Heatmaps of pSer^5 and pSer^2 occupancy in cells transduced with NonT, shINTS8, or shPP2A-C at promoters and enhancers. Rows are sorted by decreasing INTAC occupancy in the NonT condition. Color-scaled intensities are in units of RPM. For differential heatmaps, the color bars depict log2 values. (**B**) Empirical cumulative distribution function plot of the PRR distribution in cells transduced with NonT, shINTS11, shINTS8, or shPP2A-C. PRR is calculated by the ratio of Pol II occupancy between the gene body and its promoter. The promoter is defined as the region covering −500 base pairs (bp) upstream to 500 bp downstream of the transcriptional start site (TSS); the gene body is defined as the region from 500 bp downstream of the TSS to the transcriptional end site (TES). (**C**) Representative genome browser track examples of total Pol II for snRNA genes *RNU1* and *RNU4*. The *x* axis indicates the chromosome position, and the *y* axis represents RPM values. chr, chromosome. (**D**) Metagene analysis of nascent RNA levels on snRNA genes in cells transduced with NonT, shINTS11, shINTS8, or shPP2A-C.

both transcription initiation and elongation. The quantification of nascent RNA levels reveals a major increase in expression of INTAC targets upon the impairment of its phosphatase activity by depleting PP2A-C or disruption of PP2A-Integrator interaction by depleting INTS8 (fig. S10A). Notably, the knockdown of INTS8 or PP2A-C caused an even greater increase in nascent RNA levels at enhancers than at promoters, unveiling a broad impact of INTAC in regulating transcription of both coding genes and noncoding elements.

To assess the extent of the impact on transcription initiation, promoter-proximal pausing, and elongation, we calculated the ratio of Pol II occupancy between the gene body and promoter, which is referred to as the Pol II release ratio (PRR) (*57*). The depletion of INTS8 or PP2A-C had a minor effect on PRR values (Fig. 6B), indicating that disrupting the phosphatase module leads to a comparable effect on initiation, pausing, and elongation. These results agree with the nonpreferential phosphatase activity on pSer^5 and pSer^2 and suggest that the phosphatase activity of INTAC functions as a general suppressor in transcriptional regulation.

Assembly of the endonuclease module into INTAC

In the endonuclease module, INTS11 and INTS9 are structurally similar and each consists of a bilobed endonuclease domain formed by metallo-β-lactamase and β-CASP domains followed by a carboxy-terminal dimerization domain (Fig. 7, A to C, and fig. S11, A to C). INTS11 possesses highly conserved residues arranged in an octahedral geometry for the chelation of two zinc cations (Fig. 7B and fig. S12). Such zinc coordination is essential for INTS11 endonuclease activity (*47*). By contrast, INTS9 has no such equivalent residues, consistent with its lack of catalytic activity (Fig. 7C and fig. S12). The two endonuclease domains (active in INTS11 and inactive in INTS9) form a pseudo-symmetric dimer and bind to each other in a back-to-back manner with the open catalytic pockets facing outward in opposite directions (Fig. 7A).

The carboxy-terminal dimerization domains of INTS11 and INTS9 are intertwined and dimerize into a β-barrel domain and an α/β domain (Fig. 7A, top right). The eight-stranded β-barrel domain is further stabilized by an α helix of INTS11 (Fig. 7A). As described previously (*22*), the C-terminal α/β domain is formed by a nine-stranded β sheet packing against four α helices on one side. The β-barrel and α/β domains flank out of the core of the endonuclease module and bring the C termini of the two catalytic domains into proximity, possibly supporting the architectural arrangement of INTS11 and INTS9.

INTS4 binds and stabilizes the INTS11-INTS9 heterodimer and bridges the endonuclease to scaffold modules (Fig. 7, A and D, and fig. S11A). INTS4 shares a similar domain composition to that of INTS7 but exhibits a distinct overall conformation (Figs. 3E and 7E and fig. S11, D and E). The crescent-shaped α solenoid of INTS4 associates with the dimeric endonuclease domains of INTS11 and INTS9 (Fig. 7, A and D). The eight-helix bundle and β-sandwich domain of INTS4 make extensive contacts with INTS7 and stabilize the β-CASP domain of INTS9 (Fig. 7, C and D).

The phosphatase and endonuclease modules play independent roles in controlling transcription

The architectural placement of the phosphatase and endonuclease modules suggests two

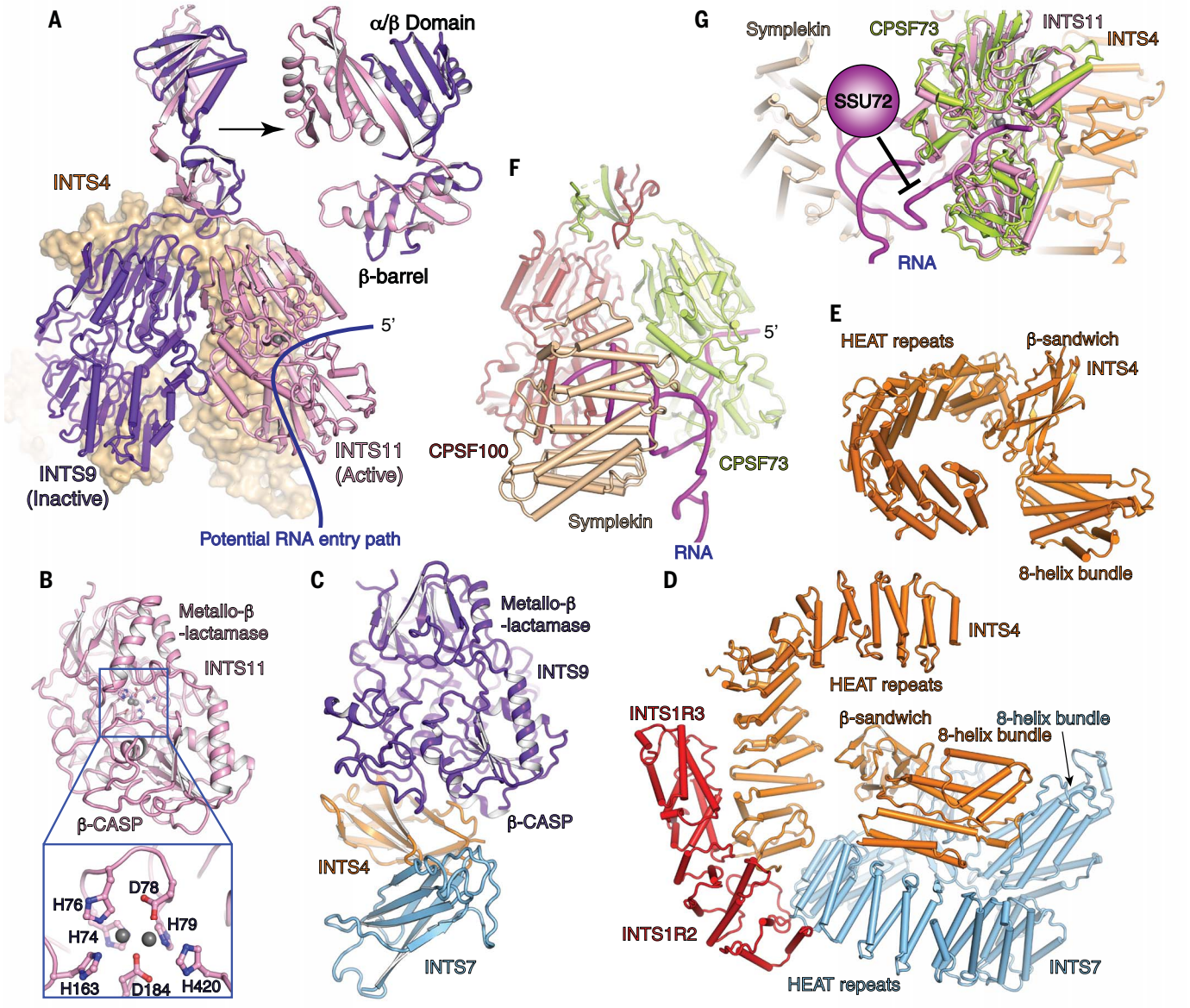


Fig. 7. Structures of the endonuclease modules and comparison with HCC structure. (**A**) Cartoon model of the INTS9-INTS11 dimer packing against INTS4 shown in surface representations. The close-up view of two carboxy-terminal dimerization domains in a different view is shown on the right. (**B** and **C**) Endonuclease domains of INTS11 (B) and INTS9 (C) shown in a similar view. A close-up view of the catalytic center of INTS11 is shown at the bottom of (B). The two zinc cations are shown as gray spheres, and critical residues involved in zinc coordination are shown as sticks. H, His; D, Asp. (**D**) The endonuclease-backbone interactions. (**E**) Structure of INTS4. (**F**) Cartoon model of symplekin-CPSF100-CPSF73 in HCC complex structure (*59*). The RNA substrate is colored in purple. CPSF73-CPSF100 is shown in an orientation similar to INTS11-INTS9 in the INTAC structure (A). (**G**) Structural comparison of symplekin-CPSF73 in HCC and INTS4-INTS11 in INTAC with CPSF73 and INTS11 superimposed. The two active pockets are similar, whereas symplekin and INTS4 are positioned on two opposite sides. SSU72 binds symplekin and inhibits RNA association (*58*), whereas PP2A-C (not shown) is far away from INTS11.

functionally separated activities in regulating transcription. To distinguish the functions of the phosphatase and endonuclease modules in transcription, we measured Pol II levels upon depleting endonuclease subunit INTS11. Interestingly, attenuation of INTS11 expression resulted in a compelling reduction of total Pol II and pSer^5 at a subset of promoters (fig. S10B), in contrast to the widespread increase in Pol II phosphorylation upon depleting PP2A or INTS8 (Fig. 5, D and E). The change of Pol II pattern by INTS11 knockdown reflects an alteration of the initiation-elongation transition or, more specifically, the release of paused Pol II into productive elongation. Indeed, the RNA endonuclease function of Integrator was recently shown to be involved in regulating Pol II pausing and its release by promoting early termination (*2*, *5*). In accordance, INTS11 knockdown induced an obvious increase in PRR values, indicating an enhanced release of Pol II from promoters to gene bodies (Fig. 6B).

We next investigated the role of INTAC phosphatase activity in transcription termination at small nuclear RNAs (snRNAs), a characteristic process that requires the endonuclease activity of INTS11. As exemplified by *RNU1* and *RNU4* (Fig. 6C), we observed prolonged snRNA transcription to about 10 kb downstream of

termination sites after depleting INTS11 (Fig. 6D and fig. S10C), confirming the function of the endonuclease module in promoting snRNA termination (*3*, *5*, *6*, *9*, *11*, *43*). By contrast, PP2A-C or INTS8 knockdown failed to cause evident termination defects of snRNA genes. These results indicate that the phosphatase activity of INTAC is not required for the endonuclease-mediated snRNA cleavage. Taken together, the RNA endonuclease (INTS11) and phosphatase (PP2A-C) play distinct roles in governing Pol II phosphorylation and pattern at protein-coding genes, enhancers, and snRNAs (Figs. 5 and 6, and fig. S10), consistent with their architectural separation in the INTAC structure.

Coexistence, but distinct architectural arrangement, of endonuclease and phosphatase in RNA-cleavage complexes

The coexistence of RNA endonuclease and protein phosphatase in the INTAC complex is reminiscent of the cleavage and polyadenylation specificity factor (CPSF), a pre-mRNA 3′-end–processing complex, and histone pre-mRNA cleavage complex (HCC). The CPSF and HCC both consist of symplekin-CPSF100-CPSF73, an RNA cleavage module equivalent to INTS4-INTS9-INTS11 (*19*, *22*, *58*, *59*). The structure of the dimeric endonuclease of INTS11-INTS9 is similar to that of RNA-bound CPSF73-CPSF100 (*58*) (Fig. 7F and fig. S11, B and C), suggesting a similar substrate entry path and RNA binding pocket (Figs. 2C and 7A).

Despite the shared endonuclease dimer, structural comparison of INTAC and HCC reveals considerable differences (Fig. 7, F and G). In the HCC complex, SSU72, another Pol II CTD phosphatase (*29*), directly binds symplekin (INTS4 equivalent) and inhibits RNA cleavage owing to steric clash (*58*, *60*, *61*), indicating a direct cross-talk between the endonuclease and phosphatase subunits and a phosphatase activity–independent function. By contrast, the RNA endonuclease and protein phosphatase subunits of INTAC are structurally apart from each other and functionally independent in transcriptional regulation. The differences between INTAC and HCC suggest distinct requirements of the dual-enzyme complexes in different cellular contexts.

A model of INTAC-mediated transcriptional regulation

Here, we present the identification, structural visualization, and functional characterization of a previously unknown, noncanonical Integrator-containing PP2A holoenzyme (Fig. 8). As a replacement of the single-protein regulatory subunit in canonical PP2A holoenzyme (Fig. 8A), the multisubunit Integrator in INTAC recruits Pol II to PP2A-C for dephosphorylation of the CTD (Fig. 8B). The depletion of phosphatase subunits or dissociation of PP2A from INTAC leads to an aberrant increase in Pol II phosphorylation and thus dysregulation of transcription (Fig. 8C). The phosphatase module of INTAC affects initiation, pausing, and elongation. By contrast, the endonuclease mainly modulates the pausing-elongation transition and plays a more evident role in regulating the transcription of snRNAs. Although the phosphatase and endonuclease function independently, the two modules within the INTAC complex may have coordinated functions under some transcriptional processes yet to be discovered. The enrichment of INTAC on genomic regions with higher Pol II occupancy also suggests that the two enzymatic activities should be tightly regulated to avoid uncontrolled suppression of transcription by INTAC.

Although PP2A has been linked to gene expression by modulating multiple signaling pathways (*38*–*42*), our findings reveal that PP2A, in the context of INTAC, has a broader impact on transcription by directly communicating with Pol II in addition to the targeted pathways (Fig. 8, A and B). The well-established dogma of PP2A-regulated gene expression and dysregulation in human diseases needs to be revised by adding the impact of INTAC-directed gene expression. Moreover, therapeutic targeting of PP2A through reactivation or blockade of the phosphatase activity needs to consider the effect of INTAC-mediated direct changes to transcription. Our structure also provides a framework for further investigation of separated and/or coordinated functions of the phosphatase and RNA endonuclease activities in modulating transcription under various physiological and pathological conditions.

Materials and methods

Antibodies and cell culture

Antibodies were as follows: INTS1 (A300-361A, Bethyl Laboratories), INTS2 (sc-514945, Santa Cruz), INTS3 (16620-1-AP, Proteintech), INTS5 (14069-1-AP, Proteintech), INTS8 (18802-1-AP, Proteintech), INTS9 (#13945, Cell Signaling), INTS11 (A301-274A, Bethyl Laboratories), PP2A-A (18802-1-AP, Proteintech), PP2A-C (18802-1-AP, Proteintech), H3K4me1 (A2355, Abclonal), H3K27ac (A7253, Abclonal), Pol II (NTD) (#14958, Cell Signaling), Pol II (pSer5) (#13523, Cell Signaling), Pol II (pSer2) (#13499, Cell Signaling), and Tubulin (AC008, Abclonal). 293T, DLD-1, and mouse embryonic fibroblast (MEF) cells were grown in Dulbecco's minimum essential medium (DMEM) supplemented with 10% fetal bovine serum (FBS). HEK Expi293 cells were grown in suspension in serum-free medium.

Cloning and protein expression

The full-length INTS1 to INTS14 open reading frames were separately cloned into a modified pCAG vector and INTS2, INTS3, INTS4, and INTS10 were tagged with an N-terminal Flag-4×Protein A. All plasmids were cotransfected to HEK Expi293 cells using polyethylenimine (PEI) (Polysciences) to a final concentration of 3 mg/liter. After being cultured at 37°C for 72 hours, cells were harvested for lysis and purification. PP2A-A, PP2A-C, and SSU72 were overexpressed in a similar way, separately. After being cultured at 37°C for 72 hours, cells were harvested for lysis and purification. In addition, PP1 was expressed in *Escherichia coli* cells (*62*), and the cells were harvested after cultivation overnight for further lysis and purification.

Protein purification

All the purification steps were performed at 4°C unless otherwise stated. The INTAC complex was purified from 16 liters of HEK Expi293 expression. Cell pellets were resuspended and lysed in lysis buffer containing 50 mM HEPES pH 7.4, 200 mM NaCl, 0.2% CHAPS, 5 mM $MgCl_2$, 5 mM adenosine triphosphate (ATP), 10% glycerol, 2 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 μg/ml Aprotinin, 1 μg/ml Pepstatin, 1 μg/ml Leupeptin for 30 min and cleared by centrifugation at 16,000 rotations per minute (rpm) for 30 min. The supernatant was incubated with immunoglobulin G (IgG) resins overnight. The resins were washed with buffer containing 50 mM HEPES pH 7.4, 200 mM NaCl, 0.1% CHAPS, 10% glycerol, 2 mM DTT. After on-column cleavage for 4 hours, the immobilized proteins were eluted out and concentrated for further purification by density gradient sedimentation. The concentrated proteins were layered on top of a 4-ml 8-40% (v/v) glycerol gradient in buffer containing 20 mM HEPES pH 7.4, 200 mM NaCl, 0.05% CHAPS, 2 mM DTT and centrifuged at 34,000 rpm for 16 hours. The fractions, 200 μl each, were collected manually from the top of the gradient and analyzed by a 4-12% Bis-Tris gel followed by Coomassie blue staining. Peak fractions corresponding to the INTAC complex were pooled and concentrated to 1 to 2 mg/ml accompanied with the removal of glycerol. The purified complex was used for cryo-EM sample preparation and in vitro assays. All subunits of the INTAC complex were confirmed by mass spectrometric analysis.

For the purification of the PP2A core enzyme, cell pellets of PP2A-A and PP2A-C were mixed and lysed in lysis buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.2% CHAPS, 2 mM $MgCl_2$, 0.2 mM EDTA, 10% glycerol, 2 mM DTT, 1 mM PMSF, 1 μg/ml Aprotinin, 1 μg/ml Pepstatin, 1 μg/ml Leupeptin for 30 min and cleared by centrifugation at 16,000 rpm for 30 min. The affinity purification was performed as described above. The protein was further purified by ion exchange chromatography (Mono Q 5/5, GE Healthcare). Peak fractions containing the PP2A core enzyme were pooled and concentrated to ~1 mg/ml, aliquoted, snap frozen,

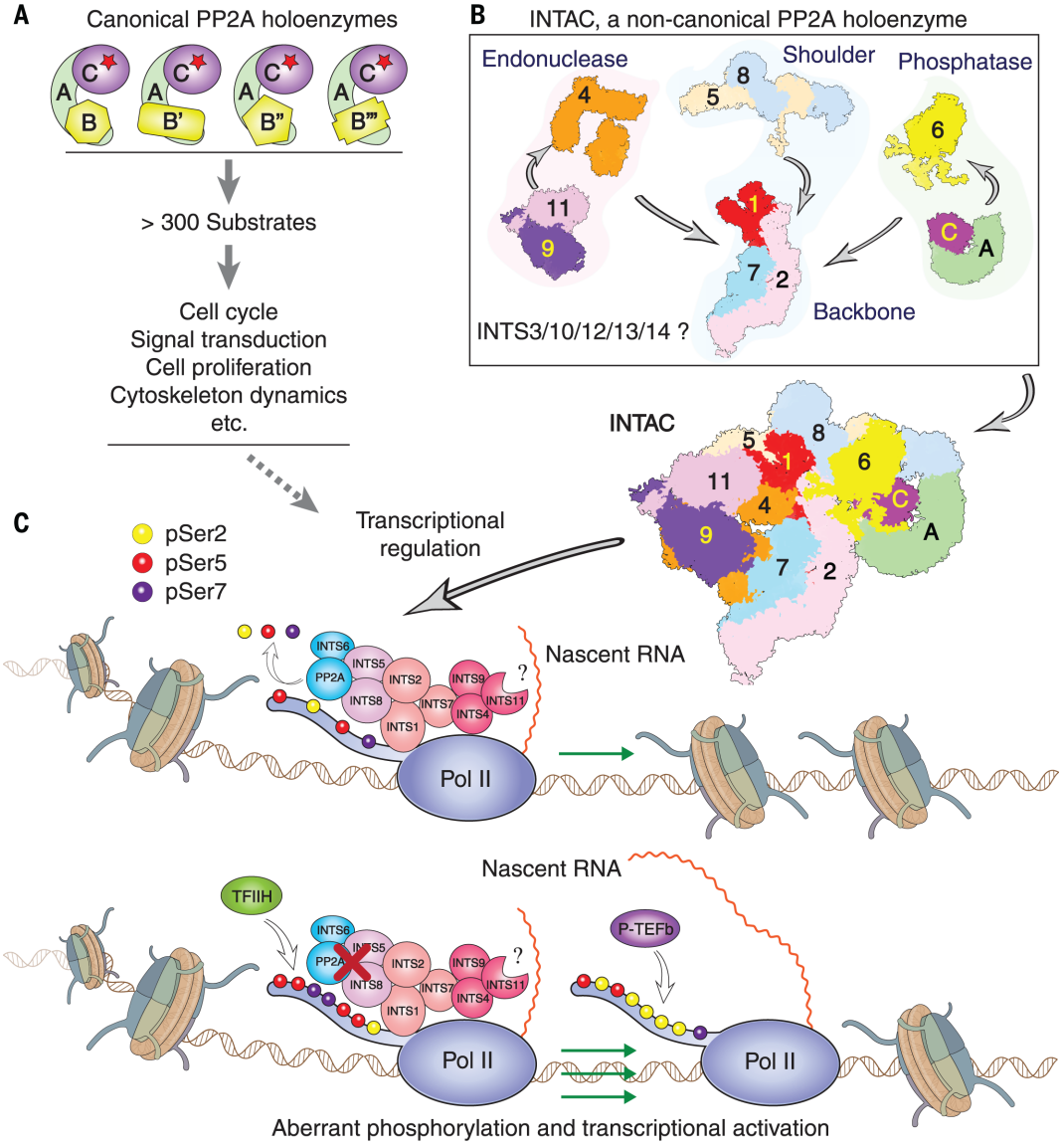


Fig. 8. Modular organization and a working model for transcriptional regulation by INTAC. (**A**) Schematic model of canonical PP2A holoenzymes, which dephosphorylate more than 300 substrates and regulate various cellular processes but do not directly regulate transcription. A, B, and C indicate PP2A scaffold (PP2A-A), regulatory (PP2A-B), and catalytic subunit (PP2A-C), respectively. The red star indicates the phosphatase catalytic site. (**B**) Modular organization of the INTAC complex. A and C represent PP2A-A and PP2A-C, respectively. The numbers indicate INTS subunits. INTS3, INTS10, INTS12, INTS13, and INTS14 were not observed in our structural model, and their roles require further study. As a noncanonical PP2A holoenzyme, INTAC directly regulates transcription through dephosphorylating Pol II. (**C**) In the INTAC complex, Integrator binds Pol II and PP2A-C dephosphorylates Pol II CTD at Ser^2, Ser^5, and Ser^7, which suppresses both transcription initiation, pausing, and elongation (top). The loss of phosphatase activity of INTAC increases the levels of Pol II phosphorylation and leads to dysregulated transcription of INTAC target genes (bottom). The phosphatase function is independent of the known RNA endonuclease activity of INTS11 in INTAC.

and stored at −80°C for biochemical analyses. In addition, PP1 (lysed by high pressure) and SSU72 were purified in a similar way.

In vitro phosphatase assay

Unphosphorylated Pol II was purified from pig thymus and was subjected to in vitro phosphorylation using purified human transcription factor IIH (TFIIH). The HEK Expi293 cells containing overexpressed TFIIH were pelleted and lysed in lysis buffer containing 50 mM HEPES pH7.4, 300 mM NaCl, 0.2% CHAPS, 5 mM $MgCl_2$, 5 mM ATP, 0.5 mM EDTA, 10% glycerol, 2 mM DTT, 1 mM PMSF, 1 μg/ml Aprotinin, 1 μg/ml Pepstatin, 1 μg/ml Leupeptin. After centrifugation, the supernatant was incubated with IgG resins at 4°C for 2 hours. The resins were extensively washed and resuspended in a kinase buffer containing 50 mM HEPES pH7.4, 150 mM NaCl, 0.1% CHAPS, 12.5 mM $MgCl_2$, 5 mM ATP, 10% glycerol, 3 mM DTT. The unphosphorylated Pol II (300 μg) was added and mixed with the resins for reaction overnight at 4°C. The phosphorylated Pol II was eluted with an elution buffer containing 50 mM HEPES pH7.4, 150 mM NaCl, 0.1% CHAPS, 5 mM $MgCl_2$, 10% glycerol, 2 mM DTT. The TFIIH remained immobilized on the resins by the N-terminal 4×Protein A tag. The eluted Pol II was aliquoted, snap frozen, and stored at −80°C for the following assays.

The phosphorylated Pol II (100 nM) was incubated with INTAC complex (10 to 40 nM), PP2A core enzyme, PP1, or SSU72 (40 nM) in a final volume of 20 μl containing 50 mM

HEPES pH 7.4, 100 mM NaCl, 0.01% CHAPS, 10 mM $MgCl_2$, 1 mM $MnCl_2$, and 2 mM DTT in the absence or presence of PP2A inhibitor Microcystin-LR (MedChemExpress) (100 nM). The reactions were performed at 30°C for 30 min and stopped by adding 5 μl of 5×SDS loading buffer. Samples (2.5 μl) were loaded on 4-12% SDS-polyacrylamide gels (PAGE) and analyzed by Western blotting with indicated antibodies.

The wild-type PP2A-C (WT) or phosphatase dead PP2A-C (H59K/H241K, termed PD) were subcloned into plasmid containing an N-terminal Flag tag followed by 4×Protein A tag. The tagged PP2A-C, untagged PP2A-A, and untagged INTS1-INTS14 were cotransfected into HEK Expi293 cells. The protein expression and purification were performed using IgG resins followed by on-column cleavage, essentially as described above. The eluted proteins, INTAC-WT and INTAC-PD, were individually incubated with antibodies against INTS5 and immobilized on Protein G resins. The resins were then extensively washed and resuspended in 20 μl of the reaction buffer containing 50 mM HEPES pH7.4, 125 mM NaCl, 0.1% CHAPS, 5% glycerol, 2 mM DTT. The phosphorylated Pol II (20 nM) was added to resins and incubated at 30°C for 30 min. The reactions were stopped by adding 5 μl of 5×SDS loading buffer. Samples (5 μl) were loaded on 4-12% SDS-polyacrylamide gels (PAGE) and analyzed by Western blotting with indicated antibodies.

In vitro phosphatase assay with pNPP

To test the phosphatase activity of the purified phosphatases, we performed an assay using the general substrate *p*-nitrophenyl phosphate (pNPP, New England Biolabs). The pNPP (50 mM) was incubated with INTAC complex, PP2A core enzyme, PP1(50 nM), or SSU72 (50 nM and 0.5 μM) in a final volume of 20 μl containing 50 mM HEPES pH 7.4, 100 mM NaCl, 0.01% CHAPS, 10 mM $MgCl_2$, 1 mM $MnCl_2$ (without $MnCl_2$ for SSU72 owing to the decreased activity in the presence of $MnCl_2$), and 2 mM DTT. After incubation at 30°C for 20 min, the reactions were stopped by adding 80 μl of 1 M NaOH, and the released pNPP was measured at optical density at 405 nm (OD 405 nm) using a Synergy H4 Hybrid Microplate Reader (BioTek Instruments). The experiments were performed in triplicates.

RNA interference

Lentivirus expressing shRNAs was prepared by transfecting PLKO.1 shRNA plasmids and packaging plasmids containing psPAX2 and pMD2.G into 293T cells using PEI (Polysciences). Collected conditional media containing virus particles were used to transduce cells in the growth media supplemented with 8 μg/ml Polybrene for 24 hours. The infected cells were selected with 2 μg/ml puromycin for an extra 48 hours. The cells were then switched into growth media without antibiotics and grown for an additional 24 hours before being harvested for further analysis.

ChIP-seq and data analysis

For each ChIP assay, 1×10^7 DLD-1 cells were used, as described (*63*), and 1×10^6 MEF cells were mixed with DLD-1 cells as the spike-in. The libraries for sequencing were prepared with VAHTS Universal Plus DNA Library Prep Kit for Illumina. Reads were aligned to the human genome (UCSC hg19) by using alignment software Bowtie (1.1.1) (*64*) with the following options: -S –m 1 –k 1. Only unique reads mapping to a single genomic location were kept. Duplicate reads were removed by using in house script. We calculated the ChIP-seq signal as reads per million mapped reads (RPM) based on sequence depth. We counted the number of spike-in reads by aligning to the mouse genome(mm10) with bowtie. We normalized the ChIP-seq signal by multiplying the RPM by the spike-in factor a = $1/N_d$, where N_d is the count of reads (in millions) mapped to mouse genome (*65*). Density genome tracks were generated by using the bedtools genomecov program to convert bam files to bigwig files (*66*). Peaks were called using macs (1.3.7.1) (*67*) with the following parameters: –nomodel –nolambda and the peaks with 10*log10 (pvalue) > 100 were kept. To produce the heatmaps, density was calculated at peaks summit and tiled the flanking upstream and downstream regions within 5 kb in 100 bins.

Nascent-seq

The nascent RNA isolation procedure was previously described (*68*). Libraries were made with the VAHTS Stranded mRNA-seq Library Prep Kit for Illumina V2. Reads were aligned to the human genome (UCSC hg19) using Tophat (v2.1.1) (*69*). The parameters used were "-m 1 -g 1 -microexon -search–no-coverage-search." The gene expression level was calculated by Cufflinks (version 2.2.1) (*70*) using the refFlat database from the UCSC genome browser. Genes with fragments per kilobase of transcript per million mapped reads (FPKM) ≥ 1 were considered expressed for subsequent analysis.

IgG pulldown assay

HEK Expi293 cells containing overexpressed INTAC complex were pelleted and lysed as described above. The supernatant of the cell lysate was incubated with IgG resins for 2 hours at 4°C. The complex was immobilized on the resins by N-terminal 4×Protein A–tagged INTS1. The resins were extensively washed and resuspended in 0.5 ml of the binding buffer containing 25 mM HEPES pH7.4, 150 mM NaCl, 0.1% CHAPS, 2 mM MgCl2, 5% glycerol, 2 mM DTT. The phosphorylated or unphosphorylated Pol II was then added and incubated with the resins at 4°C for 2 hours. The resins were extensively washed with the binding buffer, and the bound proteins were subjected to SDS-PAGE followed by Coomassie blue staining.

Immunoprecipitation assay

Nuclear extract was prepared from HEK Expi293 cells using lysis buffer containing 50 mM HEPES pH 7.4, 300 mM NaCl, 0.2% NP-40, 2 mM $MgCl_2$, 10% glycerol, 2 mM DTT. After centrifuging at 16,000 rpm for 30 min, supernatant was collected and the NaCl concentration was adjusted to 150 mM. The nuclear extract was incubated with antibodies for 2 hours, and Protein G resins for another 2 hours at 4°C. After immunoprecipitation, resins were washed five times with wash buffer containing 30 mM HEPES pH 7.4, 150 mM NaCl, 0.1% NP-40, 2 mM $MgCl_2$, 5% glycerol, 2 mM DTT. Samples were subject to SDS-PAGE and Western blotting.

Cryo-EM sample preparation

The purified INTAC complex was cross-linked and purified using Grafix (*71*). The glycerol gradient was prepared using light buffer containing 8% (v/v) glycerol, 20 mM HEPES pH 7.4, 200 mM NaCl, 0.05% CHAPS, 2 mM DTT, and heavy buffer containing 40% (v/v) glycerol, 0.0015% glutaraldehyde (Sigma), 20 mM HEPES pH 7.4, 200 mM NaCl, 0.05% CHAPS, 2 mM DTT. After centrifugation at 34,000 rpm for 16 hours at 4°C, peak fractions were pooled and the cross-linking reactions were quenched with 100 mM Tris-HCl pH 7.0. The homogeneity of peak fractions was assessed by negative-stain electron microscopy (EM). Fractions of interest were concentrated to about 1.7 mg/ml and dialyzed overnight against a buffer containing 20 mM HEPES pH 7.4, 200 mM NaCl, 0.01% CHAPS, 0.8% glycerol, 1 mM tris(2-carboxyethyl)phosphine (TCEP), followed by cryo-EM grid preparation.

For negative-stain EM, 5 μl of freshly purified protein sample was applied onto a glow-discharged copper grid supported by a thin layer of carbon film for 1 min before negative staining by 2% (w/v) uranyl formate at room temperature. The negatively stained grid was loaded onto a FEI Talos L120C microscope operated at 120 kV, equipped with a Ceta CCD camera.

For cryo-EM grid preparation, 4 μl of protein sample (about 0.73 mg/ml) was applied onto a glow-discharged holey carbon grid (Quantifoil Au, R1.2/1.3, 300 mesh). After blotting for 3 to 6 s, the grid was vitrified by plunging it into liquid ethane using a Vitrobot Mark IV (FEI) operated at 4°C and 100% humidity.

Cryo-EM data collection and image processing

Cryo-EM data were collected on a Titan Krios electron microscope (FEI) operated at 300 kV

at the Cryo-EM platform of Peking University, equipped with a K2 summit direct detector (Gatan) and a GIF quantum energy filter (Gatan) set to a slit width of 20 eV. Automated data acquisition was carried out with Serial EM software (*72*) in the superresolution counting mode (*73*) at a nominal magnification 130,000×, corresponding to a calibrated pixel size of 1.055 Å, and a defocus range from −1.5 to −2.5 μm. Each image stack was dose-fractionated to 32 frames with a total exposure dose of about 50 $e^-/Å^2$ and exposure time of 6.96 s.

The image stacks were motion-corrected and dose-weighted using MotionCorr2 (*74*). The contrast transfer function (CTF) parameters were estimated by Gctf (*75*) from non-dose-weighted micrographs. About 40,000 particles autopicked from 200 micrographs were subjected into two-dimensional (2D) classification in RELION v3.0 (*76*) and ab initio reconstruction by cryoSPARC v2 (*77*). The selected 2D templates and 3D initial model were low-passed and used as references for subsequent particle-picking and 3D classification. The following procedures of image processing were performed using RELION for dose-weighted micrographs. 1,439,347 particles were autopicked from 6826 micrographs for further data processing. After two rounds of 30-iteration 2D and 3D classification, 79,408 particles within the two selected classes were grouped and subjected to auto refinement and postprocessing, yielding a density map at 3.8-Å resolution. 68,378 particles were selected after one round of 3D classification without alignment. Finally, a reconstruction at 3.5-Å overall resolution was obtained using particles subjected to CTF refinement and Bayesian polishing.

The reported resolutions above are based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion. In addition, local-mask refinements were applied to improve the local resolution in favor of subsequent model building. All the visualization and evaluation of 3D density maps were performed with UCSF Chimera or UCSF ChimeraX (*78*, *79*), and the local resolution variations were calculated using ResMap (*80*). The directional FSC plots were calculated using 3DFSC (*81*). The above procedures of data processing are summarized in fig. S2.

Model building and structure refinement

We combined de novo modeling, homology modeling, and rigid-body fitting of subunits with known structures to generate an atomic model. INTS11 and INTS9 share structural homology with CPSF73 and CPSF100 in the CPSF complex. The crystal structures of human CPSF73 [Protein Data Bank (PDB) ID: 2I7V] (*47*) and yeast CPSF100 (PDB ID: 2I7X) (*47*) were used to guide modeling of INTS11 and INTS9 according to the local-mask refinement map at 3.8-Å resolution. The structures of human INTS11-INTS9 CTD complex (PDB ID: 5V8W) (*22*) and PP2A holoenzyme (PDB ID: 2NYL) (*46*) were placed into the density maps by rigid-body fitting in Chimera (*78*) and manually adjusted according to the locally refined maps in Coot (*82*). The obtained models fit well into the overall map.

For the remaining subunits (INTS1, INTS2, INTS5, INTS6, INTS7, and INTS8) that lack homologous structures, de novo modeling was performed. The model of INTS6 was built according to the local masked map at 3.5-Å resolution, followed by fitting into the overall map. The 3.6- and 3.7-Å local-mask refinement maps were used for sequence assignments of INTS2, INTS7, PP2A-A, and PP2A-C. Secondary structure predictions from PSIPRED (*83*) were used to assist de novo modeling. The main chain parts of these structures were manually built according to the maps in COOT (*82*). Sequence assignments and identification of subunits were guided mainly by bulky residues such as Tyr, Phe, and Trp using a script developed by our group. Maps without postprocessing were used to build the unassigned chain.

The structure model of the INTAC complex was refined against the 3.5-Å overall map in real space with PHENIX (*84*) and validated through examination of Ramachandran plot statistics, a MolProbity score (*85*), and a EMRinger score (*86*). The statistics of the map reconstruction and model refinement are summarized in table S1. The structural model was used for combining the focused maps to generate a composite map using program phenix.combine_focused_maps as previously described (*87*, *88*). The combined map is shown in Fig. 2B, fig. S3F, and movies S2 to S7. Map and model representations in the figures and movies were prepared by PyMOL (*89*) and UCSF ChimeraX (*79*).

REFERENCES AND NOTES

1. M. Rienzo, A. Casamassimi, Integrator complex and transcription regulation: Recent findings and pathophysiology. *Biochim. Biophys. Acta* **1859**, 1269–1280 (2016). doi: 10.1016/j.bbagrm.2016.07.008; pmid: 27427483
2. A. Gardini *et al.*, Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56**, 128–139 (2014). doi: 10.1016/j.molcel.2014.08.004; pmid: 25201415
3. F. Lai, A. Gardini, A. Zhang, R. Shiekhattar, Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399–403 (2015). doi: 10.1038/nature14906; pmid: 26308897
4. J. Yue *et al.*, Integrator orchestrates RAS/ERK1/2 signaling transcriptional programs. *Genes Dev.* **31**, 1809–1820 (2017). doi: 10.1101/gad.301697.117; pmid: 28982763
5. N. D. Elrod *et al.*, The Integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* **76**, 738–752.e7 (2019). doi: 10.1016/j.molcel.2019.10.034; pmid: 31809743
6. D. C. Tatomer *et al.*, The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* **33**, 1525–1538 (2019). doi: 10.1101/gad.330167.119; pmid: 31530651
7. J. R. Skaar *et al.*, The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res.* **25**, 288–305 (2015). doi: 10.1038/cr.2015.19; pmid: 25675981
8. B. Stadelmayer *et al.*, Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531 (2014). doi: 10.1038/ncomms6531; pmid: 25410209
9. D. Baillat *et al.*, Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005). doi: 10.1016/j.cell.2005.08.019; pmid: 16239144
10. S. Egloff *et al.*, Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**, 1777–1779 (2007). doi: 10.1126/science.1145989; pmid: 18079403
11. J. Yamamoto *et al.*, DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* **5**, 4263 (2014). doi: 10.1038/ncomms5263; pmid: 24968874
12. M. P. Rubtsova *et al.*, Integrator is a key component of human telomerase RNA biogenesis. *Sci. Rep.* **9**, 1701 (2019). doi: 10.1038/s41598-018-38297-6; pmid: 30737432
13. M. Xie *et al.*, The host Integrator complex acts in transcription-independent maturation of herpesvirus microRNA 3′ ends. *Genes Dev.* **29**, 1552–1564 (2015). doi: 10.1101/gad.266973.115; pmid: 26220997
14. D. Cazalla, M. Xie, J. A. Steitz, A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol. Cell* **43**, 982–992 (2011). doi: 10.1016/j.molcel.2011.07.025; pmid: 21925386
15. R. Oegema *et al.*, Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLOS Genet.* **13**, e1006809 (2017). doi: 10.1371/journal.pgen.1006809; pmid: 28542170
16. K. Y. Lui *et al.*, Integrator complex subunit 6 (INTS6) inhibits hepatocellular carcinoma growth by Wnt pathway and serve as a prognostic marker. *BMC Cancer* **17**, 644 (2017). doi: 10.1186/s12885-017-3628-3; pmid: 28899352
17. A. Yoshimi *et al.*, Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature* **574**, 273–277 (2019). doi: 10.1038/s41586-019-1618-0; pmid: 31578525
18. J. Chen *et al.*, An RNAi screen identifies additional members of the *Drosophila* Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3′-end formation. *RNA* **18**, 2148–2156 (2012). doi: 10.1261/rna.035725.112; pmid: 23097424
19. T. R. Albrecht *et al.*, Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res.* **46**, 4241–4255 (2018). doi: 10.1093/nar/gky100; pmid: 29471365
20. J. Huang, Z. Gong, G. Ghosal, J. Chen, SOSS complexes participate in the maintenance of genomic stability. *Mol. Cell* **35**, 384–393 (2009). doi: 10.1016/j.molcel.2009.06.011; pmid: 19683501
21. W. Ren *et al.*, Structural basis of SOSS1 complex assembly and recognition of ssDNA. *Cell Rep.* **6**, 982–991 (2014). doi: 10.1016/j.celrep.2014.02.020; pmid: 24630995
22. Y. Wu, T. R. Albrecht, D. Baillat, E. J. Wagner, L. Tong, Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4394–4399 (2017). doi: 10.1073/pnas.1616605114; pmid: 28396433
23. S. Buratowski, Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009). doi: 10.1016/j.molcel.2009.10.019; pmid: 19941815
24. J. P. Hsin, J. L. Manley, The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012). doi: 10.1101/gad.200303.112; pmid: 23028141
25. I. Jonkers, J. T. Lis, Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015). doi: 10.1038/nrm3953; pmid: 25693130
26. F. X. Chen, E. R. Smith, A. Shilatifard, Born to run: Control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478 (2018). doi: 10.1038/s41580-018-0010-5; pmid: 29740129
27. S. Egloff, S. Murphy, Cracking the RNA polymerase II CTD code. *Trends Genet.* **24**, 280–288 (2008). doi: 10.1016/j.tig.2008.03.008; pmid: 18457900
28. J. Archambault *et al.*, FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase IIO. *J. Biol. Chem.* **273**, 27593–27601 (1998). doi: 10.1074/jbc.273.42.27593; pmid: 9765293
29. S. Krishnamurthy, X. He, M. Reyes-Reyes, C. Moore, M. Hampsey, Ssu72 is an RNA polymerase II CTD phosphatase. *Mol. Cell* **14**, 387–394 (2004). doi: 10.1016/S1097-2765(04)00235-7; pmid: 15125841
30. S. Egloff, J. Zaborowska, C. Laitem, T. Kiss, S. Murphy, Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol. Cell* **45**, 111–122 (2012). doi: 10.1016/j.molcel.2011.11.006; pmid: 22137580

31. A. L. Mosley *et al*., Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol. Cell* **34**, 168–178 (2009). doi: 10.1016/j.molcel.2009.02.025; pmid: 19394294
32. V. Janssens, J. Goris, Protein phosphatase 2A: A highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochem. J.* **353**, 417–439 (2001). doi: 10.1042/bj3530417; pmid: 11171037
33. N. Wlodarchak, Y. Xing, PP2A as a master regulator of the cell cycle. *Crit. Rev. Biochem. Mol. Biol.* **51**, 162–184 (2016). doi: 10.3109/10409238.2016.1143913; pmid: 26906453
34. Y. Shi, Serine/threonine phosphatases: Mechanism through structure. *Cell* **139**, 468–484 (2009). doi: 10.1016/j.cell.2009.10.006; pmid: 19879837
35. P. Seshacharyulu, P. Pandey, K. Datta, S. K. Batra, Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. *Cancer Lett.* **335**, 9–18 (2013). doi: 10.1016/j.canlet.2013.02.036; pmid: 23454242
36. E. Kremmer, K. Ohst, J. Kiefer, N. Brewis, G. Walter, Separation of PP2A core enzyme and holoenzyme with monoclonal antibodies against the regulatory A subunit: Abundant expression of both forms in cells. *Mol. Cell. Biol.* **17**, 1692–1701 (1997). doi: 10.1128/MCB.17.3.1692; pmid: 9032296
37. J. Sangodkar *et al*., All roads lead to PP2A: Exploiting the therapeutic potential of this phosphatase. *FEBS J.* **283**, 1004–1024 (2016). doi: 10.1111/febs.13573; pmid: 26507691
38. C. M. O'Connor, A. Perl, D. Leonard, J. Sangodkar, G. Narla, Therapeutic targeting of PP2A. *Int. J. Biochem. Cell Biol.* **96**, 182–193 (2018). doi: 10.1016/j.biocel.2017.10.008; pmid: 29107183
39. J. Westermarck, W. C. Hahn, Multiple pathways regulated by the tumor suppressor PP2A in transformation. *Trends Mol. Med.* **14**, 152–160 (2008). doi: 10.1016/j.molmed.2008.02.001; pmid: 18329957
40. A. M. Silverstein, C. A. Barrow, A. J. Davis, M. C. Mumby, Actions of PP2A on the MAP kinase pathway and apoptosis are mediated by distinct regulatory subunits. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4221–4226 (2002). doi: 10.1073/pnas.072071699; pmid: 11904383
41. E. Kickstein *et al*., Biguanide metformin acts on tau phosphorylation via mTOR/protein phosphatase 2A (PP2A) signaling. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21830–21835 (2010). doi: 10.1073/pnas.0912793107; pmid: 21098287
42. Y. C. Kuo *et al*., Regulation of phosphorylation of Thr-308 of Akt, cell proliferation, and survival by the B55α regulatory subunit targeting of the protein phosphatase 2A holoenzyme to Akt. *J. Biol. Chem.* **283**, 1882–1892 (2008). doi: 10.1074/jbc.M709585200; pmid: 18042541
43. D. Baillat, E. J. Wagner, Integrator: Surprisingly diverse functions in gene expression. *Trends Biochem. Sci.* **40**, 257–264 (2015). doi: 10.1016/j.tibs.2015.03.005; pmid: 25882383
44. R. Sales Gil, I. J. de Castro, J. Berihun, P. Vagnarelli, Protein phosphatases at the nuclear envelope. *Biochem. Soc. Trans.* **46**, 173–182 (2018). doi: 10.1042/BST20170139; pmid: 29432143
45. S. T. Mo *et al*., Visualization of subunit interactions and ternary complexes of protein phosphatase 2A in mammalian cells. *PLOS ONE* **9**, e116074 (2014). doi: 10.1371/journal.pone.0116074; pmid: 25536081
46. Y. Xu *et al*., Structure of the protein phosphatase 2A holoenzyme. *Cell* **127**, 1239–1251 (2006). doi: 10.1016/j.cell.2006.11.033; pmid: 17174897
47. C. R. Mandel *et al*., Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* **444**, 953–956 (2006). doi: 10.1038/nature05363; pmid: 17128255
48. Y. Xing *et al*., Structure of protein phosphatase 2A core enzyme bound to tumor-inducing toxins. *Cell* **127**, 341–353 (2006). doi: 10.1016/j.cell.2006.09.025; pmid: 17055435
49. M. R. Groves, N. Hanlon, P. Turowski, B. A. Hemmings, D. Barford, The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell* **96**, 99–110 (1999). doi: 10.1016/S0092-8674(00)80963-0; pmid: 9989501
50. J. R. Walker, R. A. Corpina, J. Goldberg, Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607–614 (2001). doi: 10.1038/35088000; pmid: 11493912
51. S. Mikami *et al*., Structural insights into the recruitment of SMRT by the corepressor SHARP under phosphorylative regulation. *Structure* **22**, 35–46 (2014). doi: 10.1016/j.str.2013.10.007; pmid: 24268649
52. Y. Xu, Y. Chen, P. Zhang, P. D. Jeffrey, Y. Shi, Structure of a protein phosphatase 2A holoenzyme: Insights into B55-mediated Tau dephosphorylation. *Mol. Cell* **31**, 873–885 (2008). doi: 10.1016/j.molcel.2008.08.006; pmid: 18922469
53. U. S. Cho, W. Xu, Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature* **445**, 53–57 (2007). doi: 10.1038/nature05351; pmid: 17086192
54. N. Wlodarchak *et al*., Structure of the Ca^{2+}-dependent PP2A heterotrimer and insights into Cdc6 dephosphorylation. *Cell Res.* **23**, 931–946 (2013). doi: 10.1038/cr.2013.77; pmid: 23752926
55. Z. Xu *et al*., Structure and function of the PP2A-shugoshin interaction. *Mol. Cell* **35**, 426–441 (2009). doi: 10.1016/j.molcel.2009.06.031; pmid: 19716788
56. F. Guo *et al*., Structural basis of PP2A activation by PTPA, an ATP-dependent activation chaperone. *Cell Res.* **24**, 190–203 (2014). doi: 10.1038/cr.2013.138; pmid: 24100351
57. F. X. Chen *et al*., PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. *Cell* **162**, 1003–1015 (2015). doi: 10.1016/j.cell.2015.07.042; pmid: 26279188
58. Y. Sun *et al*., Structure of an active human histone pre-mRNA 3′-end processing machinery. *Science* **367**, 700–703 (2020). doi: 10.1126/science.aaz7758; pmid: 32029631
59. Y. Zhang, Y. Sun, Y. Shi, T. Walz, L. Tong, Structural insights into the human pre-mRNA 3′-end processing machinery. *Mol. Cell* **77**, 800–809.e6 (2020). doi: 10.1016/j.molcel.2019.11.005; pmid: 31810758
60. K. Xiang *et al*., Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**, 729–733 (2010). doi: 10.1038/nature09391; pmid: 20861839
61. K. Xiang, J. L. Manley, L. Tong, An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72. *Genes Dev.* **26**, 2265–2270 (2012). doi: 10.1101/gad.198853.112; pmid: 23070812
62. F. Salvi *et al*., Effects of stably incorporated iron on protein phosphatase-1 structure and activity. *FEBS Lett.* **592**, 4028–4038 (2018). doi: 10.1002/1873-3468.13284; pmid: 30403291
63. F. Chen, X. Gao, A. Shilatifard, Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. *Genes Dev.* **29**, 39–47 (2015). doi: 10.1101/gad.246173.114; pmid: 25561494
64. 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
65. D. A. Orlando *et al*., Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014). doi: 10.1016/j.celrep.2014.10.018; pmid: 25437568
66. A. R. Quinlan, BEDTools: The Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–11.12.34 (2014). doi: 10.1002/0471250953.bi1112s47; pmid: 25199790
67. 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
68. F. X. Chen, S. A. Marshall, Y. Deng, S. Tianjiao, Measuring nascent transcripts by Nascent-seq. *Methods Mol. Biol.* **1712**, 19–26 (2018). doi: 10.1007/978-1-4939-7514-3_2; pmid: 29224065
69. D. Kim *et al*., TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013). doi: 10.1186/gb-2013-14-4-r36; pmid: 23618408
70. C. Trapnell *et al*., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012). doi: 10.1038/nprot.2012.016; pmid: 22383036
71. B. Kastner *et al*., GraFix: Sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008). doi: 10.1038/nmeth1139; pmid: 18157137
72. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). doi: 10.1016/j.jsb.2005.07.007; pmid: 16182563
73. M. G. Campbell *et al*., Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20**, 1823–1828 (2012). doi: 10.1016/j.str.2012.08.026; pmid: 23022349
74. S. Q. Zheng *et al*., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
75. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). doi: 10.1016/j.jsb.2015.11.003; pmid: 26592709
76. J. Zivanov *et al*., New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018). doi: 10.7554/eLife.42166; pmid: 30412051
77. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
78. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
79. T. D. Goddard *et al*., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi: 10.1002/pro.3235; pmid: 28710774
80. A. Kucukelbir, F. J. Sigworth, H. D. Tagare, Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014). doi: 10.1038/nmeth.2727; pmid: 24213166
81. Y. Z. Tan *et al*., Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017). doi: 10.1038/nmeth.4347; pmid: 28671674
82. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
83. D. W. Buchan, F. Minneci, T. C. Nugent, K. Bryson, D. T. Jones, Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, W349–W357 (2013). doi: 10.1093/nar/gkt381; pmid: 23748958
84. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
85. V. B. Chen *et al*., MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010). doi: 10.1107/S0907444909042073; pmid: 20057044
86. B. A. Barad *et al*., EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015). doi: 10.1038/nmeth.3541; pmid: 26280328
87. D. Liebschner *et al*., Macromolecular structure determination using x-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019). doi: 10.1107/S2059798319011471; pmid: 31588918
88. T. H. D. Nguyen *et al*., Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016). doi: 10.1038/nature16940; pmid: 26829225
89. W. L. DeLano, The PyMOL molecular graphics system (Schrödinger, 2002); www.pymol.org.

ACKNOWLEDGMENTS

We thank the Center of Cryo-Electron Microscopy, Fudan University; the Center of Cryo-Electron Microscopy, Peking University; the Center for Biological Imaging of Institute of Biophysics (IBP) of the Chinese Academy of Sciences (CAS); and the National Center for Protein Science Shanghai (NCPSS) for support with cryo-EM data collection and data analyses. We thank B. Wen, Y. Dang, F. Lan, and Q. F. Wang for help with cell culture and sequencing experiments and X. Li at Westlake University for support with mass spectrometry analyses. **Funding:** This work was supported by grants from the National Key R&D program of China (2016YFA0500700), the National Natural Science Foundation of China (31830107, 32030055, 31821002, and 32070636), the National Ten-Thousand Talent Program (Y.X.), the National Program for support of Top-Notch Young Professionals (Y.X.), the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), the Shanghai Municipal Science and Technology Commission (19JC1411500), the Shanghai Natural Science Foundation (20ZR1412100), and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB08000000). **Author contributions:** H.Z., S.H., and C.X. conducted the experiments with help from X.C. and Y.L.; Y.Q. performed EM analyses and model building with help from J.W., J.L., and W.L.; X.C. analyzed the sequencing data with help from Z.Y.; K.L. and G.W. helped with project design; Y.X. and F.X.C. wrote the manuscript with help from H.Z.; and Y.X. and F.X.C. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The cryo-EM density map has been deposited in the Electron Microscopy Data Bank (EMDB) under accession code EMD- 30473, and the model coordinate has been deposited in the Protein Data Bank (PDB) under accession no. 7CUN. Sequencing data have been deposited at the Gene Expression Omnibus (GEO) under accession number GSE154093.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/eabb5872/suppl/DC1
Figs. S1 to S12
Table S1
References
MDAR Reproducibility Checklist
Movies S1 to S7

View/request a protocol for this paper from *Bio-protocol*.

5 March 2020; resubmitted 9 July 2020
Accepted 23 September 2020
10.1126/science.abb5872

RESEARCH ARTICLE SUMMARY

CELL BIOLOGY

Abl and Canoe/Afadin mediate mechanotransduction at tricellular junctions

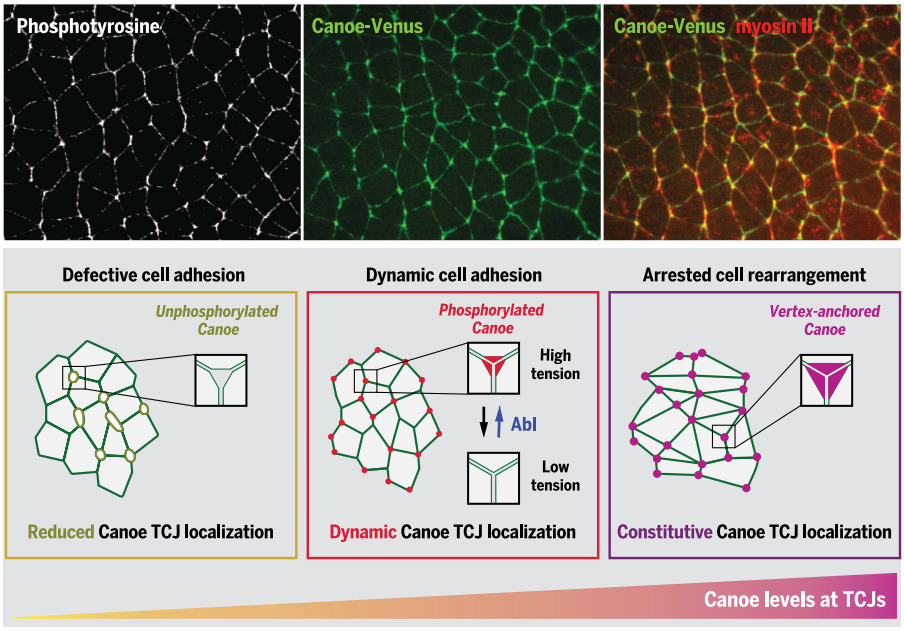
Huapeng H. Yu and Jennifer A. Zallen*

INTRODUCTION: The ability to sense and respond to mechanical force is an intrinsic property of cells that is critical for normal cell function and often deregulated in disease. Mechanical forces induce changes in protein conformation, localization, activity, and post-translational modification in vitro. An outstanding question is how mechanical signals modulate protein function and cell behavior in response to the physiological forces encountered by cells in living organisms. Tyrosine phosphorylation has long been recognized to be enriched at adherens junctions, which detect and transmit forces between epithelial cells, and can be directly enhanced by mechanical forces in vitro. However, whether mechanical forces influence tyrosine kinase signaling in vivo, and the consequences of this regulation for cell behavior, are not known.

RATIONALE: An important step in the development of epithelial organs is the remodeling of cell adhesion in response to mechanical forces, which transforms epithelial sheets into diverse structures. Tricellular junctions (TCJs) where three cells meet are dynamically assembled and disassembled during cell rearrangement and are predicted to be sites of increased tension during epithelial remodeling, providing an opportunity to investigate how cells respond to mechanical forces. However, the mechanisms that allow tricellular junctions to rapidly remodel under tension without disrupting epithelial continuity are not well understood. Identifying and characterizing proteins that localize to tricellular junctions under tension can provide insight into the mechanotransduction pathways that modulate cell adhesion in response to mechanical forces in vivo.

RESULTS: Using time-lapse imaging and a method to directly visualize tyrosine phosphorylation in living embryos, we found that tyrosine phosphorylation is increased at tricellular junctions in a tension-dependent fashion during *Drosophila* convergent extension. By screening a library of tagged proteins, we identified Canoe as a tyrosine-phosphorylated protein that localizes to tricellular junctions under tension. Canoe recruitment to tricellular junctions dynamically correlated with changes in myosin localization, and Canoe rapidly dissociated from tricellular junctions when tension was released by laser ablation. The enrichment of Canoe at tricellular junctions required Abl-dependent tyrosine phosphorylation of Canoe at Y1987, a conserved site in the Canoe actin-binding domain. Decreasing Abl expression or mutating this tyrosine to an unphosphorylatable amino acid significantly reduced Canoe enrichment at tricellular junctions and disrupted tricellular adhesion in vivo. To determine whether the dynamic recruitment of Canoe to tricellular junctions is important for cell rearrangement, we developed a "vertex trap" method to constitutively target Venus-tagged Canoe to tricellular junctions independently of mechanical inputs. Stably anchoring Canoe at tricellular junctions aberrantly stabilized adhesion at four-way vertices and arrested cell rearrangement. These results suggest that Canoe levels dynamically modulate the strength of cell adhesion at tricellular junctions to facilitate cell rearrangement.

CONCLUSION: These results identify a mechanotransduction pathway that dynamically couples tricellular adhesion with mechanical forces during epithelial remodeling. Canoe recruitment to tricellular junctions requires actomyosin contractility and Abl-dependent phosphorylation of a conserved tyrosine in the Canoe actin-binding domain. Loss of Canoe disrupts tricellular adhesion, and increasing Canoe levels at tricellular junctions slows or arrests cell rearrangement, influencing the rate of epithelial remodeling. In one model, Canoe could act as a mechanosensor, physically stretching under tension to expose a phosphorylation site for Abl. Alternatively, tension could enhance Abl activity or generate a unique configuration of actin or other molecules at tricellular junctions that is recognized by Canoe. Tricellular junctions influence many processes required for epithelial development and homeostasis, including cell rearrangement, cell division, stem cell self-renewal, and barrier function. A better understanding of the composition, organization, and dynamic properties of tricellular junctions will provide insight into how these structures sense and integrate mechanical forces in epithelia. ■



Regulation of tricellular adhesion by tension requires Canoe and the Abl tyrosine kinase. (Top) Localization of phosphotyrosine (white), Canoe-Venus (green), and myosin-mCherry (red) in the *Drosophila* embryo. (Bottom) Canoe localization at tricellular junctions is enhanced by Abl and cytoskeletal tension. This force-sensitive mechanism is important for cell adhesion and epithelial remodeling.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zallenj@mskcc.org
Cite this article as: H. H. Yu and J. A. Zallen, *Science* **370**, eaba5528 (2020). DOI: 10.1126/science.aba5528

RESEARCH ARTICLE

CELL BIOLOGY

Abl and Canoe/Afadin mediate mechanotransduction at tricellular junctions

Huapeng H. Yu and Jennifer A. Zallen*

Epithelial structure is generated by the dynamic reorganization of cells in response to mechanical forces. Adherens junctions transmit forces between cells, but how cells sense and respond to these forces in vivo is not well understood. We identify a mechanotransduction pathway involving the Abl tyrosine kinase and Canoe/Afadin that stabilizes cell adhesion under tension at tricellular junctions in the *Drosophila* embryo. Canoe is recruited to tricellular junctions in response to actomyosin contractility, and this mechanosensitivity requires Abl-dependent phosphorylation of a conserved tyrosine in the Canoe actin-binding domain. Preventing Canoe tyrosine phosphorylation destabilizes tricellular adhesion, and anchoring Canoe at tricellular junctions independently of mechanical inputs aberrantly stabilizes adhesion, arresting cell rearrangement. These results identify a force-responsive mechanism that stabilizes tricellular adhesion under tension during epithelial remodeling.

A universal property of cells is the ability to detect and respond to mechanical stimuli such as the stiffness of the extracellular environment, shear and osmotic stress from fluid flow, and forces generated by neighboring cells (*1*, *2*). Mechanical forces influence cell fate, division, adhesion, and behavior (*3*, *4*), and defects in cellular mechanotransduction pathways are implicated in many diseases, including deafness, atherosclerosis, and cancer (*5*, *6*). In epithelial tissues, adherens junctions are critical sites of force transmission between cells (*7*–*9*) and undergo dynamic assembly and disassembly under tension during cell rearrangement (*3*, *4*, *7*). Interactions between several components of adherens junction complexes are stabilized by mechanical forces in vitro and in cultured cells and have been proposed to reinforce cell adhesion under tension (*10*–*15*). However, how adherens junctions sense and respond to mechanical forces in vivo, and how these processes influence dynamic cell behaviors in tissues, are not well understood. Mechanical forces are converted into biochemical changes in cells through a variety of mechanisms, including changes in protein conformation, localization, activity, interactions, and posttranslational modifications (*1*, *2*, *9*). In particular, tyrosine phosphorylation is enhanced at cell-cell and cell-matrix adhesions, two essential force-bearing structures within cells (*16*, *17*). Tyrosine phosphorylation can be modulated by mechanical force in vitro, with evidence for force sensing at the level of tyrosine kinases (*18*–*22*) and substrates (*22*–*25*). However, whether tyrosine kinase signaling is regulated by mechanical forces in vivo, and the effects of this regulation on cell adhesion and epithelial remodeling, are not well understood.

Howard Hughes Medical Institute and Developmental Biology Program, Sloan Kettering Institute, New York, NY, USA.
*Corresponding author. Email: zallenj@mskcc.org

Tyrosine phosphorylation at tricellular junctions is tension sensitive

To investigate whether tyrosine kinase signaling is regulated by mechanical forces in vivo, we analyzed the distribution of tyrosine-phosphorylated proteins in the *Drosophila* embryo. During convergent extension, epithelial cells undergo directional cell rearrangements in response to planar polarized actomyosin contractility (*26*–*30*). We observed a tissue-wide pattern of phosphotyrosine during this process, with the highest levels of tyrosine phosphorylation at tricellular adherens junctions where three cells meet (Fig. 1A and fig. S1A). Tricellular adherens junctions are specialized structures that contain dedicated adhesion molecules (*31*–*34*) and are predicted to be sites of increased tension in epithelia (*35*–*37*). Phosphotyrosine was more strongly enriched at tricellular junctions than the core adherens junction proteins E-cadherin and α-catenin or the membrane marker Spider-GFP (Fig. 1, B and G; fig. S1, H and I; and table S1). The enrichment of phosphotyrosine was further increased at four-way junctions that form as intermediates during cell rearrangement (fig. S1, B and C). These results indicate that tyrosine phosphorylation is enhanced at sites where three or more cells meet.

To investigate the dynamics of tyrosine kinase signaling at tricellular junctions, we developed a method to visualize tyrosine phosphorylation in living embryos. Fluorophore-conjugated monoclonal antibodies to phosphotyrosine were injected into syncytial embryos and retained in cells after cellularization, allowing direct visualization of tyrosine-phosphorylated proteins (Fig. 1, C to E, and movie S1). Phosphotyrosine levels at tricellular junctions were positively correlated with myosin II regulatory light chain fused to the green fluorescent protein (MRLC-GFP), an indicator of cortical contractility (Fig. 1, D and I, and fig. S1E). Inhibiting myosin activity by njecting the Rho-kinase inhibitor Y-27632 (*27*, *30*) rapidly depleted phosphotyrosine levels at tricellular junctions within 5 min of injection, accompanied by an increase in phosphotyrosine levels at bicellular junctions (Fig. 1, F and H, and fig. S1, F and G). The acute release of tension by Y-27632 did not affect the overall level of tyrosine phosphorylation in the embryo (fig. S1D), adherens junction integrity (fig. S1, H to K), or the integrity of tricellular junctions marked by the tricellular adherens junction protein Sidekick (*31*–*34*) (fig. S1, L to O), which is sensitive to actomyosin perturbations over longer time scales (*32*). These results demonstrate that increased tyrosine phosphorylation at tricellular junctions requires cytoskeletal tension.

Canoe/Afadin is a tyrosine kinase substrate at tricellular junctions

To investigate the molecular basis of tension-sensitive tyrosine phosphorylation at tricellular junctions, we screened >100 Venus-tagged proteins (*31*) to identify factors that colocalize with phosphotyrosine at tricellular junctions. This screen identified Canoe/Afadin, a conserved actin-binding protein that links cell surface receptors to the actomyosin cytoskeleton in *Drosophila* and mammals (*38*–*43*). A functional Canoe-Venus fusion expressed from the endogenous locus (*31*) was enriched at tricellular junctions in wild-type (WT) embryos (Fig. 1G), consistent with previous findings (*40*, *42*). Canoe-Venus localization at tricellular junctions was highly dynamic, accumulating and dissipating on a time scale of seconds (movie S2), and these changes were strongly correlated with changes in phosphotyrosine (Fig. 1, J to L, and fig. S2, A to C). Canoe-Venus was tyrosine phosphorylated in embryonic lysates, although Canoe protein accounted for only a minority of tyrosine-phosphorylated proteins in the embryo (fig. S2, D, I, and J). Moreover, phosphotyrosine levels at tricellular junctions were significantly reduced in *canoe*^{R2} maternal mutants (referred to as *canoe* mutants) and these levels were restored by reintroducing wild-type N-terminally tagged Canoe protein (fig. S2, E to H). These results indicate that Canoe, or a Canoe-dependent protein, is a substrate for tyrosine kinase signaling at tricellular junctions.

Canoe/Afadin is a critical regulator of epithelial morphogenesis and animal viability in *Drosophila* and mammals (*38*–*41*). However, although mammalian Afadin can influence tricellular adhesion under certain conditions in culture (*42*), an in vivo requirement for Canoe/Afadin in regulating tricellular adhesion

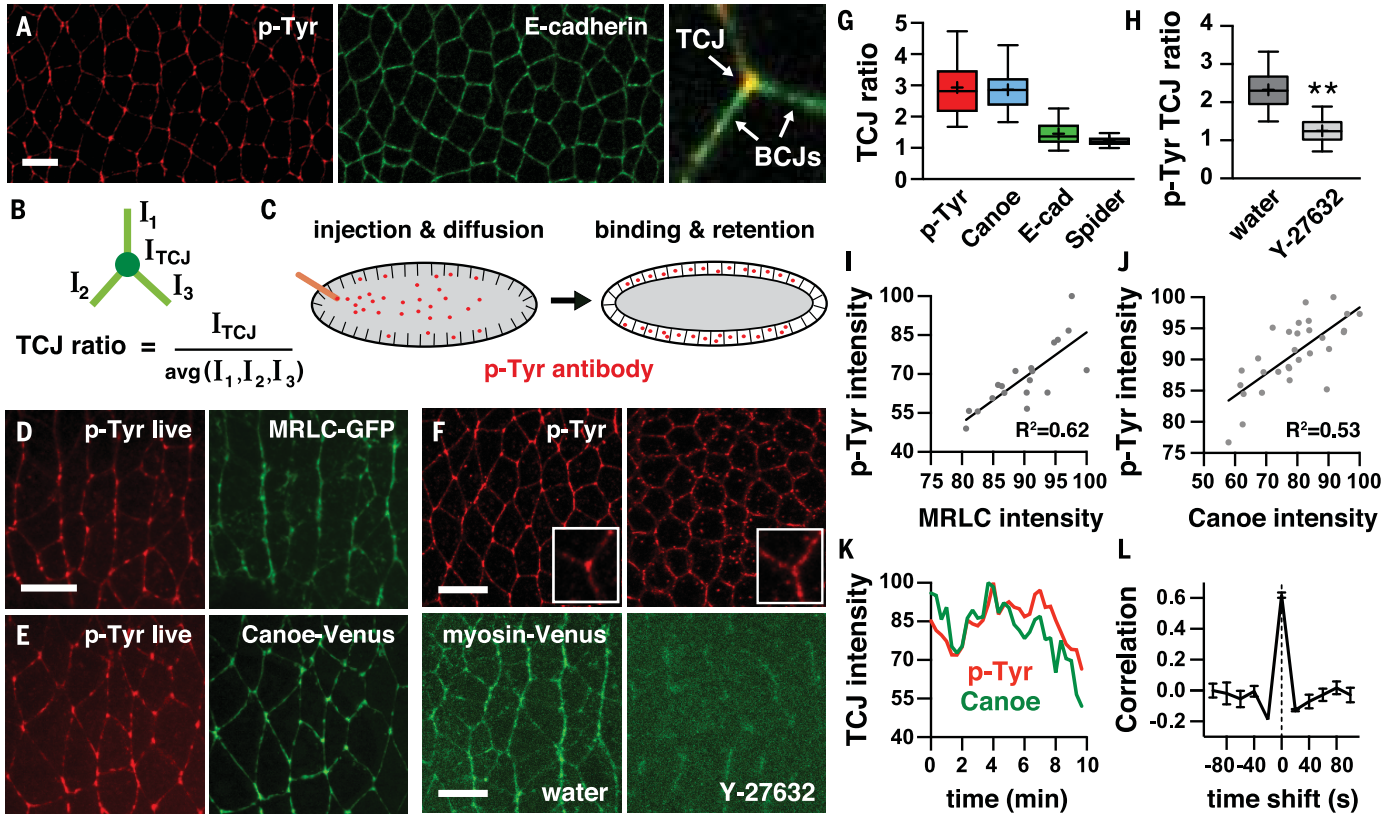


Fig. 1. Tension-sensitive localization of phosphotyrosine at tricellular junctions. (**A**) Phosphotyrosine (p-Tyr) and E-cadherin in the *Drosophila* embryo. Shown is a close-up of one tricellular junction (TCJ) and three bicellular junctions (BCJs). (**B**) TCJ ratio (ratio of the mean TCJ intensity to the mean intensity of the three connected BCJs). (**C**) Alexa Fluor 594–conjugated p-Tyr antibody was injected into syncytial embryos and retained in cells after cellularization. (**D** and **E**) p-Tyr in living embryos expressing MRLC-GFP (D) or endogenous Canoe-Venus (E). (**F**) p-Tyr and myosin-Venus (myosin II heavy chain) in embryos injected with water or Y-27632. Shown are close-ups of single TCJs. (**G**) TCJ ratios of p-Tyr, Canoe-Venus, E-cadherin, and Spider-GFP. (**H**) p-Tyr TCJ ratio in embryos injected with water or Y-27632. (**I** and **J**) p-Tyr intensity correlates with MRLC-GFP (I) and Canoe-Venus (J) at TCJs. Dots are individual TCJs in a single embryo. (**K**) Canoe-Venus and p-Tyr intensity at a single TCJ over time (imaged every 20 s). (**L**) Correlation coefficient for the rates of change in Canoe-Venus and p-Tyr intensity at TCJs (mean ± SEM between embryos). Canoe-Venus data were shifted by the times on the *x*-axis. Boxes are second and third quartiles; whiskers are 5th to 95th percentiles; horizontal line is the median; plus sign is the mean. Intensity is a percentage of the maximum value. $^{**}P < 0.003$, Welch's *t* test. Embryos are stage 7. Living embryos are shown in (D), (E), and (F), bottom panels, and in (I) to (L). Anterior is left; ventral is down. Scale bars, 10 μm. See table S1 for details.

has not been directly demonstrated. To address this, we analyzed cell adhesion in Canoe-deficient embryos using E-cadherin and Sidekick-Venus to visualize bicellular and tricellular junctions, respectively. In WT embryos, cells were in close contact at tricellular junctions and Sidekick-Venus localization in all three cells appeared as a single spot by confocal microscopy, indicating strong adhesion (Fig. 2A). By contrast, gaps in E-cadherin and Sidekick-Venus localization at tricellular junctions were detected in *canoe* mutants and *canoe* knockdown (KD) embryos that express a short hairpin RNA (shRNA) targeting *canoe* (Fig. 2, A, I, and K, and fig. S3, A to C). Defects in E-cadherin localization at tricellular junctions were first detected in stage 7 in Canoe-deficient embryos, when E-cadherin localization at bicellular junctions was largely intact. By contrast, defects in E-cadherin localization at bicellular junctions were not prevalent until stage 8 (fig. S3, D to F). These results demonstrate an early, specific requirement for Canoe in regulating cell adhesion at tricellular junctions.

Abl-dependent phosphorylation of Canoe is required for tricellular adhesion

The correlation between Canoe and phosphotyrosine localization at tricellular junctions, and the requirement for Canoe in maintaining tricellular adhesion, raise the possibility that Canoe is part of a tyrosine kinase signaling pathway that stabilizes cell adhesion at tricellular junctions. Consistent with this idea, injection of the tyrosine kinase inhibitor bosutinib rapidly decreased Canoe localization at tricellular junctions (fig. S4, A to G) and disrupted tricellular adhesion (fig. S4, H to J). Abl, a nonreceptor tyrosine kinase that is inhibited by bosutinib in mammals (*44*), has conserved functions in epithelial remodeling and phosphorylates a large number of substrates involved in actin organization, cell adhesion dynamics, and cell matrix attachment (*45*–*49*). We therefore tested whether Abl is required for Canoe localization and function at tricellular junctions using a maternally expressed shRNA that effectively inhibits *Abl* expression and activity (*48*, *49*). We found that *Abl* KD embryos displayed a significantly reduced enrichment of Canoe and phosphotyrosine at tricellular junctions (Fig. 2, B and C, and fig. S5, A to E), as well as defects in tricellular adhesion that were similar to, but not quite as severe as, the defects in Canoe-deficient embryos (Fig. 2, I to K). These results demonstrate that Abl is required for Canoe localization and cell adhesion at tricellular junctions.

Because Abl is a tyrosine kinase, we hypothesized that it could regulate Canoe localization and function by promoting Canoe tyrosine phosphorylation. In support of this model, Abl enhanced Canoe tyrosine phosphorylation when coexpressed with Canoe in *Drosophila* S2R$^+$ cells (fig. S6, A and B). Canoe

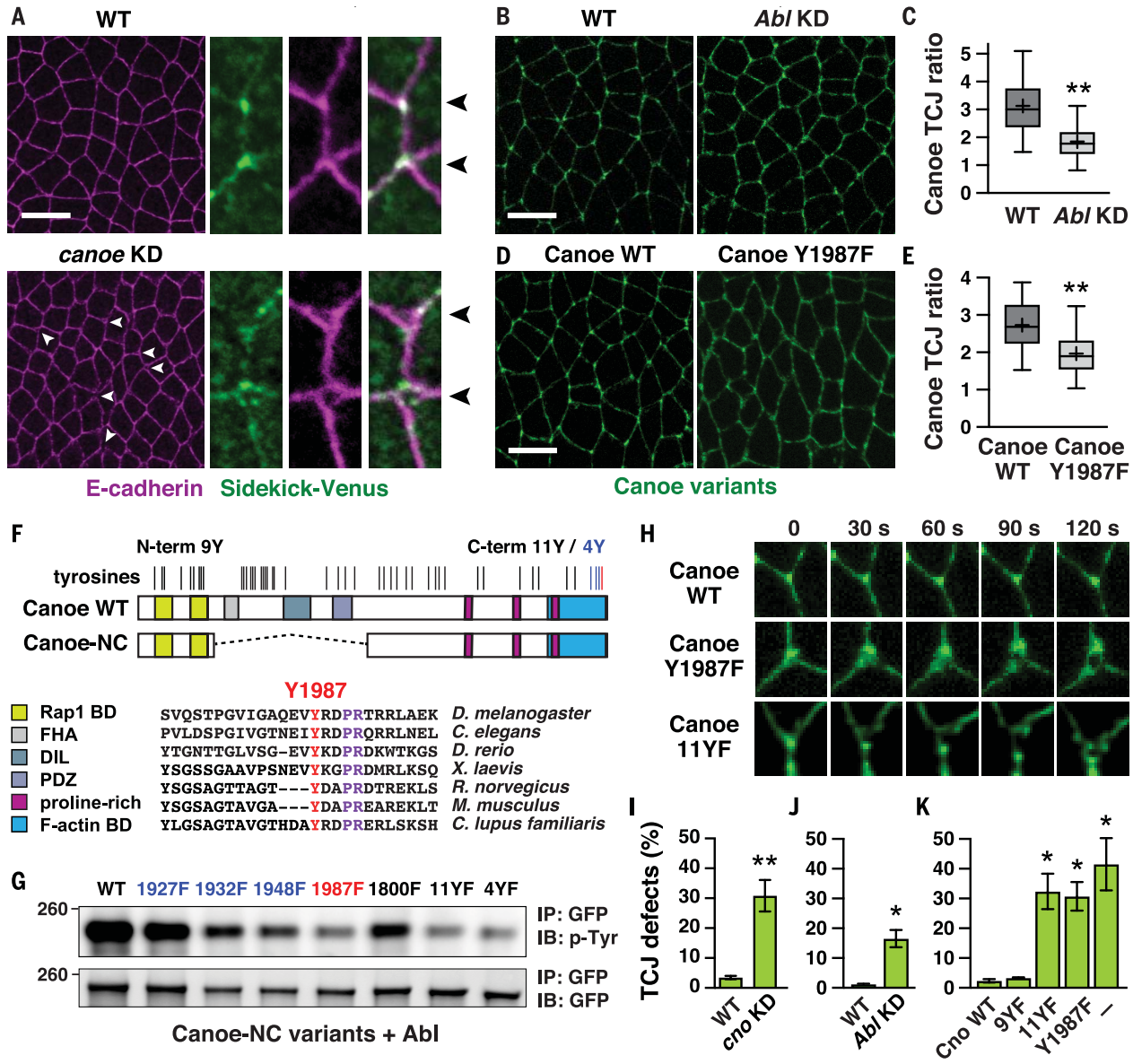


Fig. 2. Tyrosine phosphorylation is required for Canoe localization and function at tricellular junctions. (**A**) E-cadherin and Sidekick-Venus in WT and *canoe* KD embryos. Shown are close-ups of single edges (rotated 30°). Arrowheads indicate TCJs and four-way junctions. (**B**) Endogenous Canoe-Venus in WT and *Abl* KD embryos. (**C**) Canoe-Venus TCJ ratios. (**D**) N-terminally tagged Venus-Canoe variants expressed using the Gal4/UAS system in a *canoe* mutant background. (**E**) Venus-Canoe TCJ ratios. (**F**) Canoe domain structure. Vertical lines indicate tyrosines. (**G**) Venus-Canoe-NC variants in which the indicated tyrosines were mutated to unphosphorylatable phenylalanine (F) residues were coexpressed with HA-Abl in S2R$^+$ cells, immunoprecipitated (IP) with anti-GFP, and immunoblotted (IB) with anti–p-Tyr or anti-GFP. Protein size is shown in kilodaltons. The tyrosines mutated in 11YF and 4YF are indicated in (F). (**H**) Stills from movies of *canoe* mutant embryos expressing full-length Venus-Canoe variants using the Gal4/UAS system. (**I** to **K**) Percentage of TCJs with gaps in E-cadherin signal in *canoe* KD (I), *Abl* KD (J), or *canoe* mutants expressing full-length Venus-Canoe variants (K) (mean ± SEM between embryos). Cno WT, *canoe* mutant expressing Venus-Canoe-WT; –, *canoe* mutant alone. *P < 0.04, **P ≤ 0.001, Welch's t test. Embryos are stages 7 and 8. Living embryos are shown in (D), (E), and (H). Anterior is left; ventral is down. Scale bars, 10 μm.

contains 47 tyrosines, making identifying the Abl target sites in Canoe a daunting task. To expedite this process, we performed a structure-function analysis to identify the minimal sequences required for Canoe function. A truncated protein that contains the N- and C-terminal domains of Canoe (Canoe-NC), but lacks the central region including the PDZ domain that interacts with several cell-surface receptors (*39*), localized correctly to tricellular junctions (fig. S7, A and D) and fully rescued myosin localization in *canoe* mutants (fig. S7, E and F), a well-established readout for Canoe function (*40*, *41*). By contrast, further deletion of the N-terminal domain of Canoe that binds to the Rap1 guanosine triphosphatase (GTPase) (*43*, *50*) or the C-terminal domain that binds to F-actin (*38*, *40*) eliminated Canoe localization to tricellular junctions and rendered Canoe nonfunctional (fig. S7, A and D to F). Moreover, tyrosine phosphorylation of Canoe-NC, but not the full-length protein, was significantly reduced in *Abl* KD embryos (fig. S6C). Therefore, even though other tyrosine kinases can phosphorylate the nonessential Canoe central domain, Abl is required for the phosphorylation of the minimal Canoe-NC protein that is fully functional during convergent extension. Further analysis of Canoe-NC variants lacking specific tyrosines demonstrated that

Abl-dependent tyrosine phosphorylation of the Canoe-NC protein in S2R$^+$ cells was significantly reduced by mutating the 11 C-terminal tyrosines (Canoe-NC^{11YF}), the last four tyrosines (Canoe-NC^{4YF}), or the most C-terminal tyrosine of Canoe (Canoe-NC^{Y1987F}), which is located in a predicted Abl target site (L/I/V-Y-x-x-P) (*46*) (Fig. 2, F and G, and fig. S6, A to C). These findings indicate that Abl promotes the phosphorylation of the C-terminal tyrosine of Canoe, which is conserved from worms to mammals.

To determine whether tyrosine phosphorylation is required for Canoe function, we tested whether full-length Canoe phosphovariants rescue the defects in *canoe* mutant embryos. Both WT Canoe and a full-length Canoe variant lacking the N-terminal tyrosines (Canoe^{9YF}) localized correctly to tricellular junctions (fig. S7, A to D) and fully rescued tricellular adhesion in *canoe* mutants (Fig. 2K). By contrast, full-length Canoe proteins lacking the 11 C-terminal tyrosines (Canoe^{11YF}) or the most C-terminal tyrosine (Canoe^{Y1987F}) were less enriched at tricellular junctions and failed to rescue tricellular adhesion and myosin localization in *canoe* mutants (Fig. 2, D, E, H, and K, and fig. S7, A and D to F). Canoe^{Y1987F} embryos had more severe defects in tricellular adhesion than did *Abl* KD embryos (Fig. 2, J and K), likely because of residual Abl expression or the ability of other kinases to phosphorylate the Canoe C-terminal domain. All Canoe phosphovariants localized correctly to bicellular junctions (fig. S7, A and B), indicating that a general localization to adherens junctions in the absence of a selective enrichment at tricellular junctions is not sufficient for Canoe function. These results demonstrate that phosphorylation of the most C-terminal tyrosine of Canoe is necessary for its localization and function in vivo.

Canoe localization to tricellular junctions is mechanosensitive

Because Canoe is dynamically recruited to and stabilizes adhesion at tricellular junctions, we hypothesized that it could be part of a force-regulated mechanism that strengthens adhesion at tricellular junctions under tension. To test this model, we used three approaches to determine whether Canoe localization is regulated by force. First, we analyzed the relationship between Canoe and myosin localization at tricellular junctions in living embryos. Canoe-Venus localization to tricellular junctions was pulsatile (Fig. 3, A and B), a characteristic feature of contractile actomyosin networks (*51*, *52*). Moreover, Canoe recruitment to tricellular junctions was strongly correlated with changes in myosin localization, with no delay between Canoe-Venus and MRLC-mCherry recruitment by cross-correlation analysis (Fig. 3C and movie S3). Consistent with this relationship, the enrichment of Canoe at tricellular junctions was significantly reduced in stage 15 embryos that have less cortical myosin (fig. S8, A to C). By contrast, less Canoe protein was recruited with each myosin pulse in *Abl* KD embryos, even though myosin localization and dynamics at tricellular junctions occurred normally (fig. S5, H to N, and movie S4). Thus, Canoe recruitment correlates with bursts of myosin accumulation at tricellular junctions in WT embryos, and Canoe localization appears to be uncoupled from myosin in the absence of Abl activity.

To further investigate this model, we tested whether Canoe localization to tricellular junctions

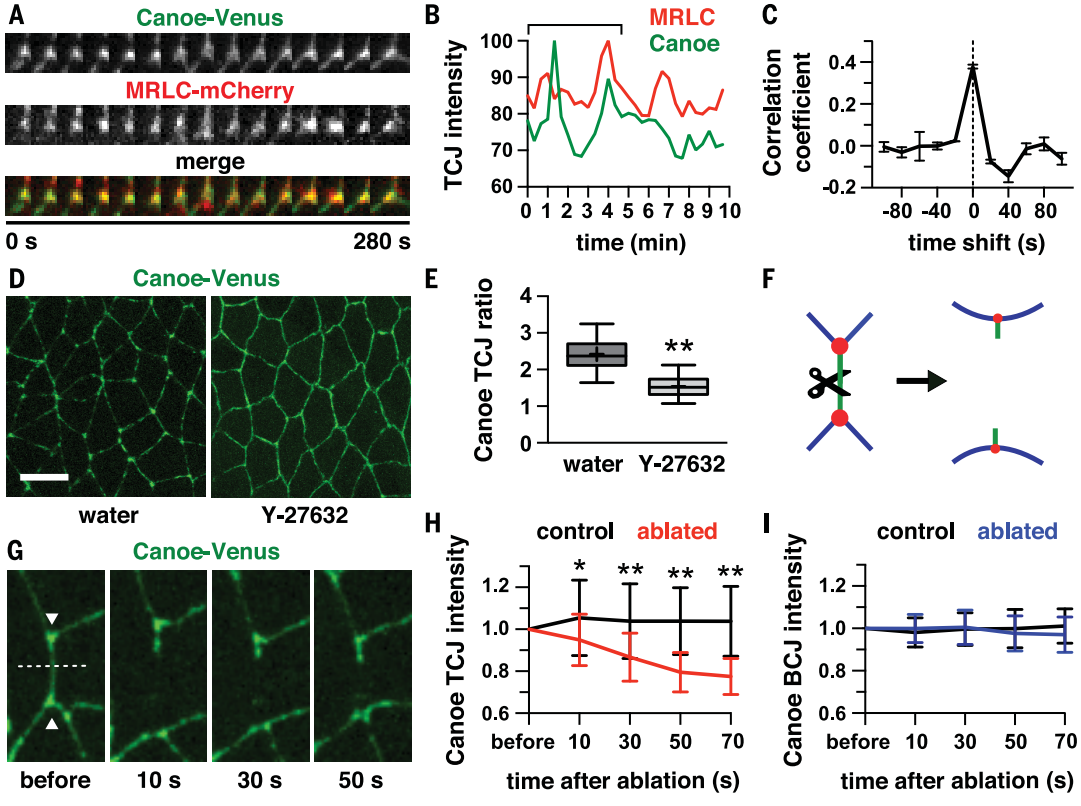


Fig. 3. Canoe is recruited to tricellular junctions by cytoskeletal tension. (**A** and **B**) Kymograph (A) and plot (B) of Canoe-Venus and MRLC-mCherry intensity at a single TCJ over time (imaged every 20 s). Intensity is a percentage of the maximum value. Bracket indicates the time window in (A). (**C**) Correlation coefficient for the rates of change in Canoe-Venus and MRLC-mCherry intensity at TCJs. Canoe-Venus data were shifted by the times on the *x*-axis (mean ± SEM between embryos). (**D**) Canoe-Venus in embryos injected with water or Y-27632. (**E**) Canoe-Venus TCJ ratios. (**F**) Laser ablation schematic. (**G**) Stills from movies of Canoe-Venus before and after ablation of an edge connecting two TCJs (arrowheads) (rotated 30°). (**H** and **I**) Canoe-Venus intensity at TCJs (H) and BCJs (I) before and after ablation, normalized to the value at t = 0 (mean ± SD between ablations). *P = 0.01, **$P \leq 0.0001$, Welch's t test (E), one-way ANOVA [(H) and (I)]. Embryos are stage 7. Living embryos expressing endogenous Canoe-Venus are shown in all panels. Anterior is left; ventral is down. Scale bars, 10 μm.

requires actomyosin activity. Reducing actomyosin contractility by injecting a Rho-kinase inhibitor resulted in a rapid loss of Canoe-Venus from tricellular junctions within 5 min of injection and suppressed fluctuations in Canoe-Venus intensity (Fig. 3, D and E; fig. S8, D to F; and movie S5). By contrast, Rho-kinase inhibitor did not decrease, and in fact slightly increased, Canoe-Venus levels at bicellular junctions (Fig. 3D and fig. S8, G, I, and J). These results indicate that actomyosin contractility is not necessary for Canoe cortical localization but is required for the selective accumulation of Canoe at tricellular junctions.

To directly test whether Canoe localization is regulated by mechanical force, we used laser ablation methods (*30*, *35*) to acutely release tension at individual tricellular junctions by severing one of the three connected bicellular junctions (Fig. 3, F and G, and movie S6). To quantify the effects of releasing tension, we measured Canoe-Venus intensity at the tricellular junctions and bicellular junctions attached to the cut edge before and after ablation. To control for photobleaching, we analyzed Canoe-Venus intensity in cells located two or more cells away from the ablation site for comparison (fig. S8H). Laser ablation led to a 23 ± 9% (mean ± SD) decrease in Canoe-Venus intensity at tricellular junctions within 70 s after ablation, whereas Canoe-Venus localization at bicellular junctions was unaffected (Fig. 3, H and I). Taken together, these results demonstrate that Canoe-Venus enrichment at tricellular junctions requires mechanical forces generated by actomyosin contractility.

Dynamic Canoe localization to tricellular junctions is required for cell intercalation

The findings that Canoe is recruited to tricellular junctions by tyrosine phosphorylation and actomyosin contractility raise two questions. First, what is the physiological relevance of coupling Canoe recruitment with changes in myosin activity? Second, does tyrosine phosphorylation only regulate Canoe localization to tricellular junctions, or is it also required for Canoe function at these sites? To address these questions, we developed a method to target Canoe to tricellular junctions independently of endogenous mechanical inputs. This "vertex trap" method involved expressing VhhGFP4 (VHH), a genetically encoded nanobody with high affinity for GFP (*53*), fused to the C terminus of the tricellular junction protein Sidekick (Sdk) (*32*–*34*). Sdk-VHH expression is predicted to recruit GFP- and Venus-tagged proteins to tricellular junctions through interactions with the localized nanobody (Fig. 4B). The localization of Canoe-Venus was not significantly altered by Sdk-VHH expression in WT embryos (Fig. 4, A and E), and cell adhesion and myosin localization in these embryos occurred normally (fig. S9, A to E). By contrast, Sdk-VHH allowed Canoe-Venus to remain associated with tricellular junctions when myosin activity was inhibited by Y-27632 injection, a situation in which this enrichment is normally lost in WT embryos (Fig. 4, A and F). Thus, the vertex trap system uncouples Canoe localization from cytoskeletal tension, providing an opportunity to investigate the effects of eliminating Canoe mechanosensitivity on cell rearrangement.

To investigate the effects of uncoupling Canoe localization from mechanical inputs, we performed time-lapse imaging of embryos expressing vertex-trapped Canoe. During convergent extension, epithelial cells undergo cell rearrangements that are driven by planar polarized actomyosin contractility (*26*–*30*). In a stereotypical cell rearrangement during this process (*27*), tricellular junctions merge to form four-way vertices that subsequently resolve to form two new tricellular junctions, a process that requires active remodeling of cell adhesion (Fig. 4, C and I, and fig. S10A). Protein localization is highly dynamic during this process, with MRLC-mCherry and Canoe-Venus reproducibly dissociating from four-way vertices just before resolution (Fig. 4C and fig. S10, B and C). In embryos that express vertex-trapped Canoe, cells were frequently arrested at the four-way vertex stage, resulting in an increase in the average vertex duration time to 9.1 ± 0.7 min compared with 5.1 ± 0.4 min in WT embryos (mean ± SEM) (Fig. 4, D, G, and I, and movie S7). Embryos that expressed vertex-trapped Canoe also often failed to complete resolution and displayed increased cell stretching, consistent with a reduction in cell rearrangement (fig. S9, F and G). Further increasing the levels of Canoe at tricellular junctions by coexpressing Sdk-VHH with overexpressed Canoe extended the vertex duration time to 17.8 ± 0.9 min, more than three times longer than in WT embryos (Fig. 4H and movie S8). By contrast, *Abl* KD embryos that recruited less Canoe to tricellular junctions accelerated vertex resolution, reducing the average vertex duration time to 2.6 ± 0.2 min (fig. S10, D and E) (*48*). It was not possible to analyze vertex duration in Canoe-deficient embryos because of the widespread defects in tricellular adhesion. These results indicate that varying Canoe levels at tricellular junctions modulates the timing of cell rearrangement.

We next tested whether tyrosine phosphorylation is solely required for Canoe localization to tricellular junctions, or if it is also necessary for Canoe function at these sites. If tyrosine phosphorylation is only required to recruit Canoe to tricellular junctions, then trapping Canoe phosphovariants at these sites would be predicted to bypass the requirement for phosphorylation. Using the vertex trap assay, we found that Canoe^{11YF}-Venus, which lacks the 11 C-terminal tyrosines of Canoe including Y1987, stabilized cell adhesion at four-cell vertices and delayed cell rearrangement to the same extent as WT Canoe (Fig. 4H and movie S9). By contrast, deleting the entire C-terminal region of Canoe (Canoe-ΔC), including the F-actin–binding domain, eliminated the effects of vertex-trapped Canoe on cell adhesion (Fig. 4H and fig. S9, H and I). Thus, tyrosine phosphorylation is essential for Canoe localization to tricellular junctions but is not required for Canoe to stabilize adhesion at these sites when Canoe is targeted to tricellular junctions through a heterologous mechanism. Together, these results demonstrate that tyrosine phosphorylation modulates Canoe levels at tricellular junctions, which influences the rate of cell rearrangement during epithelial remodeling.

Discussion

Here, we show that Canoe and Abl function in a mechanotransduction pathway that regulates dynamic changes in cell adhesion at tricellular junctions during epithelial remodeling. Canoe localization to tricellular junctions is acutely disrupted by laser ablation, demonstrating that Canoe localization is rapidly modulated by mechanical perturbation. Canoe mechanosensitivity requires Abl-dependent phosphorylation of a conserved tyrosine (Y1987) in the Canoe actin-binding domain, and Canoe localization to tricellular junctions is required to stabilize tricellular adhesion. Conversely, constitutively anchoring Canoe at tricellular junctions arrests cell rearrangement, indicating that Canoe levels can tune the rate of junctional remodeling (Fig. 4J). These results demonstrate that the mechanosensitivity of this critical junctional regulator is modulated by phosphorylation of a single tyrosine residue and reveal an essential role of Canoe in coupling tricellular adhesion with mechanical forces during epithelial remodeling.

Mechanical forces trigger a cascade of molecular events in cells that translate biophysical signals into altered cellular behaviors. However, the mechanosensors that directly change conformation under tension in vivo are not well defined. Cell surface receptors are well positioned to detect forces generated by neighboring cells, but Canoe localization and function at tricellular junctions do not require the PDZ domain that mediates its interaction with known receptors (*39*). One possibility is that the Canoe protein itself could act as a mechanosensor. Canoe is predicted to be anchored to the membrane through interactions with Rap1 and the actin cytoskeleton (*38*–*40*, *43*, *50*). Thus, cytoskeletal tension could stretch the Canoe protein and expose tyrosine 1987 to phosphorylation by Abl. Alternatively, Abl or its upstream activators could be regulated by tension during this process, because the activity of Abl and other tyrosine kinases such as Src and FAK has been shown

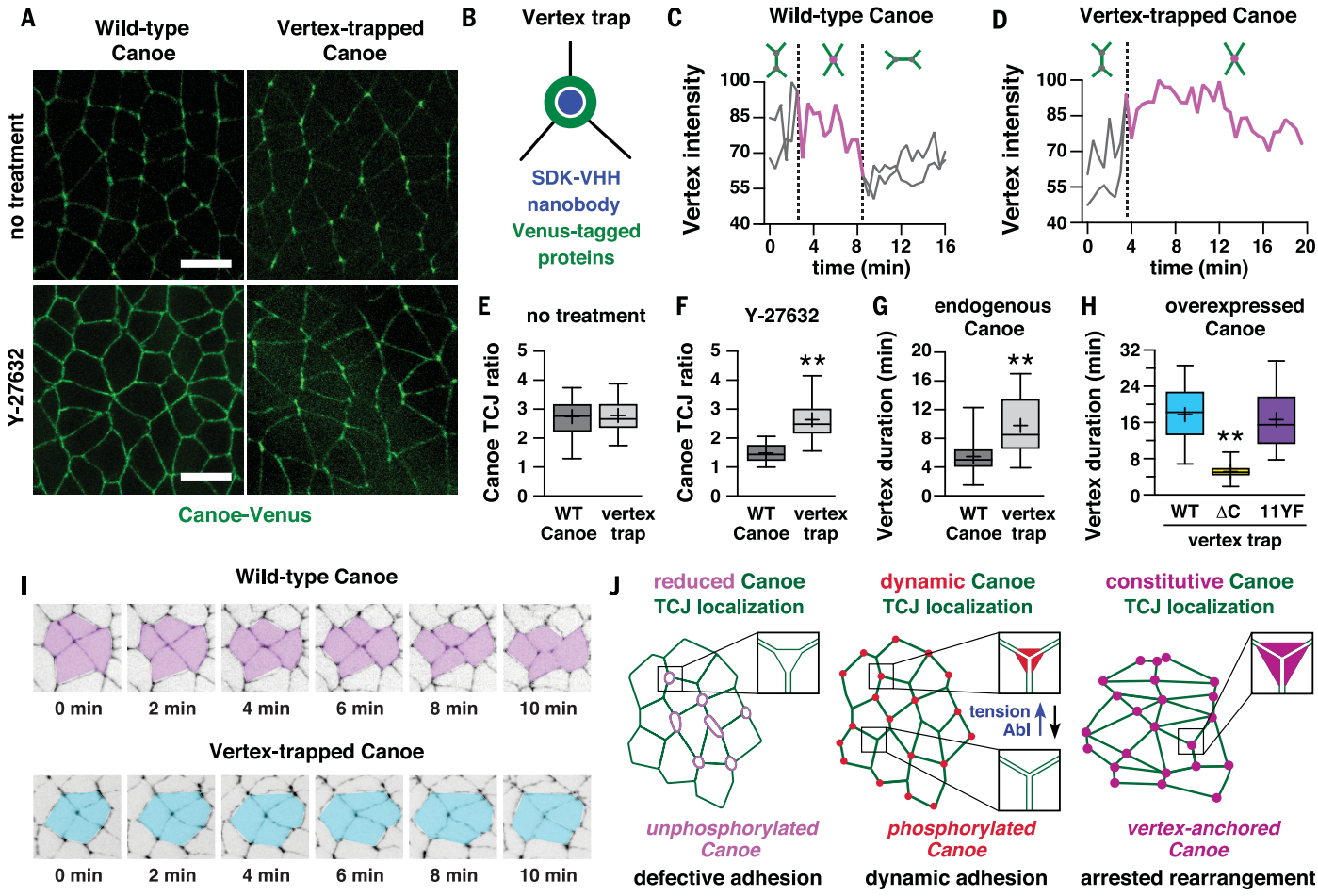


Fig. 4. Anchoring Canoe at tricellular junctions impedes cell rearrangement. (**A**) Embryos expressing endogenous Canoe-Venus (WT Canoe, left) or Canoe-Venus trapped at tricellular junctions by Sdk-VHH (vertex-trapped Canoe, right). Top panels are uninjected embryos. Bottom panels are embryos injected with Y-27632. (**B**) Schematic of the vertex trap technique. The GFP nanobody (VHH) fused to the TCJ protein Sdk recruits GFP- and Venus-tagged proteins to tricellular junctions. (**C** and **D**) Endogenous Canoe-Venus intensity in the absence (C) or presence (D) of Sdk-VHH. Two TCJs (gray) converge to form a four-way vertex (magenta), which resolves to form two new TCJs (imaged every 30 s). (**E** to **G**) Canoe-Venus TCJ ratio [(E) and (F)] and four-way vertex duration (G) in the absence (WT Canoe) or presence (vertex trap) of Sdk-VHH. (**H**) Four-way vertex duration in embryos that overexpress vertex-trapped Venus-Canoe variants in a WT background. (**I**) Stills from movies of embryos expressing endogenous Canoe-Venus in the absence (WT Canoe) or presence (vertex-trapped Canoe) of Sdk-VHH. (**J**) Model. Actomyosin contractility and Abl-mediated tyrosine phosphorylation coordinate Canoe levels with mechanical tension at tricellular junctions, promoting the remodeling of cell adhesion during cell rearrangement. $^{**}P < 0.0001$, Welch's t test. Embryos are stages 7 and 8. Living embryos are shown in all panels. Anterior is left; ventral is down. Scale bars, 10 μm.

to be regulated by tension in vitro (*18*–*22*). In a third possibility, the regulation of Canoe by Abl could allow Canoe to detect force-induced changes in other molecules at tricellular junctions. Because tyrosine 1987 is in the Canoe actin-binding domain, phosphorylation at this site could allow Canoe to recognize distinct actin structures at tricellular junctions, force-induced conformational changes in the cadherin-catenin complex (*54*, *55*), or other specialized features of tricellular junction composition or geometry (*37*). Once recruited to tricellular junctions, Canoe could stabilize adhesion at these sites by reinforcing the connection between adherens junctions and the actomyosin cytoskeleton (*40*–*42*).

Abl tyrosine kinases influence many structural changes that are driven by mechanical forces in tissues, including epithelial remodeling, tissue invagination, axon guidance, and cell migration (*45*, *46*). Abl has been shown to regulate a number of proteins that act at tension-bearing structures in cells in addition to Canoe/Afadin recruitment to tricellular junctions, including vinculin localization (*20*, *22*, *25*) and β-catenin recycling (*48*) at bicellular junctions and regulation of membrane curvature by BAR-domain proteins (*21*). Together, these results raise the possibility that Abl could transduce mechanical forces into a wide range of structural changes within cells. For example, during cell rearrangement, Abl could simultaneously stabilize tricellular adhesion by recruiting Canoe and destabilize bicellular adhesion by enhancing β-catenin turnover (*48*), which could allow bicellular junctions to complete contraction before tricellular junctions are remodeled. Tricellular junctions serve many important roles in epithelial development and homeostasis, including modulating cell rearrangement (*32*–*34*, *52*), orienting mitotic spindles (*56*), balancing stem cell proliferation and differentiation (*57*), and maintaining epithelial barrier function (*58*). An understanding of how mechanical inputs affect the conformation, localization, activity, and interactions of proteins at tricellular junctions will provide insight into how these structures sense and integrate mechanical forces in epithelial tissues.

Methods summary

The methods used in this study are described in the supplementary materials, including *Drosophila* stocks and genetics, cloning and transgenic lines, immunofluorescence, time-lapse imaging, laser ablation, drug and antibody injection, immunoprecipitation, immunoblotting,

mass spectrometry, and quantification and statistical analysis.

REFERENCES AND NOTES

1. M. A. Schwartz, Integrins and extracellular matrix in mechanotransduction. *Cold Spring Harb. Perspect. Biol.* **2**, a005066 (2010). doi: 10.1101/cshperspect.a005066; pmid: 21084386
2. C.-P. Heisenberg, Y. Bellaïche, Forces in tissue morphogenesis and patterning. *Cell* **153**, 948–962 (2013). doi: 10.1016/j.cell.2013.05.008; pmid: 23706734
3. N. I. Petridou, Z. Spiró, C.-P. Heisenberg, Multiscale force sensing in development. *Nat. Cell Biol.* **19**, 581–588 (2017). doi: 10.1038/ncb3524; pmid: 28561050
4. D. Pinheiro, Y. Bellaïche, Mechanical force-driven adherens junction remodeling and epithelial dynamics. *Dev. Cell* **47**, 3–19 (2018). doi: 10.1016/j.devcel.2018.09.014; pmid: 30300588
5. C. Hahn, M. A. Schwartz, Mechanotransduction in vascular physiology and atherogenesis. *Nat. Rev. Mol. Cell Biol.* **10**, 53–62 (2009). doi: 10.1038/nrm2596; pmid: 19197332
6. D. Schramek *et al.*, Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. *Science* **343**, 309–313 (2014). doi: 10.1126/science.1248627; pmid: 24436421
7. M. Takeichi, Dynamic contacts: Rearranging adherens junctions to drive epithelial remodelling. *Nat. Rev. Mol. Cell Biol.* **15**, 397–410 (2014). doi: 10.1038/nrm3802; pmid: 24824068
8. W. J. Nelson, W. I. Weis, 25 years of tension over actin binding to the cadherin cell adhesion complex: The devil is in the details. *Trends Cell Biol.* **26**, 471–473 (2016). doi: 10.1016/j.tcb.2016.04.010; pmid: 27166091
9. A. S. Yap, K. Duszyc, V. Viasnoff, Mechanosensing and mechanotransduction at cell-cell junctions. *Cold Spring Harb. Perspect. Biol.* **10**, a028761 (2018). doi: 10.1101/cshperspect.a028761; pmid: 28778874
10. Q. le Duc *et al.*, Vinculin potentiates E-cadherin mechanosensing and is recruited to actin-anchored sites within adherens junctions in a myosin II-dependent manner. *J. Cell Biol.* **189**, 1107–1115 (2010). doi: 10.1083/jcb.201001149; pmid: 20584916
11. S. Yonemura, Y. Wada, T. Watanabe, A. Nagafuchi, M. Shibata, α-Catenin as a tension transducer that induces adherens junction development. *Nat. Cell Biol.* **12**, 533–542 (2010). doi: 10.1038/ncb2055; pmid: 20453849
12. C. D. Buckley *et al.*, Cell adhesion. The minimal cadherin-catenin complex binds to actin filaments under force. *Science* **346**, 1254211–1254211 (2014). doi: 10.1126/science.1254211; pmid: 25359979
13. K. Manibog, H. Li, S. Rakshit, S. Sivasankar, Resolving the molecular mechanism of cadherin catch bond formation. *Nat. Commun.* **5**, 3941 (2014). doi: 10.1038/ncomms4941; pmid: 24887573
14. M. Yao *et al.*, Force-dependent conformational switch of α-catenin controls vinculin binding. *Nat. Commun.* **5**, 4525 (2014). doi: 10.1038/ncomms5525; pmid: 25077739
15. N. Ishiyama *et al.*, Force-dependent allostery of the α-catenin actin-binding domain controls adherens junction dynamics and functions. *Nat. Commun.* **9**, 5121 (2018). doi: 10.1038/s41467-018-07481-7; pmid: 30504777
16. G. Giannone, M. P. Sheetz, Substrate rigidity and force define form through tyrosine phosphatase and kinase pathways. *Trends Cell Biol.* **16**, 213–223 (2006). doi: 10.1016/j.tcb.2006.02.005; pmid: 16529933
17. R. W. McLachlan, A. S. Yap, Not so simple: The complexity of phosphotyrosine signaling at cadherin adhesive contacts. *J. Mol. Med. (Berl.)* **85**, 545–554 (2007). doi: 10.1007/s00109-007-0198-x; pmid: 17429596
18. Y. Wang *et al.*, Visualizing the mechanical activation of Src. *Nature* **434**, 1040–1045 (2005). doi: 10.1038/nature03469; pmid: 15846350
19. J. Seong *et al.*, Distinct biophysical mechanisms of focal adhesion kinase mechanoactivation by different extracellular matrix proteins. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19372–19377 (2013). doi: 10.1073/pnas.1307405110; pmid: 24222685
20. P. Sehgal *et al.*, Epidermal growth factor receptor and integrins control force-dependent vinculin recruitment to E-cadherin junctions. *J. Cell Sci.* **131**, jcs206656 (2018). doi: 10.1242/jcs.206656; pmid: 29487179
21. A. Echarri *et al.*, An Abl-FBP17 mechanosensing system couples local plasma membrane curvature and stress fiber remodeling during mechanoadaptation. *Nat. Commun.* **10**, 5828 (2019). doi: 10.1038/s41467-019-13782-2; pmid: 31862885
22. J. L. Bays *et al.*, Vinculin phosphorylation differentially regulates mechanotransduction at cell-cell and cell-matrix adhesions. *J. Cell Biol.* **205**, 251–263 (2014). doi: 10.1083/jcb.201309092; pmid: 24751539
23. M. Tamada, M. P. Sheetz, Y. Sawada, Activation of a signaling cascade by cytoskeleton stretch. *Dev. Cell* **7**, 709–718 (2004). doi: 10.1016/j.devcel.2004.08.021; pmid: 15525532
24. Y. Sawada *et al.*, Force sensing by mechanical extension of the Src family kinase substrate p130Cas. *Cell* **127**, 1015–1026 (2006). doi: 10.1016/j.cell.2006.09.044; pmid: 17129785
25. C. Bertocchi *et al.*, Nanoscale architecture of cadherin-based cell adhesions. *Nat. Cell Biol.* **19**, 28–37 (2017). doi: 10.1038/ncb3456; pmid: 27992406
26. J. A. Zallen, E. Wieschaus, Patterned gene expression directs bipolar planar polarity in *Drosophila*. *Dev. Cell* **6**, 343–355 (2004). doi: 10.1016/S1534-5807(04)00060-7; pmid: 15030758
27. C. Bertet, L. Sulak, T. Lecuit, Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* **429**, 667–671 (2004). doi: 10.1038/nature02590; pmid: 15190355
28. J. T. Blankenship, S. T. Backovic, J. S. P. Sanny, O. Weitz, J. A. Zallen, Multicellular rosette formation links planar cell polarity to tissue morphogenesis. *Dev. Cell* **11**, 459–470 (2006). doi: 10.1016/j.devcel.2006.09.007; pmid: 17011486
29. M. Rauzi, P. Verant, T. Lecuit, P.-F. Lenne, Nature and anisotropy of cortical forces orienting *Drosophila* tissue morphogenesis. *Nat. Cell Biol.* **10**, 1401–1410 (2008). doi: 10.1038/ncb1798; pmid: 18978783
30. R. Fernandez-Gonzalez, Sde. M. Simoes, J.-C. Röper, S. Eaton, J. A. Zallen, Myosin II dynamics are regulated by tension in intercalating cells. *Dev. Cell* **17**, 736–743 (2009). doi: 10.1016/j.devcel.2009.09.003; pmid: 19879198
31. C. M. Lye, H. W. Naylor, B. Sanson, Subcellular localisations of the CPTI collection of YFP-tagged proteins in *Drosophila* embryos. *Development* **141**, 4006–4017 (2014). doi: 10.1242/dev.111310; pmid: 25294944
32. A. Letizia *et al.*, Sidekick is a key component of tricellular adherens junctions that acts to resolve cell rearrangements. *Dev. Cell* **50**, 313–326.e5 (2019). doi: 10.1016/j.devcel.2019.07.007; pmid: 31353315
33. H. Uechi, E. Kuranaga, The tricellular junction protein sidekick regulates vertex dynamics to promote bicellular junction extension. *Dev. Cell* **50**, 327–338.e5 (2019). doi: 10.1016/j.devcel.2019.06.017; pmid: 31353316
34. T. M. Finegan *et al.*, The tricellular vertex-specific adhesion molecule Sidekick facilitates polarised cell intercalation during *Drosophila* axis extension. *PLOS Biol.* **17**, e3000522 (2019). doi: 10.1371/journal.pbio.3000522; pmid: 31805038
35. R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, F. Jülicher, The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Curr. Biol.* **17**, 2095–2104 (2007). doi: 10.1016/j.cub.2007.11.049; pmid: 18082406
36. G. Trichas *et al.*, Multi-cellular rosettes in the mouse visceral endoderm facilitate the ordered migration of anterior visceral endoderm cells. *PLOS Biol.* **10**, e1001256 (2012). doi: 10.1371/journal.pbio.1001256; pmid: 22346733
37. T. Higashi, A. L. Miller, Tricellular junctions: How to build junctions at the TRICkiest points of epithelial cells. *Mol. Biol. Cell* **28**, 2023–2034 (2017). doi: 10.1091/mbc.e16-10-0697; pmid: 28705832
38. K. Mandai *et al.*, Afadin: A novel actin filament-binding protein with one PDZ domain localized at cadherin-based cell-to-cell adherens junction. *J. Cell Biol.* **139**, 517–528 (1997). doi: 10.1083/jcb.139.2.517; pmid: 9334353
39. K. Mandai, Y. Rikitake, Y. Shimono, Y. Takai, Afadin/AF-6 and canoe: Roles in cell adhesion and beyond. *Prog. Mol. Biol. Transl. Sci.* **116**, 433–454 (2013). doi: 10.1016/B978-0-12-394311-8.00019-4; pmid: 23481206
40. J. K. Sawyer, N. J. Harris, K. C. Slep, U. Gaul, M. Peifer, The *Drosophila* afadin homologue Canoe regulates linkage of the actin cytoskeleton to adherens junctions during apical constriction. *J. Cell Biol.* **186**, 57–73 (2009). doi: 10.1083/jcb.200904001; pmid: 19596848
41. J. K. Sawyer *et al.*, A contractile actomyosin network linked to adherens junctions by Canoe/afadin helps drive convergent extension. *Mol. Biol. Cell* **22**, 2491–2508 (2011). doi: 10.1091/mbc.e11-05-0411; pmid: 21613546
42. W. Choi *et al.*, Remodeling the zonula adherens in response to tension and the role of afadin in this response. *J. Cell Biol.* **213**, 243–260 (2016). doi: 10.1083/jcb.201506115; pmid: 27114502
43. T. T. Bonello, K. Z. Perez-Vale, K. D. Sumigray, M. Peifer, Rap1 acts via multiple mechanisms to position Canoe and adherens junctions and mediate apical-basal polarity establishment. *Development* **145**, dev157941 (2018). doi: 10.1242/dev.157941; pmid: 29361565
44. J. M. Golas *et al.*, SKI-606, a 4-anilino-3-quinolinecarbonitrile dual inhibitor of Src and Abl kinases, is a potent antiproliferative agent against chronic myelogenous leukemia cells in culture and causes regression of K562 xenografts in nude mice. *Cancer Res.* **63**, 375–381 (2003). pmid: 12543790
45. W. D. Bradley, A. J. Koleske, Regulation of cell migration and morphogenesis by Abl-family kinases: Emerging mechanisms and physiological contexts. *J. Cell Sci.* **122**, 3441–3454 (2009). doi: 10.1242/jcs.039859; pmid: 19759284
46. J. Colicelli, ABL tyrosine kinases: Evolution of function, regulation, and specificity. *Sci. Signal.* **3**, re6 (2010). doi: 10.1126/scisignal.3139re6; pmid: 20841568
47. D. T. Fox, M. Peifer, Abelson kinase (Abl) and RhoGEF2 regulate actin organization during cell constriction in *Drosophila*. *Development* **134**, 567–578 (2007). doi: 10.1242/dev.02748; pmid: 17202187
48. M. Tamada, D. L. Farrell, J. A. Zallen, Abl regulates planar polarized junctional dynamics through β-catenin tyrosine phosphorylation. *Dev. Cell* **22**, 309–319 (2012). doi: 10.1016/j.devcel.2011.12.025; pmid: 22340496
49. J. N. Jodoin, A. C. Martin, Abl suppresses cell extrusion and intercalation during epithelium folding. *Mol. Biol. Cell* **27**, 2822–2832 (2016). doi: 10.1091/mbc.e16-05-0336; pmid: 27440923
50. B. Boettner *et al.*, The AF-6 homolog canoe acts as a Rap1 effector during dorsal closure of the *Drosophila* embryo. *Genetics* **165**, 159–169 (2003). pmid: 14504224
51. A. C. Martin, M. Kaschube, E. F. Wieschaus, Pulsed contractions of an actin-myosin network drive apical constriction. *Nature* **457**, 495–499 (2009). doi: 10.1038/nature07522; pmid: 19029882
52. T. E. Vanderleest *et al.*, Vertex sliding drives intercalation by radial coupling of adhesion and actomyosin networks during *Drosophila* germband extension. *eLife* **7**, e34586 (2018). doi: 10.7554/eLife.34586; pmid: 29985789
53. D. Saerens *et al.*, Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. *J. Mol. Biol.* **352**, 597–607 (2005). doi: 10.1016/j.jmb.2005.07.038; pmid: 16095608
54. S. Pokutta, F. Drees, Y. Takai, W. J. Nelson, W. I. Weis, Biochemical and structural definition of the l-afadin- and actin-binding sites of α-catenin. *J. Biol. Chem.* **277**, 18868–18874 (2002). doi: 10.1074/jbc.M201463200; pmid: 11907041
55. S. Sakakibara *et al.*, Afadin regulates actomyosin organization through αE-catenin at adherens junctions. *J. Cell Biol.* **219**, e201907079 (2020). doi: 10.1083/jcb.201907079; pmid: 32227204
56. F. Bosveld *et al.*, Epithelial tricellular junctions act as interphase cell shape sensors to orient mitosis. *Nature* **530**, 495–498 (2016). doi: 10.1038/nature16970; pmid: 26886796
57. M. Resnik-Docampo *et al.*, Tricellular junctions regulate intestinal stem cell behaviour to maintain homeostasis. *Nat. Cell Biol.* **19**, 52–59 (2017). doi: 10.1038/ncb3454; pmid: 27992405
58. R. E. Stephenson *et al.*, Rho flares repair local tight junction leaks. *Dev. Cell* **48**, 445–459.e5 (2019). doi: 10.1016/j.devcel.2019.01.016; pmid: 30773490

ACKNOWLEDGMENTS

We thank M. Tamada and C. Fincher for initially observing the effects of increased myosin activity on tyrosine phosphorylation, B. Glick for msVenus, M. Laurie for the *UASp* vector, M. Gredler for help with graphic illustration, H. Molina for help with mass spectrometry, and E. Brooks, M. Gredler, A. Paré, W. Razzell, G. Sabio, M. Tamada, and R. Zallen for comments on the manuscript. Some stocks were obtained from the Bloomington *Drosophila* Stock Center (NIH P40OD018537) and the Kyoto Stock Center. **Funding:** This work was funded by NIH/NIGMS R01 grant no. GM079340 to J.A.Z. J.A.Z. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** H.H.Y. performed the experiments. H.H.Y. and J.A.Z. conceived and designed the project, analyzed the data, and wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/eaba5528/suppl/DC1
Materials and Methods
Figs. S1 to S10
References (*59*–*69*)
Movies S1 to S9
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

12 December 2019; accepted 11 October 2020
10.1126/science.aba5528

RESEARCH ARTICLE

FUNDAMENTAL PHYSICS

Two-photon frequency comb spectroscopy of atomic hydrogen

Alexey Grinin[1]*, Arthur Matveev[1], Dylan C. Yost[1]†, Lothar Maisenbacher[1], Vitaly Wirthl[1], Randolf Pohl[1]‡, Theodor W. Hänsch[1,2], Thomas Udem[1,2]

We have performed two-photon ultraviolet direct frequency comb spectroscopy on the 1S-3S transition in atomic hydrogen to illuminate the so-called proton radius puzzle and to demonstrate the potential of this method. The proton radius puzzle is a significant discrepancy between data obtained with muonic hydrogen and regular atomic hydrogen that could not be explained within the framework of quantum electrodynamics. By combining our result [$f_{1S\text{-}3S}$ = 2,922,743,278,665.79(72) kilohertz] with a previous measurement of the 1S-2S transition frequency, we obtained new values for the Rydberg constant [R_∞ = 10,973,731.568226(38) per meter] and the proton charge radius [r_p = 0.8482(38) femtometers]. This result favors the muonic value over the world-average data as presented by the most recent published CODATA 2014 adjustment.

Quantum electrodynamics (QED) is widely considered to be the most developed fundamental theory. It served as the blueprint for other quantum field theories and is capable of predicting the energy levels of atomic hydrogen and the electron *g*-factor with an accuracy of 12 digits. It is of fundamental interest to experimentally verify all of these digits or to discover tiny discrepancies that may lead to new physics.

The QED theory for the energy levels of atomic hydrogen can be cast in the form

$$E_{n,\ell,j} = chR_\infty\left(-\frac{1}{n^2}+f_{n,\ell,j}\left(\alpha,\frac{m_e}{m_p},...\right)+\delta_{\ell,0}\frac{C_{NS}}{n^3}r_p^2\right) \quad (1)$$

with n, ℓ and j being the principle quantum number and the orbital and total angular momentum, respectively. The well-known leading term is obtained from both the Bohr and Schrödinger theories. The function $f_{n,\ell,j}$ is a lengthy but mostly analytic expression (*1*) that depends, among other things, on the fine structure constant α and the electron-to-proton mass ratio m_e/m_p. These constants can be obtained with sufficient accuracy from other experiments where they do not merely represent a correction to a much larger term (*2*, *3*). The last term is a small but important correction to S-states owing to the finite root-mean-square (RMS) charge radius of the proton r_p. To convert to SI units (International System of Units) used in the experiment, the full theory expression needs to be multiplied by the Rydberg energy, chR_∞.

By accurately measuring transition frequencies, i.e., differences of energy levels, values for R_∞ and r_p can be obtained simply by treating them as parameters and adjusting their values to match the experimental observations. Two measured transition frequencies are required to determine the values of the two parameters, and the second most accurate measurement sets the limit on the uncertainty of R_∞ and r_p. If QED, its usage, and the experiments are all correct, a self-consistent set of the parameter values is expected when using all available measured transition frequencies (*4*). However, a serious discrepancy to existing hydrogen data was found when high-resolution laser spectroscopy on muonic hydrogen first became possible (*5*). This exotic atom is analogous to regular atomic hydrogen, but with the electron replaced by its heavier, short-lived sibling, the muon. Its energy levels should obey the same QED formalism but with an approximately four orders of magnitude larger coefficient C_{NS} and a corresponding increased sensitivity to r_p. A discrepancy of four standard deviations (4σ) was found when comparing r_p obtained in this way with the values obtained from regular atomic hydrogen. This fact has motivated additional high-precision spectroscopy of atomic hydrogen (*6*–*8*). These more recent measurements are discussed below.

The proton charge radius can also be obtained from electron-proton scattering with results that either support the muonic value (*9*, *10*) or elevate the discrepancy associated with the proton radius puzzle (*1*). Because there is no QED expression like Eq. 1 for the proton charge distribution, electron scattering data cannot directly contribute to testing QED. Using consistent scattering data could improve the values of R_∞ and r_p. However, here we are less interested in these values and more interested in checking for consistency of their QED-dependent determinations.

We report a measurement that considerably reduces the uncertainty of the 1S-3S transition frequency, which is the second most precisely known transition frequency in atomic hydrogen. It is surpassed only by the 1S-2S transition (*11*), which has a natural line width that is several orders of magnitude smaller than any other relevant transition in atomic hydrogen. We have substantially reduced the systematic uncertainties and required corrections by using a cold atomic beam and by utilizing the advantages of the direct frequency comb technique. An almost shot noise limited statistical uncertainty of only 110 Hz has been obtained with small systematic effects whose compensation does not critically rely on model

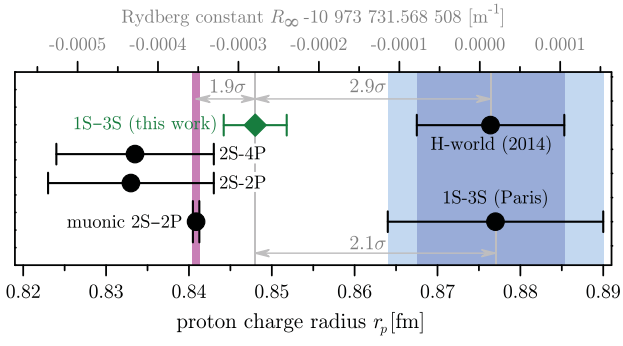


Fig. 1. Proton charge radius measurements. The proton charge radius measurements (bottom axis), as obtained from various experiments, partially disagree, preventing a verification of quantum electrodynamics at the level of experimental uncertainties. Combining the 1S-2S and the 1S-3S (this work) transition frequencies in atomic hydrogen gives a significantly smaller value than the previous world data [H-world (2014)] obtained by using all hydrogen measurements (15 lines) available for the most recent published CODATA 2014 adjustment [Adj #8, table XXIX in (*1*)]. In contrast, a recent continuous wave measurement of the 1S-3S transition frequency [1S-3S (Paris)] (*7*) confirms the CODATA value. Our result is in reasonable agreement with a value derived from the previous 2S-4P measurement (*6*), with a radio frequency measurement of the 2S-2P transition (*8*), and with the value obtained from muonic hydrogen (*32*). Because of the strong correlation (98.91%) between R_∞ and r_p, the data can also be displayed in terms of the Rydberg constant (same graph, upper axis).

[1]Laser Spectroscopy Division Max-Planck-Institut für Quantenoptik, Garching, Germany. [2]Department of Physics, Ludwig-Maximilians-Universität, München, Germany.
†Present address: Department of Physics, Colorado State University, Fort Collins, CO, USA.
‡Present address: Institut für Physik, Johannes Gutenberg-Universität Mainz, Mainz, Germany.
*Corresponding author. Email: alexey.grinin@mpq.mpg.de

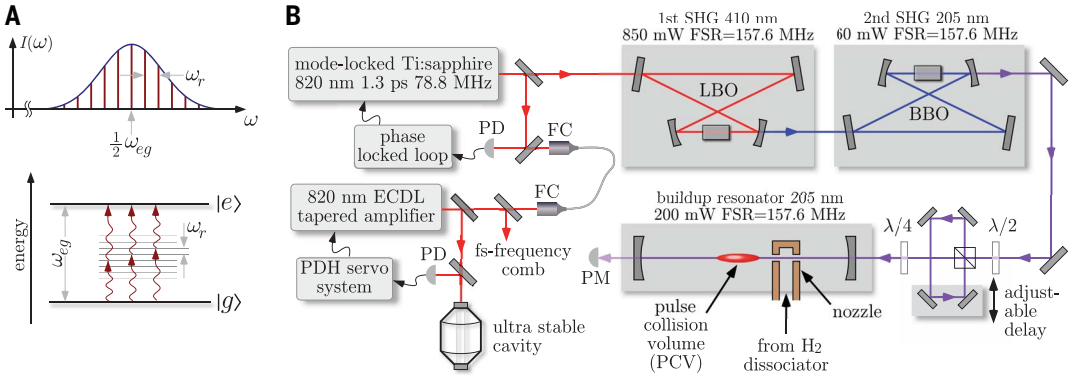


Fig. 2. Principle and experimental setup for two-photon direct frequency comb spectroscopy. (**A**) Spectral envelope of the frequency comb with repetition rate ω_r (not to scale) tuned to excite a two-photon transition between $|g\rangle$ and $|e\rangle$ at the frequency ω_{eg}. On resonance, pairwise addition of properly phased modes provides an efficient excitation of the atoms. (**B**) A mode-locked titanium sapphire laser (78.8 MHz, 1.3 ps, 2.8 W) is referenced to a transfer laser that is itself locked to an ultrastable cavity and referenced to a femtosecond-frequency comb. This frequency comb is then frequency quadrupled in two successive intracavity doubling stages to generate a deep ultraviolet frequency comb at 205 nm. The optical cavities used for frequency doubling are built with half the length of the fundamental laser cavity, which effectively doubles the repetition rate of the quadrupled frequency comb to 157.6 MHz. The pulse train is then sent to a beam splitter and delay line used to generate counterpropagating pulses within a final enhancement cavity where the hydrogen spectroscopy takes place. PM, power meter; FC, fiber coupler; FSR, free spectral range; PDH, Pound-Drever-Hall stabilization (*33*); ECDL, extended cavity diode laser; SHG, second-harmonic generation; LBO/BBO, lithium triborate and β–barium borate crystals; PD, photodetector.

assumptions. In addition, we used the simple Lorentzian line shape model and found the line center within 10^{-3} of the line width, which is a rather moderate value.

With this experiment, we also demonstrate high-resolution laser spectroscopy with a harmonic frequency comb in the ultraviolet region. The associated short pulses of the frequency comb make the harmonic generation process in crystals and gas targets more efficient, while avoiding the photorefractive effect (*12*). In the future, this method may allow precision spectroscopy at even shorter unexplored wavelength regions using high harmonic generation—hopefully enabling high-resolution laser spectroscopy of hydrogen-like ions (*13*).

Combining the results for the 1S-3S and the 1S-2S transitions, we extracted values for the Rydberg constant R_∞ and the proton charge radius r_p. These new values are two times more accurate than the ones obtained from all previous hydrogen data combined. By using only two measurements to determine two constants (treated as parameters), nothing can be said about the validity of QED. It does not matter whether we use the Rydberg constant or the proton charge radius for a consistency check, because the values of these parameters are strongly correlated through Eq. 1. This work favors the data from muonic hydrogen and a recently improved measurement of the 2S-2P Lamb shift in regular hydrogen (*8*). Further, our value is in good agreement with a recent measurement of the 2S-4P interval (*6*) but disagrees by 2.9σ with the hydrogen world data obtained by CODATA 2014 (*1*). Figure 1 summarizes the situation.

Two-photon direct frequency comb spectroscopy

An optical frequency comb is a regularly spaced array of laser frequencies $\omega_n = n\omega_r + \omega_0$ that is readily generated with a mode-locked laser (*14*). The pulse repetition rate, ω_r, is around $2\pi \times 315.2$ MHz, and the offset frequency, ω_0, is always smaller than ω_r. The integer n labels the ≈3000 laser lines (or modes) around $n \approx 1.16 \times 10^6$. During the measurements, the frequency of one of the laser modes and the exact repetition rate were constantly recorded with the help of a second, self-referenced frequency comb ("fs–frequency comb" in Fig. 2) (*15*) and a radio frequency counter, respectively. Both were referenced to a global positioning system (GPS)–disciplined hydrogen maser.

Our spectrometer is sketched in Fig. 2 and described in more detail in (*16*–*18*). Briefly, the carrier frequency of a mode-locked titanium: sapphire laser (78.8 MHz, 1.3 ps, 2.8 W) was doubled in an enhancement resonator with half the optical length of the laser resonator. This arrangement also doubled the repetition rate. After a second doubling stage that left the repetition rate constant, we obtained up to 50 mW of 205-nm radiation with a full width half maximum (FWHM) pulse duration of 2.0 ps (*18*). A delay line generated σ^+/σ^- circular-polarized double pulses that met twice per round trip at the center of the enhancement cavity housed in a vacuum chamber. This doubled the repetition rate once more to 315.2 MHz. Our general effort to produce a high repetition rate, both through the doubling stages and with the use of the delay line, led to a relatively short two-mirror enhancement cavity that supported two counterpropagating pulses. Our choice of polarization avoided loss of power in the delay line and reduced the excitation of the 1S-3D transitions relative to the 1S-3S transitions. The time-averaged circulating power in this cavity was ~60 mW per propagation direction with a beam radius at the focus of w_0 = 80 μm. Hydrogen atoms were produced by dissociation of hydrogen molecules in a radio frequency discharge tube, which were then guided through a Teflon tube to a cryogenic copper nozzle. In this nozzle, they thermalized and escaped into the vacuum through two adjacent holes. Some of the atoms made it into the cigar-shaped laser pulse collision volume (PCV) and were excited to the 3S state, from which they decayed within 0.16 μs to the 2P state, releasing a detectable 656-nm Balmer–α photon (Fig. 3).

To drive a two-photon transition with a frequency comb, the photon energies associated with the laser modes add up pairwise to produce the transition energy $h\omega_{eg}$ (*19*). By tuning ω_r such that a particular comb mode ω_n corresponds to half the transition frequency, all mode pairs m that satisfy the relation $\omega_{eg} = \omega_{n-m} + \omega_{n+m}$ contribute to the excitation rate. This is sketched on the left-hand side of Fig. 2. The two-photon resonance condition is also satisfied when the transition occurs exactly between two comb modes. Therefore, the spectroscopic signal repeats with the repetition rate when measured at the atomic frequency. All transitions accessible with the spectral bandwidth of the comb are observed to be convoluted within the range of ω_r.

An important feature of two-photon spectroscopy is that the first-order Doppler effect is suppressed when the atom absorbs a pair of counterpropagating photons. This can take place only within the PCV, where the excitation paths (shown in Fig. 2) add constructively. Outside the PCV, only Doppler-broadened absorption of two photons from either side takes place. The transition matrix elements for all four combinations of absorbed photon directions are the same but must be weighted by the number of atoms in the proper velocity class for Doppler-broadened absorption and by appropriately accounting for the polarization and the hyperfine structure (*20*).

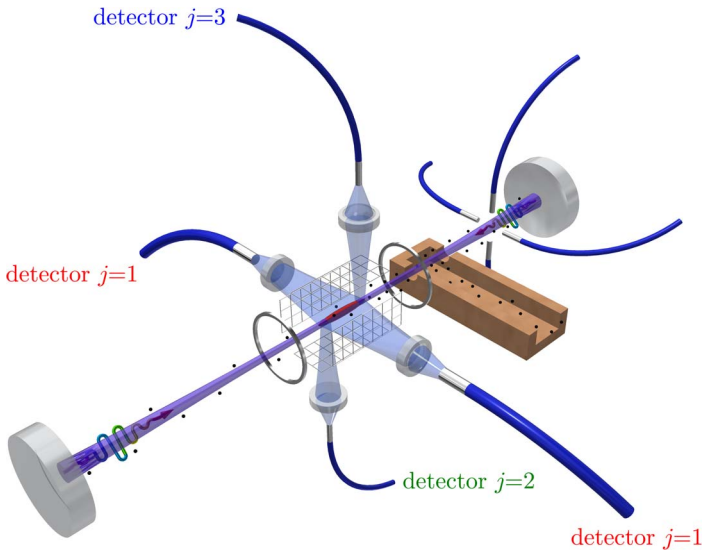


Fig. 3. Spectroscopy setup. The pulse collision volume (PCV; shown in red here and in Fig. 2) resembles an ellipsoid with semi-axes w_0 = 80 μm and $cT_{1/2}$ = 600 μm. It is surrounded by a Faraday cage made of a highly transmissive mesh and two end-cap electrodes. In the actual apparatus, the end caps cover a much larger surface around the PCV but are drawn as thin rings here for clarity. Using the quadratic DC-Stark shift, we can put tight limits on stray electric fields by applying voltages to the cage in all three directions and determining the minima of the resulting line shifts. Four lenses image the fluorescence from the whole PCV and its ends to multimode fibers (1 mm and 600 μm diameter) that guide the light through interference filters onto three independent single-photon counting modules: one main detector (j = 1) and two auxiliary detectors (j = 2, 3). With this arrangement, we can interpolate the chirp-induced residual first-order Doppler shift (CIFODS). At the other side of the nozzle, the Doppler-broadened signal is collected with four bare fibers of 1 mm core diameter that are in close proximity to the laser and atomic beam. As the Doppler-broadening is well in excess of the mode spacing, this signal is independent of the laser frequency and used for normalization. In this way, we remove substantial fluctuations of the laser power and the atom number flux. Because the Doppler-free and the Doppler-broadened signals scale in the same way with laser power, the normalized line amplitudes can be used as a measure of the atomic flux.

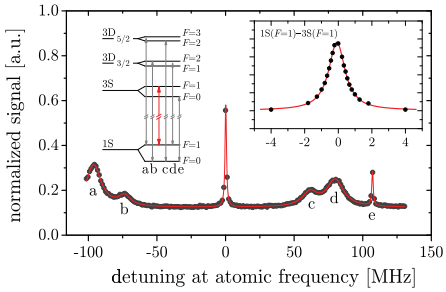


Fig. 4. Line components. Multiple 1S-3S/D two-photon transitions folded with the frequency comb so that the separation of the line components appears modulo $\omega_r = 2\pi \times 315.2$ MHz. By operating a well-chosen ω_r and with circular polarization, we obtain a well-isolated 1S(F=1)–3S(F=1) component. Other components that are allowed by the selection rules are: a, 1S(F=1)–3D$_{5/2}$; b, 1S(F=0)–3D$_{3/2}$(F=2); c, 1S(F=0)–3D$_{5/2}$(F=2); d, 1S(F=1)–3D$_{3/2}$; e, 1S(F=0)–3S(F=0). The hyperfine splitting of the 3D states is smaller than the natural line width and thus unresolved. The baseline offset is from the Doppler-broadened absorption of two copropagating photons. The inset shows an average of five line scans (3 min) within ±4 MHz of the main component, normalized to the Doppler-broadened signal, together with a Lorentzian fit. The sampling of the frequency axis was chosen to yield an approximately equidistant spacing on the signal axis to maximize the sensitivity to the center frequency per unit of measurement time. The line pulling by other 1S-3D lines and others (not visible) is negligible and discussed in detail in the SM. The detuning in this graph is measured at the atomic frequency, i.e., around the frequency given in Eq. 4. a.u., arbitrary units.

Two-photon direct comb spectroscopy is not fully free of the first-order Doppler effect. In the frequency domain, the contributing mode pairs do not have exactly the same frequencies and, thus, do not have exactly opposite Doppler shifts. However, for each pair m there is a pair $-m$ such that the Doppler shift is balanced provided that the spectral envelopes of the counterpropagating pulses are identical. This condition is best fulfilled within an enhancement resonator. As a result, the Doppler effect only broadens the line but does not shift it. In the time domain, this broadening is understood as time-of-flight broadening. Atoms with the most probable thermal velocity of v_0 = 340 m/s contribute the most to the signal at 7 K [see the supplementary materials (SM)]. When traveling along the laser beam axis, the time-of-flight broadening amounts to $\Delta\omega = 8\ln(2)v_0/cT_{1/2} = 2\pi \times 500$ kHz. By using a pulse duration (FWHM) of $T_{1/2}$ = 2 ps, we adapt the lengths of the PCV $cT_{1/2}$ such that we reach the natural line width of 1.0 MHz despite the 148 GHz wide spectral envelope of the comb (see SM).

The frequency comb in this application is, in many ways, equivalent to a continuous laser—the laser-limited line width is given by the width of a single comb mode rather than by the spectral envelope of the comb, while the line strength is given by the total power of all modes (*21*). Similarly, the AC-Stark shift derives from the time-averaged laser intensity rather than from the much larger peak intensity (*22*). Two-photon direct comb spectroscopy offers several advantages, such as access to shorter wavelengths through more efficient nonlinear processes. The small PCV allows for better control over several systematic effects and makes light collection simple and efficient. In addition, the small PCV allows for good shielding from stray electric and magnetic fields that would otherwise be challenging to compensate over a larger excitation volume. Moreover, the distance of the PCV from the nozzle can be varied to control the atomic density in a defined way to determine the pressure shift, keeping the inhomogeneities of the fields and pressure small. Although Doppler-free excitation takes place only within the PCV, Doppler-broadened fluorescence can be collected separately outside of the PCV. This signal is almost independent of the laser frequency and can be used to normalize fluctuations of laser power and atomic flux. By utilizing this feature, we could reach an almost shot-noise limited statistics (see SM). Figure 3 shows a close look at the PCV (red region) together with the shields and the detection optics that have been left out of Fig. 2. Figure 4 shows a typical line scan obtained with this spectrometer.

Data analysis and systematics

Our dataset consists of i = 1...4450 line scans like the one shown in the inset of Fig. 4, each of ≈36 sec duration. A large fraction of the data (2020 line scans) have been recorded with a nozzle temperature of 7 K. As a first step of the evaluation, we fit Lorentzians to the normalized signal to find the line widths, amplitudes, constant offsets, and the center frequencies $f_{j,i}$ for the three detectors (j = 1...3). The statistical uncertainties of the center frequencies $\sigma_{j,i}(f)$ are dominated by shot noise.

A small amount of excess noise is well explained by limitations of the normalization procedure and is taken into account as described in the SM. The lines recorded with the downstream detector ($j = 2$) at the lowest nozzle temperatures show the smallest time-of-flight broadening and reach the natural line width (see SM for details).

A systematic frequency shift that is intrinsic to the two-photon pulsed excitation of an atomic beam arises owing to a possible chirp of the laser frequency. To illustrate this effect, we consider Gaussian pulses counterpropagating along the $\pm z$ axis with a linear frequency chirp

$$E_{\pm}(t) = E_0 e^{-(1+ib)(t\pm z/c)^2/\tau^2 - i\omega_c t} \qquad (2)$$

Here, ω_c is the carrier frequency, b is the chirp parameter, and the FWHM pulse duration is given by $T_{1/2} = \tau\sqrt{2\ln(2)}$. The Doppler-free complex Rabi frequency of this transition is proportional to $E_+(t)E_-(t) = E_0^2 e^{-2(1+ib)(t^2+z^2/c^2)/\tau^2 - i2\omega_c t}$, where the fast oscillating carrier drops out within the usual rotating wave approximation. Any remaining time-dependent phase can translate into a frequency shift. This does not apply to the first phase term in the exponent, because in a repetitive pulse train, it merely leads to a periodic phase modulation without affecting the comb structure. The second phase term, however, acquires a time dependence with the changing position of the atom and leads to an instantaneous position-dependent frequency shift of $4bzv_z/c^2\tau^2$. With its dependence on velocity v_z, this frequency shift appears as a residual chirp-induced first-order Doppler shift (CIFODS) (*16*). In practice, the CIFODS is largely reduced because atoms are probed at positions of positive and negative z [in an isotropic gas sample also with both signs of v_z (*23*, *24*)]. However, as illustrated in Fig. 5, the divergence of the atomic beam leads to an imbalance of density and hence to an imbalance of the signals received from either ends of the PCV. Assuming the distance d of the PCV to the emerging point of the atomic beam is large compared to the dimensions of the PCV, the resulting CIFODS is estimated to be $\Delta\omega \approx -bv_0/d$. From the experimental data, we obtain a mean of $\langle b\rangle = -0.119$ with a substantially larger variation between the line scans. Data has been taken with $d = 19.1(1)$ mm and $d = 27.1(1)$ mm, so that the mean CIFODS is estimated to 0.34 kHz with $v_0 = 340$ m/s. A more detailed discussion is given in the SM and in (*16*).

Although the pulses from the laser had a negative chirp, the frequency-doubling stages introduced a positive chirp, mainly through self-phase modulation (*25*). Experimentally, we can vary the latter contribution by changing the crystal position along the focused laser beam. Unfortunately, standard methods such as frequency-resolved optical gating (FROG) cannot provide reliable values for small chirp parameters of picosecond pulses. In addition, the above arguments that lead to the CIFODS are based on a largely simplified model. However, under quite general assumptions, the CIFODS is an antisymmetric function of the position z relative to the center of the PCV. We use the position dependence of the CIFODS to measure and compensate for it. To this end, we collected light from different parts of the PCV (Fig. 3). The main detector $j = 1$ showed a small sensitivity to CIFODS, because it averaged the signal over the entire PCV. In contrast, the auxiliary detectors ($j = 2, 3$) that pointed at the ends of the PCV were maximally sensitive to CIFODS with opposite signs and hence provide a sensitive probe. CIFODS can be shown to be linear in the chirp parameter for any chirp function, as long as it is given by a single parameter (see SM). Therefore we can write $f_{j,i} = f_0 + \alpha_j b_i$, where $f_{j,i}$ is the measured frequency of the i-th scan with chirp parameter b_i by the j-th detector. This frequency is shifted from the unaffected transition frequency f_0 by an amount that is proportional to the chirp parameter b_i, which is unknown but assumed to be constant during a line scan. The CIFODS sensitivities α_j are geometry dependent and therefore insufficiently known, but fixed. In principle, the α_j can be determined through simulations, which include the geometry of the detectors and the beam divergence. The expressions for the three detectors may be combined to $f_{1,i} = f_0 + \kappa_{DS}(f_{2,i} - f_{3,i})$, where $\kappa_{DS} = \alpha_1/(\alpha_2 - \alpha_3)$, $|\alpha_1| \ll |\alpha_2|$, and $\alpha_2 \approx -\alpha_3$. The unperturbed frequency f_0 can then be obtained by making a linear fit of the measured frequencies of the main detector, $f_{1,i}$, versus the difference frequency obtained with the two auxiliary detectors, $f_{2,i} - f_{3,i}$. The determination of the chirp parameters is not required, nor do we need to know the geometry, i.e., α_j. Because our data cover both signs of b, we could interpolate to the proper transition frequency. Note that, for a similar dataset, interpolation is more accurate and reliable than extrapolation.

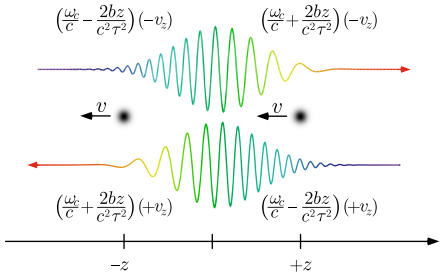


Fig. 5. Principle of the chirp-induced residual first-order Doppler shift. Two chirped counterpropagating pulses with carrier frequency ω_c (sketched spatially separated for clarity) meet within the pulse collision volume, where they drive a two-photon transition in an atom that moves parallel to the z axis with a velocity v_z. The instantaneous frequency shifts at the edges of the pulses are $\pm 2bv_z/c^2\tau^2$ at positions $\pm z$, with the chirp parameter b and with the value of the shift doubling for a two-photon transition. For the Doppler shift, the velocity has to be measured relative to the pulse propagation direction. The shifts cancel for atoms that travel exactly parallel to the z axis. In a divergent atomic beam, however, this cancellation is no longer perfect, and the CIFODS might be estimated by integrating over z with a weight given by the square of the magnitude of the Rabi frequency $\propto \exp(-4(z-d)^2/c^2\tau^2)$ with distance d between the PCV and the emerging point of the atomic beam. This results in $\Delta\omega \approx -bv/d$.

Similarly to the CIFODS interpolation, we extrapolate the AC-Stark shift by adding a term $\kappa_{AC}P_i$, where P_i are the scan-averaged cavity-transmitted laser powers (see "PM" in Fig. 2). Nonlinear effects due to saturation and ionization were negligible (<1 Hz) at the power levels used. Under average experimental conditions, the AC-Stark effect shifts the resonance by 4.6 kHz.

An additional systematic shift is the pressure shift, which is due to collisions within the atomic beam and with the residual background gas. Because the latter was estimated to be negligible (see SM), we determined the effects of the intrabeam collisions by taking advantage of the small excitation volume. By changing the PCV-to-nozzle distance, we can vary the density of atoms and molecules by about a factor of 2. The properly normalized and scaled line amplitudes (see SM for details), A_i, were used to measure the local density of atoms independently of the other fluctuations such as laser power variations. We introduce another term, $\kappa_{PS}A_i$, to extrapolate the pressure shift.

The final and somewhat larger systematic shift is due to the second-order Doppler (SOD) effect, which is given by $-(v/2c)^2$ in relative units for an atom at a velocity v. We estimate this shift to be −2.9 kHz at a nozzle temperature of $T = 7$ K by computing the average squared velocity of $\bar{v}^2 = (416\ \mathrm{m/s})^2$ for atoms with the largest contribution to the signal (see SM). Because $\bar{v}^2 \sim T$, the SOD is expected to be linear with the nozzle temperature. For a thermal atomic beam at room temperature, the SOD shift would be about −79 kHz in our geometry. The motional Stark effect can be used to compensate for this rather large shift (*26*). In our case, we instead took data at a wide range of temperatures and introduced another parameter, κ_{SOD}, that modeled the SOD shift to be linear in temperature.

In total, we adjusted

$$f_{1,i} = f_0 + \kappa_{DS}(f_{2,i} - f_{3,i}) + \kappa_{SOD}T_i + \kappa_{AC}P_i + \kappa_{PS}A_i \qquad (3)$$

to the experimental data ($f_{j,i}$, T_i, P_i, and A_i) weighted by their uncertainties and correlations, by adjusting the parameters κ_{DS}, κ_{SOD}, κ_{AC}, κ_{PS}, and f_0. As discussed earlier, we believe that the CIFODS compensation is very robust. On the basis of simple perturbation theory, the AC-Stark shift is expected to be linear. The pressure shift was less well controlled, but it is the smallest of the shifts compensated by fitting Eq. 3. The effective SOD shift, even at low temperatures, required the second largest correction. In addition, linearity with temperature may be compromised by the complex velocity-dependent excitation probability and expected deviations from the Maxwellian velocity distribution. To make sure that we fully understand this effect to the required level, we took data at seven different nozzle temperatures. Because a larger SOD requires a more precise model, we decided to use only the $T \leq 30$ K data in Eq. 3. As shown in Fig. 6, this fit correctly predicts the data at higher temperatures. Therefore we are confident that the linearity assumption holds for the lower temperatures but take possible deviations into account by adding temperature-dependent model uncertainties to the uncertainties of the measured frequencies $f_{1,i}$ (see SM for details). Further corrections applied to the transition frequency, f_0, were the DC-Stark shift, the Zeeman shift, quantum interference line pulling (*27*, *28*), line pulling through other fine- and hyperfine components, and the maser calibration (the details are given in the SM).

Rydberg constant and proton charge radius

In all steps of the evaluation, the uncertainties were propagated, and correlations were taken into account. All corrections and the error budget are given in Table 1. After applying them, we obtain the unperturbed frequency of the 1S(F=1)–3S(F=1) transition

$$f_{1S\text{-}3S}(F=1) = 2{,}922{,}742{,}936{,}716.72(72)\ \text{kHz} \quad (4)$$

Subtracting the hyperfine shifts of –341,949,069.6(8) Hz (*29*) gives the hyperfine centroid

$$f_{1S\text{-}3S}(\text{centroid}) = 2{,}922{,}743{,}278{,}665.79(72)\ \text{kHz} \quad (5)$$

This measurement is in good agreement but more than 20 times more accurate than our previous room temperature measurement (*16*) and 3.6 times more accurate than a recent 1S-3S room-temperature measurement with a continuous-wave laser (*7*). However, it deviates by 2.1 combined standard deviations from the latter. Although the combined standard deviation of our result from the muonic value is comparable, the disagreement between measurements of the same transition frequency obtained with mutually distinct leading systematic effects is of a different kind. It does not disprove QED but points to some further, yet undiscovered, systematic effects in either (or both) of these measurements.

We use our result, the measured 1S-2S transition frequency (*11*), and QED in form of Eq. 1 to improve the value for the Rydberg constant

$$R_\infty = 10{,}973{,}731.568226(38)\ \text{m}^{-1} \quad (6)$$

and obtain an independent value for the RMS proton charge radius

$$r_p = 0.8482(38)\ \text{fm} \quad (7)$$

In doing so, we have used the collection of terms provided in (*1*) as well as the values for the fine structure constant and the electron-to-proton mass ratio given there. More accurate values of these constants are available online (*30*), but that has no influence on the results presented here. An update of the theory (*31*) changes this result only within a small fraction of the error bar. The contribution to the error bar by the theory and other constants to R_∞ and r_p is 2.2×10^{-6} m^{-1} and 0.00097 fm, respectively. Our value for the Rydberg constant is more accurate than the most recent published CODATA adjustment (*1*) but differs by 3.7 combined standard deviations from it. It is, however, in agreement with the latest CODATA value that is so far available online only (*30*), without details on how this value is obtained. Our proton charge radius disagrees by 2.9 combined standard deviations from the hydrogen world data as of 2014 but is in better agreement with the most recent muonic value (*32*), a recent measurement of the Lamb shift (*8*) and the 2S-4P transition frequency (*6*) in atomic hydrogen, as well as a recent result from elastic electron-proton scattering at very low momentum transfer (*10*).

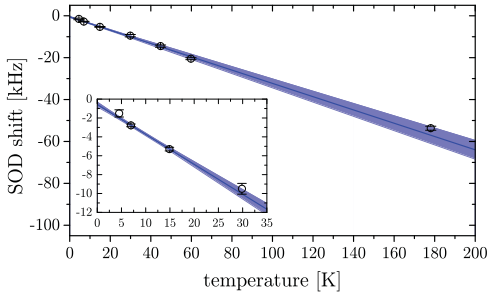


Fig. 6. Second-order Doppler. Second-order Doppler shift of the experimental data (black circles) as a function of the nozzle temperature, T, after correcting the shifts included in Eq. 3—i.e., CIFODS (κ_{DS}), AC-Stark (κ_{AC}), and pressure shift (κ_{PS}). Additional small corrections of 0.57 kHz are listed in Table 1. The error bars are obtained by assuming detection shot noise and model uncertainties as described in the text with the usual error propagation. The blue region is a linear fit together with its 1σ confidence band using the $T \leq 30$ K data only. The inset shows the low-temperature data used for evaluation in greater detail. The observed linearity justifies the use of Eq. 3. More details are given in the SM.

Table 1. Error budget of the 1S(F=1)–3S(F=1) measurement. All values are given in kilohertz. The average effect is the weighted mean of all evaluated data and quantifies the applied corrections. The multiparameter CIFODS (MP CIFODS) is an estimation of the uncertainty that may result from several independent sources of the laser chirp. This error budget is discussed in detail in the SM.

Contribution	Average effect	Correction	Uncertainty
Statistics	—	—	0.11
CIFODS	+0.79	—	0.08
SOD	–3.20	—	0.26
AC-Stark	+4.60	—	0.30
Pressure shift	+0.93	—	0.30
Residual Doppler	—	—	0.48
DC-Stark	+0.031	–0.031	0.015
Zeeman shift	–0.002	+0.002	0.002
Line pulling	–0.30	+0.30	0.050
MP CIFODS	—	—	0.10
Maser	–0.30	+0.30	0.030
Total		**+0.57**	**0.72**

REFERENCES AND NOTES

1. P. J. Mohr, D. B. Newell, B. N. Taylor, *Rev. Mod. Phys.* **88**, 035009 (2016).
2. R. H. Parker, C. Yu, W. Zhong, B. Estey, H. Müller, *Science* **360**, 191–195 (2018).
3. F. Heiße *et al.*, *Phys. Rev. Lett.* **119**, 033001 (2017).
4. Th. Udem, *Nat. Phys.* **14**, 632 (2018).
5. R. Pohl *et al.*, *Nature* **466**, 213–216 (2010).
6. A. Beyer *et al.*, *Science* **358**, 79–85 (2017).
7. H. Fleurbaey *et al.*, *Phys. Rev. Lett.* **120**, 183001 (2018).
8. N. Bezginov *et al.*, *Science* **365**, 1007–1012 (2019).

9. I. T. Lorenz, H. W. Hammer, U. G. Meißner, *Eur. Phys. J. A* **48**, 151 (2012).
10. W. Xiong *et al*., *Nature* **575**, 147–150 (2019).
11. A. Matveev *et al*., *Phys. Rev. Lett.* **110**, 230801 (2013).
12. G. Hagel, F. Nez, F. Biraben, *Appl. Opt.* **41**, 7702–7706 (2002).
13. M. Herrmann *et al*., *Phys. Rev. A* **79**, 052505 (2009).
14. M. C. Stowe *et al*., *Adv. At. Mol. Opt. Phys.* **55**, 1–60 (2008).
15. T. Udem, R. Holzwarth, T. W. Hänsch, *Nature* **416**, 233–237 (2002).
16. D. C. Yost *et al*., *Phys. Rev. A* **93**, 042509 (2016).
17. E. Peters, D. C. Yost, A. Matveev, T. W. Hänsch, Th. Udem, *Ann. Phys.* **525**, L29–L34 (2013).
18. E. Peters *et al*., *Opt. Express* **17**, 9183–9190 (2009).
19. J. N. Eckstein, A. I. Ferguson, T. W. Hänsch, *Phys. Rev. Lett.* **40**, 847–850 (1978).
20. G. Grynberg, F. Biraben, E. Giacobino, B. Cagnac, *J. Phys.* **38**, 629–640 (1977).
21. Ye. V. Baklanov, V. P. Chebotayev, *Appl. Phys.* **12**, 97–99 (1977).
22. P. Fendel, S. D. Bergeson, T. Udem, T. W. Hänsch, *Opt. Lett.* **32**, 701–703 (2007).
23. S. Reinhardt, E. Peters, T. W. Hänsch, Th. Udem, *Phys. Rev. A* **81**, 033427 (2010).
24. A. Ozawa, Y. Kobayashi, *Phys. Rev. A* **86**, 022514 (2012).
25. H. Wang, A. M. Weiner, *IEEE J. Quantum Electron.* **39**, 1600–1618 (2003).
26. G. Hagel, R. Battesti, F. Nez, L. Julien, F. Biraben, *Phys. Rev. Lett.* **89**, 203001 (2002).
27. D. C. Yost *et al*., *Phys. Rev. A* **90**, 012512 (2014).
28. H. Fleurbaey, F. Biraben, L. Julien, J. P. Karr, F. Nez, *Phys. Rev. A* **95**, 052503 (2017).
29. U. D. Jentschura, V. A. Yerokhin, *Phys. Rev. A* **73**, 062503 (2006).
30. The NIST Reference on Constants, Units, and Uncertainty, CODATA Internationally recommended 2018 values of the Fundamental Physical Constants; https://physics.nist.gov/cuu/Constants/index.html.
31. V. A. Yerokhin, K. Pachucki, V. Patkóš, *Ann. Phys.* **531**, 1800324 (2019).
32. A. Antognini *et al*., *Science* **339**, 417–420 (2013).
33. R. W. P. Drever *et al*., *Appl. Phys. B* **31**, 97–105 (1983).
34. A. Grinin, Data for the 1S-3S measurement, Version 1, Zenodo (2020); https://doi.org/10.5281/zenodo.4046497.

ACKNOWLEDGMENTS

We thank J. Weitenberg, S. Karshenboim, K. Pachucki, and V. Yerokhin for helpful discussions. **Funding:** Support was provided by the Deutsche Forschungsgemeinschaft (German Research Foundation DFG) under Germany's Excellence Strategy, project EXC-2111-390814868, and the DFG Cluster of Excellence PRISMA+ (EXC 2118/1 project 39083149). **Author contributions:** T.U. planned the experiment, and T.U., R.P., and T.W.H. supervised the experiment. A.G. and A.M. obtained the data. A.G., A.M., D.C.Y., L.M., and V.W. worked on setting up the experiment. A.M., A.G., and T.U. did the modeling. A.G., A.M., and T.U. carried out the data analysis. All authors contributed to the analysis of the systematic uncertainties and to writing the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. The raw data and description of the data are available at Zenodo (*34*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1061/suppl/DC1
Supplementary Text
Figs. S1 to S34
Tables S1 to S3
References (*35*–*63*)

14 May 2020; accepted 14 October 2020
10.1126/science.abc7776

FOREST ECOLOGY

Increased growing-season productivity drives earlier autumn leaf senescence in temperate trees

Deborah Zani[1], Thomas W. Crowther[1], Lidong Mo[1], Susanne S. Renner[2], Constantin M. Zohner[1]*

Changes in the growing-season lengths of temperate trees greatly affect biotic interactions and global carbon balance. Yet future growing-season trajectories remain highly uncertain because the environmental drivers of autumn leaf senescence are poorly understood. Using experiments and long-term observations, we show that increases in spring and summer productivity due to elevated carbon dioxide, temperature, or light levels drive earlier senescence. Accounting for this effect improved the accuracy of senescence predictions by 27 to 42% and reversed future predictions from a previously expected 2- to 3-week delay over the rest of the century to an advance of 3 to 6 days. These findings demonstrate the critical role of sink limitation in governing the end of seasonal activity and reveal important constraints on future growing-season lengths and carbon uptake of trees.

The phenological cycles of trees exert a strong control on biological interactions (*1*, *2*), the global carbon cycle, surface albedo, and the climate system (*3*). Warming trends over recent decades have led to extended growing seasons in temperate forests (*4*). As a result, primary productivity of temperate forests is expected to increase, with each day advance in spring leaf-out translating to an increase in net carbon uptake of 45 kg carbon per hectare forest and each day delay in autumn senescence translating to an increase of 98 kg ha^{-1} (*5*). Accurate projections of growing-season duration are, therefore, essential to forecast future changes in forest carbon balance (*6*). Yet, whereas spring leaf emergence has received much scientific attention (*7*–*9*), autumn phenology is relatively understudied (*10*), and existing models exhibit high uncertainty due to our limited understanding of the underlying environmental triggers (*11*). Understanding the physiological mechanisms governing the timing of autumn leaf senescence is critical if we are going to constrain the uncertainty in future projections of temperate tree phenology and productivity.

It has traditionally been accepted that autumn temperature and day length are the main determinants of autumn phenology, leading to the assumption that warming temperatures will delay autumn leaf senescence in the future (*12*). However, a growing body of evidence suggests that autumn delays will be counteracted by lagged effects of changes in spring and summer temperatures (*11*, *13*). For instance, earlier leaf unfolding in spring has been found to advance autumn leaf senescence (*14*–*16*). A likely explanation is sink limitation of photosynthesis, mediated through interactions between photosynthate supply, phytohormones, and nutrient supply (*17*–*19*). Increased photosynthetic activity causes plants to progress through their seasonal cycle more rapidly, ultimately resulting in earlier leaf senescence (*17*). This role of photosynthesis in governing the timing of leaf senescence through source-sink feedbacks has long been established in herbaceous plants (*20*, *21*) but has never been tested in trees. Alternatively, direct constraints on leaf life span (*22*), increased drought stress (*23*), and extended herbivory (*24*) might explain lagged effects of spring leaf-out and summer climate on tree autumn phenology. Because the underlying mechanisms remain unclear, phenological lagged effects are not, or only partially, accounted for in current autumn phenology models (*11*, *12*). Testing the relative importance and magnitude of these effects and evaluating the implications for future autumn trajectories require quantitative, empirical evidence about these effects from a combination of controlled experiments and long-term in situ observations on mature trees exposed to real-world environmental changes.

In this work, we used climate-manipulation experiments and direct leaf phenology observations on dominant Central European tree species to test the degree to which the timing of autumn senescence is determined by preceding seasonal productivity. These effects were evaluated relative to the individual effects of spring leaf-out, water availability, vapor pressure deficit, CO_2 concentration ($[CO_2]$), and summer and autumn temperatures. We hypothesized that, if growing-season duration is constrained by trees' carbon-sink capacity (*17*), then increased source strength (carbon uptake in spring and summer) should drive earlier leaf senescence in a warming, CO_2-enriched world (*18*). We further expected that spring leaf-out, $[CO_2]$, summer temperature, and water availability will indirectly affect the timing of leaf senescence by modulating growing-season carbon uptake.

Our in situ analysis is underpinned by a database of phenological observations of spring leaf-out and autumn leaf-senescence dates collected from the Pan European Phenology Project (*25*). After initial filtering, we obtained 434,226 phenological observations of six deciduous tree species at 3855 sites across Central Europe, resulting in a total of 14,626 individual time series (lasting 15 to 65 years) during the period 1948 to 2015 (table S1) (*26*). We then tested the relative effects of six parameters, i.e., leaf-out time, photosynthesis, vapor pressure deficit, summer and autumn temperature, and precipitation, on autumn senescence dates. To further support the results obtained from the observational data, we performed a set of controlled experiments, designed to modify plant carbon uptake. Finally, we developed an autumn phenology model that explicitly accounts for both seasonal carbon uptake and autumn environmental cues by incorporating mechanistic representations of these drivers. To represent seasonal productivity in our analyses, we used direct photosynthesis measurements for our experiments and a parameterized photosynthesis model commonly used in dynamic global vegetation models (*27*) as well as a simpler growing season index (*28*) for the in situ analysis (*26*). We compared model performance with existing autumn phenology models and used climate and spring leaf-out forecasts to project future changes in autumn senescence under a business-as-usual climate scenario. Ultimately, this combination of long-term observations, experiments, and modeling approaches enables us to evaluate the mechanisms governing the timing of autumn leaf senescence and to improve our confidence in future forecasts of leaf phenology, carbon cycling, and climate.

Analysis of long-term observations

The long-term observations showed that across all study species, years with elevated spring-summer productivity [estimated from a parameterized photosynthesis model (*27*) including information on atmospheric CO_2 levels (*29*), temperature, irradiance, and soil moisture] were associated with earlier autumn senescence (mean time series-level $R^2 \pm SD = 0.34 \pm 0.18$ Fig. 1A); after controlling for autumn temperature, each 10% increase in seasonal photosynthetic activity (relative to the mean time seasonal activity of each time series) translated to senescence 8 days (8.1 ± 3.9) earlier, on average. Across all observations [see Fig. S1 for univariate correlations among all potential drivers of leaf senescence (*30*, *31*)] and within time series (Fig. 1 and Fig. S2), growing-season photosynthesis had the greatest effect on autumn senescence dates, relative to other physiological or environmental cues. Specifically, whereas growing-season photosynthesis explained, on average, 34% of the

[1]Institute of Integrative Biology, ETH Zurich (Swiss Federal Institute of Technology), Universitätsstrasse 16, 8092 Zurich, Switzerland. [2]Systematic Botany and Mycology, University of Munich (LMU), Menzinger Str. 67, 80638 Munich, Germany.
*Corresponding author. Email: constantin.zohner@gmail.com

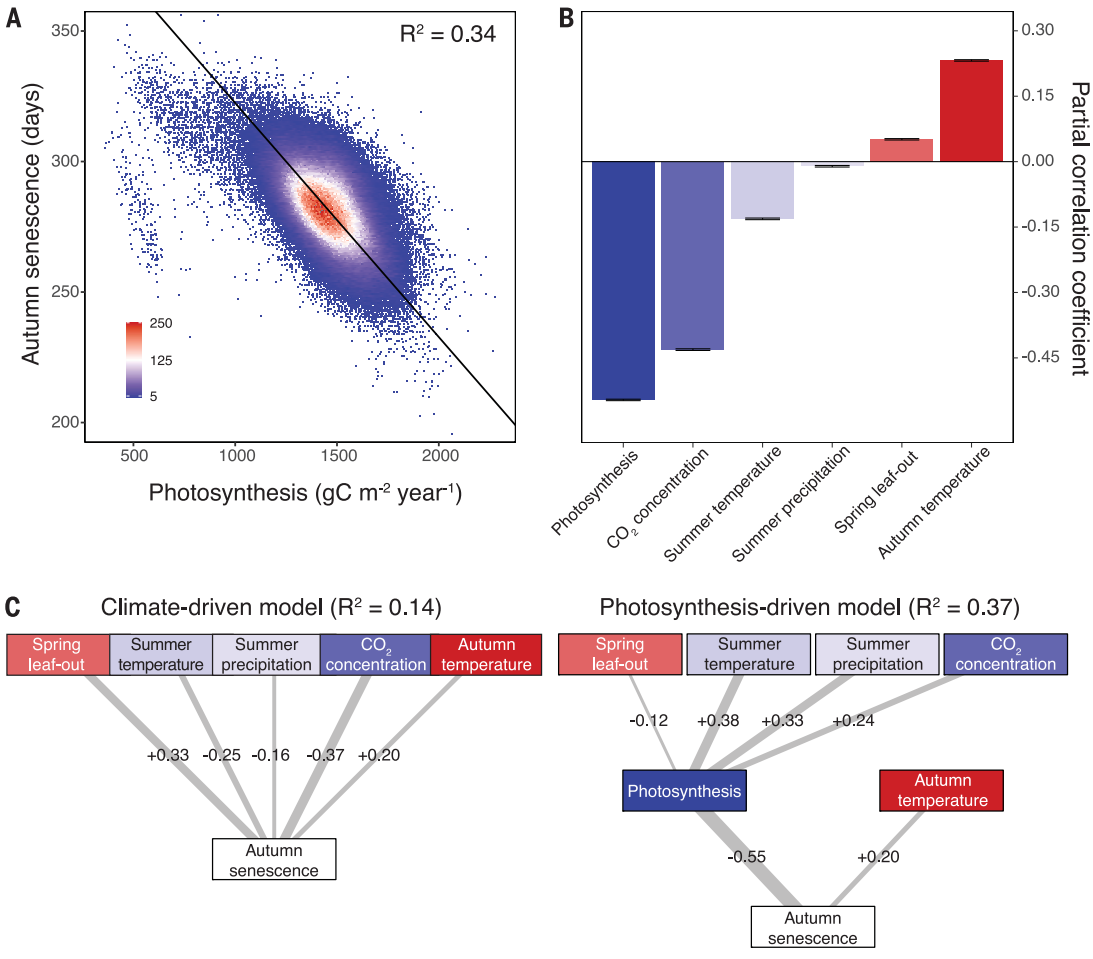


Fig. 1. The effects of seasonal photosynthesis, autumn and summer temperatures, summer precipitation, growing-season vapor pressure deficit (VPD_{GS}), and spring phenology on autumn senescence dates. (**A**) Univariate effect of seasonal photosynthesis (cA_{tot}) on the timing of autumn leaf senescence across 14,626 analyzed time series in six species. To visualize the average trend across time series, mixed-effects models were applied, including each time series (species-site combination) as a random effect. The R^2 represents the mean coefficient of determination across all time series. (**B**) Mean partial correlation coefficients (±2 standard errors) between each parameter and senescence dates across all time series from multiple linear regression. (**C**) Causality networks for the association among environmental parameters and autumn phenology as inferred from structural equation models (SEMs) excluding (climate-driven model) or including (photosynthesis-driven model) information on seasonal photosynthesis. R^2 values represent the adjusted coefficients of variation correcting for the number of predictors. Standardized correlation coefficients between variables are shown next to lines, blue boxes indicate a negative net effect of the variable on autumn senescence, and red boxes indicate a positive effect. See table S2 for details on SEM analysis and section 1.5 in (*26*) for detailed variable description. To standardize among time series, variables were computed as anomalies within each time series.

inter-annual variation in senescence dates within time series, autumn temperatures, summer temperatures, growing-season vapor pressure deficit, and precipitation individually explained 10, 7, 7, and 4% of the variation, respectively. To explore the mechanistic links between growing-season photosynthesis, spring leaf-out, climate drivers, and autumn senescence dates, we applied a structural equation model, including or excluding growing-season photosynthesis as a predictor (Fig. 1C and table S2). These analyses show the importance of photosynthesis by revealing that a combination of environmental variables alone cannot explain observed autumn senescence dates (R^2 including or excluding growing-season photosynthesis = 0.55 and 0.13, respectively). This model suggests that spring phenology, growing-season vapor pressure deficit, summer temperature, and precipitation affect the timing of leaf senescence indirectly, by modifying seasonal photosynthesis, whereas autumn temperature directly affects autumn phenology.

Experimental tests of the productivity–autumn phenology relationship

To isolate the mechanisms driving the correlations between growing-season productivity and autumn phenology, we ran a series of manipulative experiments. First, we shaded whole plants (shade treatment) or only half of a plant (half-shade treatment) of *Fagus sylvatica* and *Spiraea japonica* during the growing season and compared them, respectively, with Sun-exposed plants or the Sun-exposed halves (see materials and methods, Fig. 2A, and fig. S3 for detailed setup). Although measured leaf-level growing-season photosynthesis was strongly reduced in both shaded treatments (fig. S4A), plant-level photosynthesis was constant between the Sun-exposed and shaded halves of plants (Fig. 2B). In agreement with the sink-limitation hypothesis, senescence in fully shaded plants was delayed by, on average, 13 ± 7 and 8 ± 5 days (mean ± 95% CIs) relative to Sun-exposed individuals for *Fagus* and *Spiraea*, respectively, whereas no significant difference in senescence dates between the shaded and Sun-exposed halves of plants could be observed (Fig. 2C). This suggests that leaf-level photosynthesis (source activity) does not directly affect the timing of leaf senescence, and, instead, autumn senescence appears to be triggered by a systemic response mediated by the interaction between plant-level sink strength and photosynthetic activity.

In a second experiment, we tested the effects of CO_2 fertilization and increased temperatures on autumn senescence dates by studying variations in spring-to-summer photosynthesis and autumn senescence dates of *Betula pendula* between three controlled climate chamber treatments (elevated temperature, elevated [CO_2], both elevated) and an ambient control. The results match our predictions based on plant sink limitation, showing that an increase in measured spring-to-summer photosynthesis of 5, 35, and 42% under elevated temperatures, elevated [CO_2], or elevated temperatures and

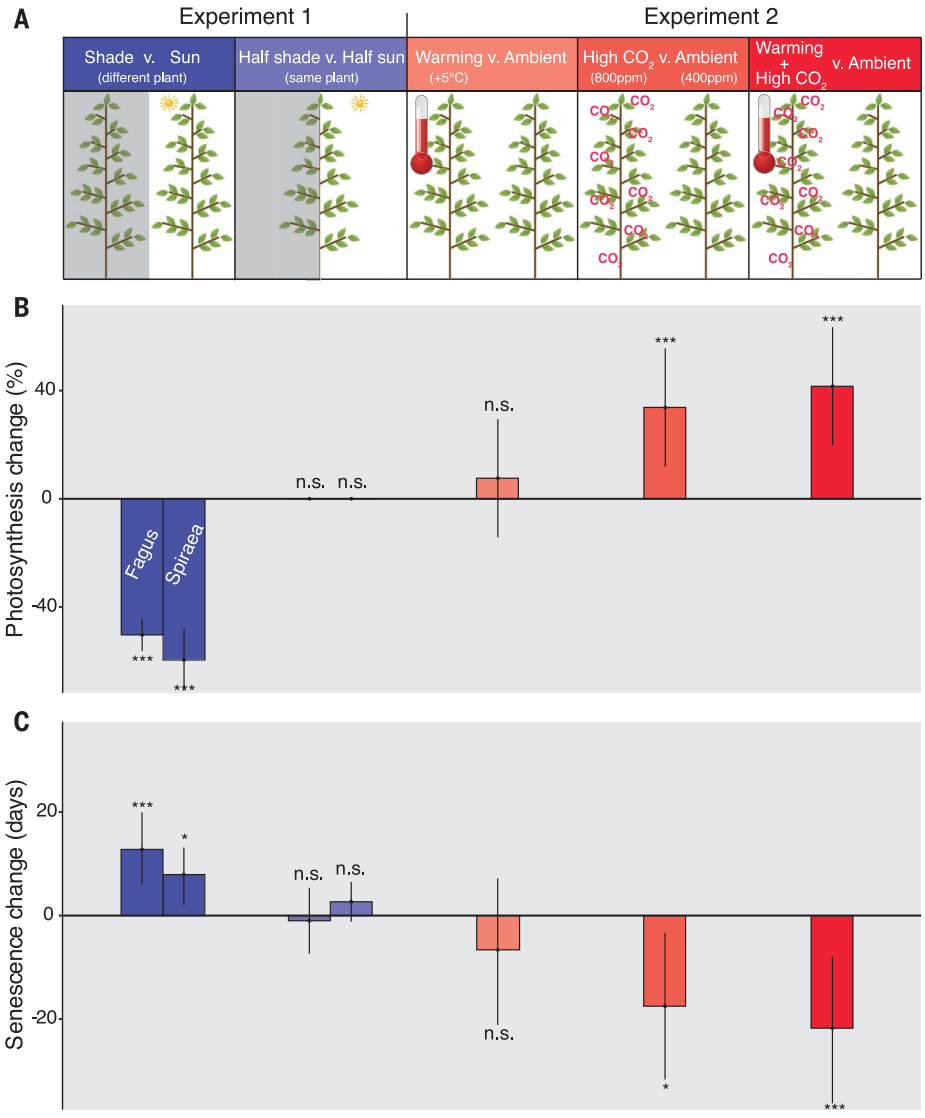
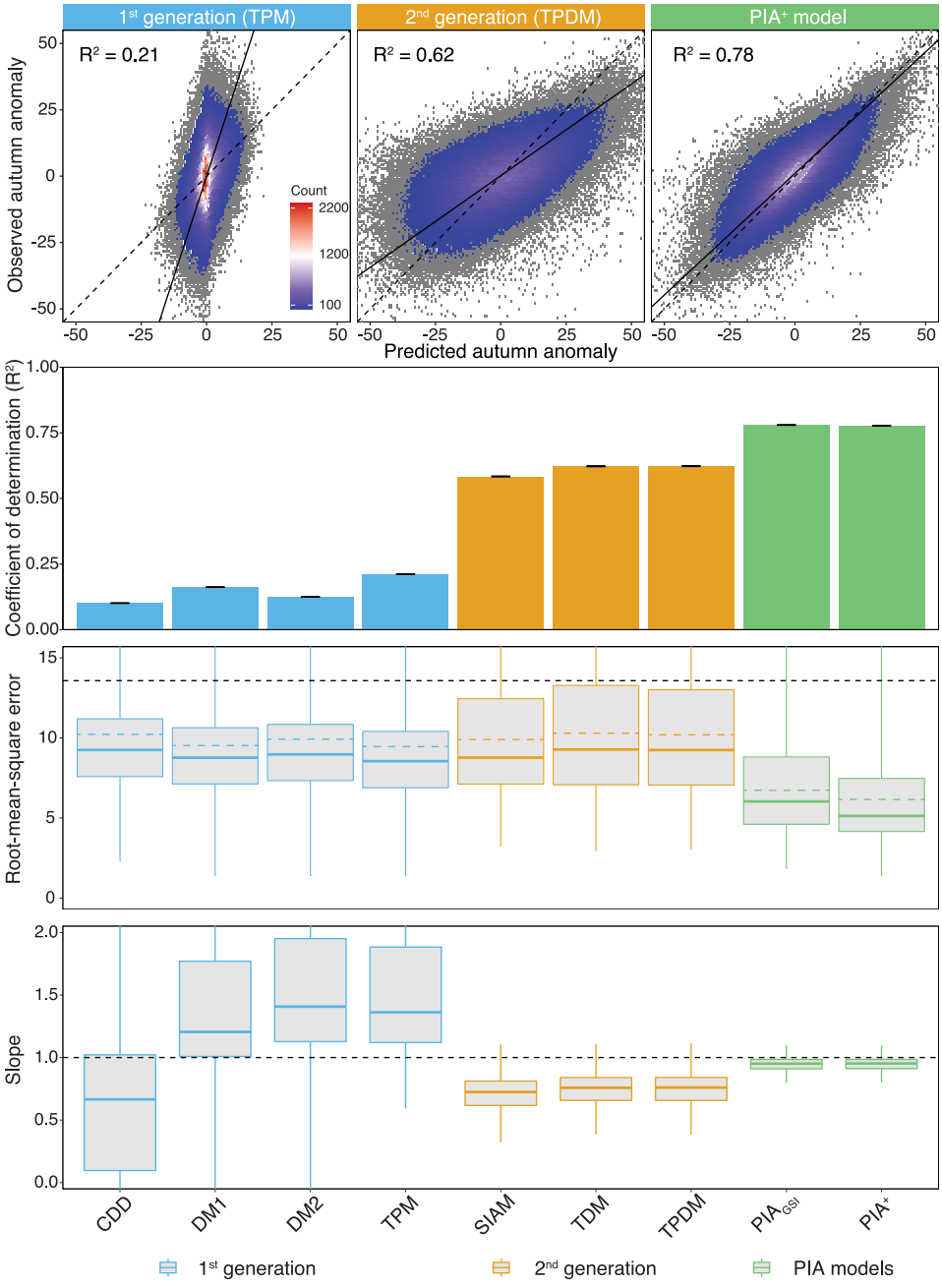


Fig. 2. Experimental tests of the effects of plant-level photosynthesis on autumn senescence dates. (**A**) Experimental setup. In experiment 1, we shaded whole plants (shade treatment) or only half of each plant (half-shade treatment) of *F. sylvatica* and *S. japonica* during the 2017 growing season [left and right blue bars in (B) and (C), respectively]. These treatments were compared with unshaded plants for the shade treatment or to the unshaded halves for the half-shade treatment. In experiment 2, *B. pendula* individuals kept in climate chambers were exposed to elevated temperatures (+5 °C), elevated [CO_2] [800 parts per million (ppm)], or a combination of elevated temperatures and [CO_2] and compared with an ambient control (400 ppm [CO_2]). v., versus. (**B**) Relative changes in plant-level spring and summer photosynthesis compared with the controls (mean ± 95% CIs). Whereas leaf-level photosynthesis of the shaded halves of plants was reduced relative to the Sun-exposed halves (see fig. S4A), there was no difference in plant-level photosynthesis between the shaded and Sun-exposed halves of plants. Note that photosynthesis reflects electron transport rates in experiment 1 and net daytime CO_2 uptake in experiment 2 (*26*). (**C**) Changes (in days) in autumn senescence dates relative to the controls. n.s., nonsignificant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

[CO_2], respectively, translated to leaf senescence 7, 17, and 22 days earlier (Fig. 2 and fig. S5). Ultimately, across all experimental manipulations, treatments that enhanced growing-season photosynthesis caused proportionally similar advances in autumn senescence, whereas treatments that restricted growing-season photosynthesis delayed leaf senescence. Both experiments lend support for the effects of growing-season productivity and sink limitation on the timing of autumn senescence dates in temperate woody plants.

The photosynthesis-influenced autumn phenology model

To represent the observed effects of growing-season productivity on the timing of autumn senescence in phenological model forecasts, we used the in situ observational data (*25*) to develop a set of photosynthesis-influenced autumn phenology (PIA) models and tested their performance against existing models from the literature (see table S3 for a summary of autumn phenology models). To account for the linear effect of seasonal photosynthesis on autumn senescence dates (Fig. 1A), our PIA models predict the critical cold-degree-day threshold in autumn as a linear function of seasonal photosynthesis (*26*) (Eqs. 53 to 55), assuming that increased growing-season productivity decreases the amount of autumn cooling required to induce leaf senescence. Previous models solely account for autumn temperature and day length (*12*) (first-generation models) and can additionally represent the potential influence of changing spring leaf-out (*15*), growing-season temperature, and precipitation (*11*) (second-generation models). Across all species, we found that our PIA models had higher power to predict observed leaf-senescence anomalies than did first- and second-generation models (Fig. 3 and table S4). Across the entire study period, the coefficient of determination (R^2 ± 2 standard errors) for the PIA models was 0.78 ± 0.01, relative to 0.62 ± 0.01 and 0.21 ± 0.01 for the best-performing second-generation and first-generation model, respectively (see Fig. 3, A and B, and fig. S6 for a comparison of model fit across time). Model error in the PIA models was reduced by more than one-third compared with previous models [mean root mean square error (RMSE) ± 2 standard errors = 10.87 ± 0.05 and 6.41 ± 0.03 for the best-performing second-generation (temperature- and precipitation-influenced Delpierre model, [TPDM]) and PIA model, respectively; Fig. 3C]. The slope components of observed versus predicted autumn anomalies were close to 1 for the PIA models, indicating that the models are unbiased, whereas the slopes strongly differed from 1 for the previous models (Fig. 3D). The best-performing PIA model was our photosynthesis model accounting for water stress (PIA^+), and the simpler growing season–index model (PIA_{GSI}) performed only marginally worse (Fig. 3). K-fold cross-validation at the time series level confirmed the stability of our predictions (Fig. 3C and table S4). The high



A

B

C

D

Fig. 3. Model comparison and validation of the PIA model. (**A**) Observed versus predicted autumn senescence dates of the PIA models and the best-performing first-generation and second-generation models. Solid lines show linear regression fit; dashed lines show the 1:1 line. To standardize among time series, observed and predicted senescence dates are shown as anomalies, i.e., as deviation from the mean observed senescence date of each time series. The R^2 values represent the mean coefficient of determination across all time series. (**B** to **D**) Comparison of the nine models tested in this study. (B) Mean coefficient of determination (R^2) ± 2 standard errors of observed versus predicted senescence dates across the 14,626 time series. (C) RMSEs within each time series. The colored dashed lines show the median RMSEs from fivefold cross-validation within time series; the black dashed line shows the median RMSE expected under a null model in which senescence dates do not differ among years. (D) Slope estimates of observed versus predicted senescence dates. CDD, Cold-degree-day model; DM, Delpierre model; SIAM, Spring-influenced autumn phenology model; TDM, Temperature-influenced Delpierre model; TPM, Low Temperature and Photoperiod Multiplicative model; TPDM, Temperature- and Precipitation-influenced Delpierre model.

predictive power and accuracy of our autumn phenology models provide further evidence for the important role of spring and summer plant activity in regulating the timing of autumn phenology.

Future projections of autumn phenology

Accounting for the effect of seasonal photosynthesis on autumn senescence dates in forests will be integral to improving the performance of global biogeochemical models. Current vegetation models, such as the Lund-Potsdam-Jena general ecosystem simulator (LPJ-GUESS) (*32*), assume a unidirectional relationship between carbon uptake and growing-season length. The assumption is that growing-season length, which itself is modeled on the basis of temperature parameters in summergreen vegetation types (*33*), determines seasonal carbon capture, not vice versa. As a result of this assumption, vegetation models predict increases in seasonal tree productivity due to ever-longer growing seasons under future warming conditions (*6*). By contrast, the direct control of the growing-season end by the extent of preceding carbon capture (Figs. 1 and 2) predicts that ever-earlier start dates of photosynthetic activity (*4*, *5*) and increased activity during the season (*34*) will drive earlier autumn senescence in the future, placing a constraint on growing-season lengths and the seasonal carbon-capture potential of temperate trees.

We further explored these implications of changes in seasonal productivity on future projections of autumn phenology (Fig. 4, A and B) and growing-season length (Fig. 4. C and D) by running the two best-performing first-generation, second-generation, and PIA models based on climate projections from a business-as-usual emissions scenario (representative concentration pathway [RCP] 8.5) over the rest of the century (see figs. S7 and S8 for species-level results). On average, across all time series (species–site combinations), the first-generation models projected senescence delays ranging between 17 and 19 days by the end of the 21st century (see Fig. 4B and fig. S8 for species-level results), which results from

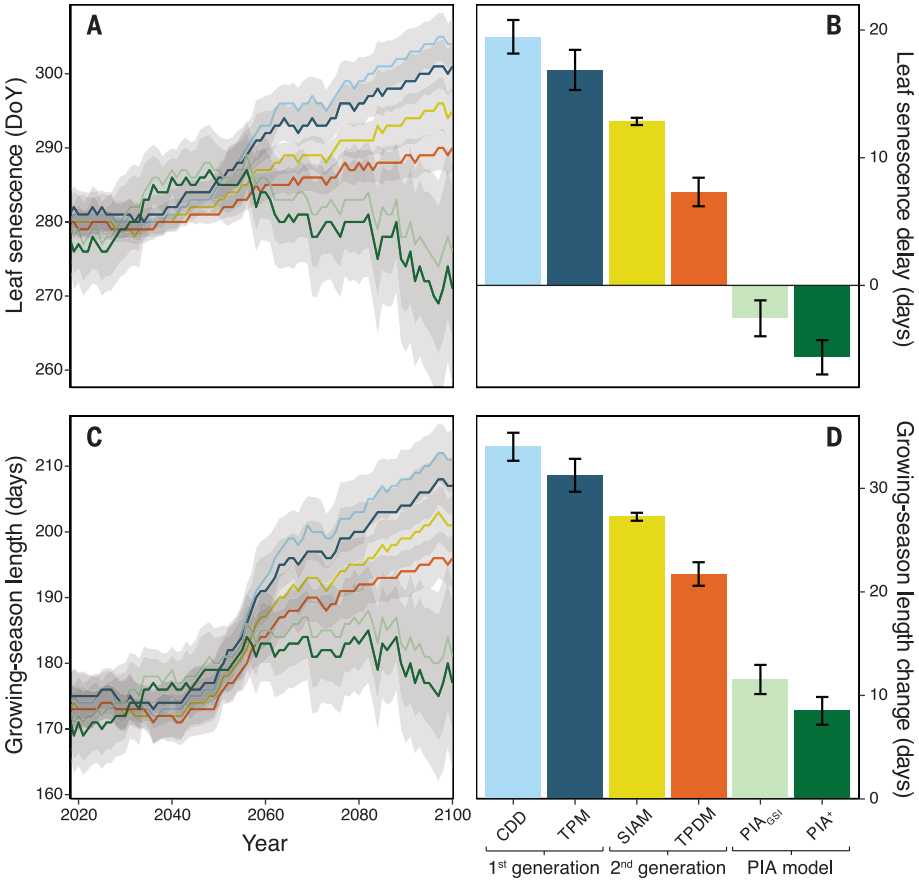


Fig. 4. Future projections of autumn senescence dates and growing-season lengths in Central European deciduous trees. (**A**) Senescence projections (15-year moving averages for six species) under the RCP 8.5 ("business-as-usual") climate scenario, based on the first-generation (CDD and TPM), second-generation (SIAM and TPDM), and PIA models. The gray area indicates 1 SD on either side of the mean. DoY, day of year. (**B**) Estimated delays in leaf senescence by the end of the 21st century (2080 to 2100) compared with the average senescence dates between 1990 and 2010. (**C** and **D**) Same as (A) and (B) but showing overall growing-season length instead of autumn senescence dates. Future leaf-out date predictions to calculate growing-season lengths are based on the M1 spring phenology model.

the large importance of autumn temperatures in these models. Second-generation models projected less pronounced senescence delays (7 to 13 days) because the earlier spring leaf-out and elevated summer temperatures constrain the response of leaf senescence to future autumn warming. In contrast to the delayed senescence predicted by both sets of models, our PIA models project a slight advance of senescence, suggesting that, if growing-season photosynthesis continues to increase, leaves may senesce 3 to 6 days earlier in the year by the end of the century. Our models, therefore, predict that the effect of autumn temperature predominant in first- and second-generation models will be outweighed by increasing spring and summer productivity in the future, reversing expectations of future changes in autumn phenology. With a predicted increase in overall growing-season length by 31 to 34 days at the end of the 21st century in first-generation models, by 22 to 27 days in second-generation models, and by only 8 to 11 days in our PIA models (Fig. 4D), these new predictions lower our expectations of the extent to which longer growing seasons will increase seasonal carbon uptake in forests.

Discussion

The most likely explanation for the negative relationship between seasonal productivity and autumn senescence dates is the control of photosynthesis by sink activity (*17*). The seasonal carbon-sink capacity of plants is limited by nutrient (in particular nitrogen) supply (*17*, *35*) and developmental and translocation constraints, such as tissue maturation and phloem loading (*36*, *37*). In sink-limited plants, carbon uptake during the growing season should, therefore, act as a self-regulatory mechanism constraining the length of the productive season. By affecting carbon uptake, preceding phenological stages, such as spring leaf-out, should be indirectly linked to autumn phenology (*14*–*16*, *38*). These predictions are met in both our experiments and the in situ analyses (Figs. 1C and 2). The effect of growing-season plant productivity on autumn phenology thus offers a mechanistic explanation for how increased carbon capture in response to elevated atmospheric CO_2 is likely to be constrained in sink-limited ecosystems (*39*).

Molecular evidence for sink limitation comes from studies on herbaceous plants that highlight an important effect of source-sink regulatory networks on autumn leaf senescence (*40*–*42*). Carbon saturation due to a critical concentration of sugar in the sink organs has been shown to shift the metabolism of the plant to nutrient remobilization and induce leaf senescence through photosynthetic feedback inhibition (*21*, *41*). We found little evidence to support the alternative hypothesis that lagged effects on autumn phenology are explained by constraints on leaf longevity or water stress (*15*, *23*), as changes in leaf-level photosynthesis did not directly affect autumn senescence (see half-shade treatment in Fig. 2) and low water availability in summer tended to delay, not hasten, autumn senescence (Fig. 1B). Spring leaf-out timing per se was a weaker predictor of autumn phenology than was growing-season productivity (Fig. 1B and fig. S2).

Our results now offer a framework for unifying previous results that earlier spring leaf emergence and warmer summers advance the timing of autumn phenology, whereas drier leafy seasons delay it (*14*, *30*, *31*) (Fig. 1, B and C, and fig. S2): Earlier leaf emergence and warmer summer periods increase growing-season photosynthesis and should thus accelerate sink saturation and leaf senescence (Fig. 1, B and C). By contrast, repeated seasonal droughts impair sugar transport from the leaf to the sink, thus delaying the senescence process (*43*). These effects are explicitly captured by our photosynthesis models because we use leaf-out dates to determine the start of seasonal photosynthesis and include a drought-stress index in our photosynthesis calculation.

Our analysis of the environmental and internal controls of autumn phenology suggests that seasonal productivity is an important driver of leaf senescence in temperate deciduous trees. Increased growing-season productivity led to earlier leaf senescence in our climate-manipulation experiments and across our 434,226 tree observations over the past 6 decades. Our autumn senescence model representing this internal control mechanism outperformed previous models in the literature and reversed expectations of future changes in autumn phenology: Although autumn warming is likely to increase over the rest of the

century, our model forecasts that there might be slight advances, not delays, in autumn senescence dates. This can be explained by sink limitation of plant growth, predicting that carbon uptake during the growing season imposes strong constraints on the length of the productive season through feedbacks between source and sink organs in plants. These results highlight physiological constraints on growing-season lengths and plant productivity in a warming, CO_2-enriched world, which has direct implications for future carbon-cycle and climate projections. An important avenue of research is the implementation of such growing-season length constraints in Earth system and vegetation models, which currently do not consider the role of source-sink feedbacks when predicting the seasonal CO_2 uptake of plants (*44*). Furthermore, given that ecosystems with high nutrient availability and minimal carbon-sink limitation, such as regions dominated by nitrogen-fixing trees, are not expected to show the observed negative relationship between seasonal productivity and senescence dates (*45*), a major research challenge will be to generate a thorough spatial understanding of the extent of sink limitation to forecast plant phenology and forest productivity over space and time.

REFERENCES AND NOTES

1. S. S. Renner, C. M. Zohner, *Annu. Rev. Ecol. Evol. Syst.* **49**, 165–182 (2018).
2. C. M. Zohner, S. S. Renner, *Nat. Ecol. Evol.* **1**, 1655–1660 (2017).
3. A. D. Richardson *et al*., *Agric. For. Meteorol.* **169**, 156–173 (2013).
4. A. Menzel, P. Fabian, *Nature* **397**, 659 (1999).
5. T. F. Keenan *et al*., *Nat. Clim. Chang.* **4**, 598–604 (2014).
6. A. D. Richardson *et al*., *Glob. Change Biol.* **18**, 566–584 (2012).
7. C. M. Zohner, L. Mo, T. A. M. Pugh, J. F. Bastin, T. W. Crowther, *Glob. Change Biol.* **26**, 4042–4055 (2020).
8. M. Chen, E. K. Melaas, J. M. Gray, M. A. Friedl, A. D. Richardson, *Glob. Change Biol.* **22**, 3675–3688 (2016).
9. C. M. Zohner *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12192–12200 (2020).
10. A. S. Gallinat, R. B. Primack, D. L. Wagner, *Trends Ecol. Evol.* **30**, 169–176 (2015).
11. G. Liu, X. Chen, Y. Fu, N. Delpierre, *Ecol. Modell.* **394**, 34–43 (2019).
12. N. Delpierre *et al*., *Agric. For. Meteorol.* **149**, 938–948 (2009).
13. L. Chen *et al*., *Nat. Clim. Chang.* **10**, 777–780 (2020).
14. Y. S. H. Fu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7355–7360 (2014).
15. T. F. Keenan, A. D. Richardson, *Glob. Change Biol.* **21**, 2634–2641 (2015).
16. C. M. Zohner, S. S. Renner, *Oecologia* **189**, 549–561 (2019).
17. M. J. Paul, C. H. Foyer, *J. Exp. Bot.* **52**, 1383–1400 (2001).
18. A. Miller, C. Schlagnhaufer, M. Spalding, S. Rodermel, *Photosynth. Res.* **63**, 1–8 (2000).
19. Y. H. Fu *et al*., *Tree Physiol.* **39**, 1277–1284 (2019).
20. M. R. Guitman, P. A. Arnozis, A. J. Barneix, *Physiol. Plant.* **82**, 278–284 (1991).
21. R. Kumar, E. Bishop, W. C. Bridges, N. Tharayil, R. S. Sekhon, *Plant Cell Environ.* **42**, 2597–2611 (2019).
22. P. B. Reich, M. B. Walters, D. S. Ellsworth, *Ecol. Monogr.* **62**, 365–392 (1992).
23. W. Buermann *et al*., *Nature* **562**, 110–114 (2018).
24. J. U. Jepsen *et al*., *Glob. Change Biol.* **17**, 2071–2083 (2011).
25. B. Templ *et al*., *Int. J. Biometeorol.* **62**, 1109–1113 (2018).
26. Materials and methods are available as supplementary materials.
27. S. Sitch *et al*., *Glob. Change Biol.* **9**, 161–185 (2003).
28. W. M. Jolly, R. Nemani, S. W. Running, *Glob. Change Biol.* **11**, 619–632 (2005).
29. M. Meinshausen *et al*., *Geosci. Model Dev.* **10**, 2057–2116 (2017).
30. Y. Xie, X. Wang, A. M. Wilson, J. A. Silander Jr., *Agric. For. Meteorol.* **250–251**, 127–137 (2018).
31. G. Liu, X. Chen, Q. Zhang, W. Lang, N. Delpierre, *Glob. Change Biol.* **24**, 3537–3545 (2018).
32. B. Smith *et al*., *Biogeosciences* **11**, 2027–2054 (2014).
33. M. T. Sykes, I. C. Prentice, W. Cramer, *J. Biogeogr.* **23**, 203–233 (1996).
34. J. Mao *et al*., *Nat. Clim. Chang.* **6**, 959–963 (2016).
35. A. Fangmeier, B. Chrost, P. Högy, K. Krupinska, *Environ. Exp. Bot.* **44**, 151–164 (2000).
36. S. C. Wong, *Photosynth. Res.* **23**, 171–180 (1990).
37. C. Körner, *Plant Cell Environ.* **18**, 1101–1110 (1995).
38. C. M. Zohner, A. Rockinger, S. S. Renner, *New Phytol.* **221**, 789–795 (2019).
39. A. P. Walker *et al*., *New Phytol.* nph.16866 (2020).
40. P. O. Lim, H. J. Kim, H. G. Nam, *Annu. Rev. Plant Biol.* **58**, 115–136 (2007).
41. N. Sade, M. Del Mar Rubio-Wilhelmi, K. Umnajkitikorn, E. Blumwald, *J. Exp. Bot.* **69**, 845–853 (2018).
42. H. R. Woo, H. J. Kim, P. O. Lim, H. G. Nam, *Annu. Rev. Plant Biol.* **70**, 347–376 (2019).
43. B. D. Hesse, M. Goisser, H. Hartmann, T. E. E. Grams, *Tree Physiol.* **39**, 192–200 (2019).
44. C. Terrer *et al*., *Nat. Clim. Chang.* **9**, 684–689 (2019).
45. G. Taylor *et al*., *Glob. Change Biol.* **14**, 264–275 (2008).
46. zanid90/AutumnPhenology: Autumn Phenology repository, Zenodo (2020).

ACKNOWLEDGMENTS

We are grateful to DOB Ecology for support and to L. Bernegger, J. Maschler, A. Rockinger, H. Schmitt, V. Sebald, M. Wenn, and W. Zohner for help with the experiments. L.M. was supported by the China Scholarship Council. **Funding:** DOB Ecology. **Author contributions:** C.M.Z. conceived and developed the study and conducted the experiments. D.Z., C.M.Z., and L.M. performed the analyses. C.M.Z. and D.Z. wrote the manuscript. T.W.C. and S.S.R. gave input to the manuscript. **Competing interests:** The authors declare that there are no competing interests. **Data and materials availability:** All source code, processed covariates, and models are available at Zenodo (*46*). The observational data used for this study are freely available at www.PEP725.eu. The experimental data are available as supplementary data files S1 and S2.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1066/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S5
References (*47–72*)
Data Files S1 and S2
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

25 July 2020; accepted 6 October 2020
10.1126/science.abd8911

REPORT

CHEMICAL PHYSICS

Capturing roaming molecular fragments in real time

Tomoyuki Endo[1,2], Simon P. Neville[3], Vincent Wanie[1]*, Samuel Beaulieu[1]†, Chen Qu[4], Jude Deschamps[1]‡, Philippe Lassonde[1], Bruno E. Schmidt[5], Hikaru Fujise[6], Mizuho Fushitani[6], Akiyoshi Hishikawa[6,7], Paul L. Houston[8,9], Joel M. Bowman[10], Michael S. Schuurman[3,11], François Légaré[1]§, Heide Ibrahim[1]§

Since the discovery of roaming as an alternative molecular dissociation pathway in formaldehyde (H_2CO), it has been indirectly observed in numerous molecules. The phenomenon describes a frustrated dissociation with fragments roaming at relatively large interatomic distances rather than following conventional transition-state dissociation; incipient radicals from the parent molecule self-react to form molecular products. Roaming has been identified spectroscopically through static product channel–resolved measurements, but not in real-time observations of the roaming fragment itself. Using time-resolved Coulomb explosion imaging (CEI), we directly imaged individual "roamers" on ultrafast time scales in the prototypical formaldehyde dissociation reaction. Using high-level first-principles simulations of all critical experimental steps, distinctive roaming signatures were identified. These were rendered observable by extracting rare stochastic events out of an overwhelming background using the highly sensitive CEI method.

Although fragments within a highly excited dissociating molecule generally move along minimum energy paths, they sometimes take a little extra "walk" (*1*, *2*). Traditionally, two possible pathways for dissociation of a molecule are considered: Dissociation to closed-shell products commonly has to overcome a large barrier, which corresponds to a tight transition state originating from electronic rearrangements. By contrast, bond-breaking dissociation to radicals is commonly electronically barrierless. The third channel, roaming, first discovered in formaldehyde (H_2CO), was observed experimentally and explained in theory 16 years ago (*3*). In addition to the production of H_2 + CO through the tight transition state (molecular dissociation channel) and H + HCO (radical dissociation channel), there is a roaming channel involving trajectories of H and HCO with pathways that are far removed from the conventional ones and ultimately lead to the abstraction of H_2. After the first detection of roaming in H_2CO, the phenomenon has been observed to occur in a broad variety of molecular systems and with different roaming fragments (*2*, *4*–*6*).

Roaming has been widely reported to be almost ubiquitous (*2*, *6*). Nevertheless, it maintains its elusive character because one cannot easily describe it in terms of chemical reactivity nor directly observe it in experiments. The current work helps to solve these experimental challenges.

So far, the outcomes of roaming processes can be observed using high-resolution spectroscopic determination of the photoproducts, which is the basis for almost all experimental roaming studies (*2*–*4*, *7*, *8*). In the example of H_2CO, this is evident in a characteristic combination of hot H_2 vibrational– and cold CO rotational–level populations (*3*, *8*, *9*), which is not present in the conventional dissociation channels.

There have been a few time-resolved experiments, but none involved the direct observation of roamers, the roaming fragments, in real time. Rather, previous work has focused on the time-resolved measurement of the outcome of processes involving roaming (*10*–*13*) or on its identification using photoelectron spectroscopy (*14*). For example, Ekanayake *et al.* (*11*, *12*) succeeded in observing the formation of H_3^+ upon strong-field double ionization in a series of alcohols whose creation can involve roaming of the neutral H_2. However, the roamer H_2 itself is invisible to those detection methods.

To the best of our knowledge, no time-resolved measurement has been accomplished to date for the prototype molecule H_2CO (*15*), on which by far most of the theoretical studies were performed (*2*). One reason is that in general roaming cannot be induced in a coherent manner (i.e., localized in time). In the case of H_2CO, roaming occurs on the vibrationally hot electronic ground state, which is fairly long-lived and not known to dissociate promptly. As verified by trajectory calculations (*9*), both the onset and duration of roaming are of a strongly stochastic character. Even though roaming itself is ultrafast, the overall time scales involved (excitation, relaxation, roaming, and dissociation) span hundreds of femtoseconds to tens of nanoseconds. Covering such a time span and thereby isolating rare roaming events is a severe challenge for any ultrafast time-resolved experiment and requires support from numerical simulations.

How can these dynamics be observed experimentally? Because the involved time scales extend over several orders of magnitude and exceed by far the probe-pulse duration, they cannot be sampled in the same way as, e.g., the well-structured coherent vibrational wavepacket motion of a diatomic molecule (*16*) that separates well from any time-independent background. The challenge is to separate a weak structureless signal from a strong structureless background when both evolve with time. A sensitive probe is required to catch each molecule of the ensemble along its individual progress during the roaming reaction. Contributions from the small number of roamers are strongly superimposed by molecules following the molecular (conventional saddle-point) or radical dissociation pathways.

To enhance hidden coherent dynamics and to suppress a dominating incoherent background, one can use coherent control techniques (*17*). However, detecting rare quasi-statistical events such as roaming hidden in an overwhelming background requires the elimination of the background, which can be accomplished by using, e.g., a reconstruction process (*18*). In our approach, we take advantage of Coulomb explosion imaging (CEI) being a quasi-background-free technique, similar to the recently discussed cold-target recoil-ion-momentum spectroscopy of NO_2 (*19*).

Here, we extend conventional femtochemistry based on coherent ensemble motions to track an ultrafast statistical process occurring in an ensemble of isolated molecules: roaming dynamics in deuterated formaldehyde (D_2CO) based on the direct and real-time observation of roamers. We present a comprehensive picture of the ultrafast dynamics triggered in neutral D_2CO molecules upon femtosecond ultraviolet (UV) photoexcitation, including the

[1]Centre Énergie Matériaux Télécommunications, Institut National de la Recherche Scientifique, Varennes, Quebec J3X 1S2, Canada. [2]Kansai Photon Science Institute, National Institutes for Quantum and Radiological Science and Technology, Kizugawa, Kyoto 619-0215, Japan. [3]Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. [4]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. [5]few-cycle Inc., Montreal, Quebec H1L 5W5, Canada. [6]Department of Chemistry, Graduate School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan. [7]Research Center for Materials Science, Nagoya University, Nagoya, Aichi 464-8602, Japan. [8]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14852, USA. [9]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA. [10]Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, GA 30322, USA. [11]National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada. *Present address: Center for Free-Electron Laser Science, DESY, 22607 Hamburg, Germany. †Present address: Fritz Haber Institute of the Max Planck Society, 14195 Berlin, Germany. ‡Present address: Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
§Corresponding author. Email: legare@emt.inrs.ca (F.L.); ibrahim@emt.inrs.ca (H.I.)

reaction competition between the roaming and the dominant D_2 + CO (molecular) and D + DCO (radical) dissociation channels.

Because the detection method that we used catches transient features rather than the product outcome, we found that roaming not only led to the well-known D_2 + CO molecular roaming products, but also sometimes preceded dissociation to radical products (D + DCO). This alternative pathway to dissociation products is shown in fig. S7 and discussed in section 2.4 of the supplementary materials. Such roaming was also captured by the experiment, so that we generalized the definition of roaming as motion over an extended period of time where D and DCO are separated from each other and where the potential is high and the local kinetic energy is low.

The experimental scheme for single-photon excitation with a UV pulse is depicted in Fig. 1 and detailed in section 1.1 of the supplementary materials. Energy was mostly deposited into two vibrational modes on the excited electronic state $\tilde{A}$, which couple to the ground electronic state $\tilde{X}$ and initiate an ultrafast population transfer (wavy arrows at t_1), as discussed later and in section 2.1 of the supplementary materials. The observation of dissociation through triplet electronic states and through multiphoton excitation pathways was excluded (see sections 2.1.5 and 4 of the supplementary materials). Once the population has reached the vibrationally hot ground state, it can follow the three different paths described above: direct dissociation to D, direct dissociation to D_2, or roaming.

At the termination of the roaming path, the D that roamed returns close to its equilibrium distance with respect to the DCO but at a different azimuthal angle than the original position (*8*). From there, it does one of the following: (i) dissociates to give radical products (D + DCO), (ii) returns to the D_2CO configuration to react later, or (iii) returns to the reactive end of DCO, where abstraction to form D_2 occurs. The experiment detected all three of these as roaming regardless of subsequent processes.

All of the indicated time steps describe a statistical process for which chronological order remains, but for which the duration of each step behaves stochastically. The photoexcited ensemble was ionized by a time-delayed probe pulse into the dicationic and tricationic electronic state manifolds. The experiment could neither tell where a specific "trajectory" was coming from nor where it was going, contrary to the simulations. Rather, the probe step provided a snapshot of the current molecular configurations. The resulting charged fragments were assigned to different breakup channels, as described in the supplementary materials. Figure 1 exemplifies the three-body breakup into D^+, D^+, and CO^+. In the following, we link the discussed reaction channels to specific kinetic energy release (KER) regions that are shifted in absolute energy for the rough and the high-level theory and for the experiment (for an overview, see table S5).

We first examined the time-resolved total KER spectra of the correlated D^+, D^+, and CO^+ fragments (Fig. 2A). The low-energy side (<12 eV) exhibited a time evolution that occurred on the hundreds of femtoseconds time scale. Such ultrafast onset was surprising given the conventional assumption that the population transfer from the $\tilde{A}$ to the $\tilde{X}$ state occurs on a time scale of picoseconds to nanoseconds (*20*, *21*), not including the additional actual roaming and dissociation times. Unfortunately, we could not directly assign the reaction channels linked to this observation. We could however roughly approximate the

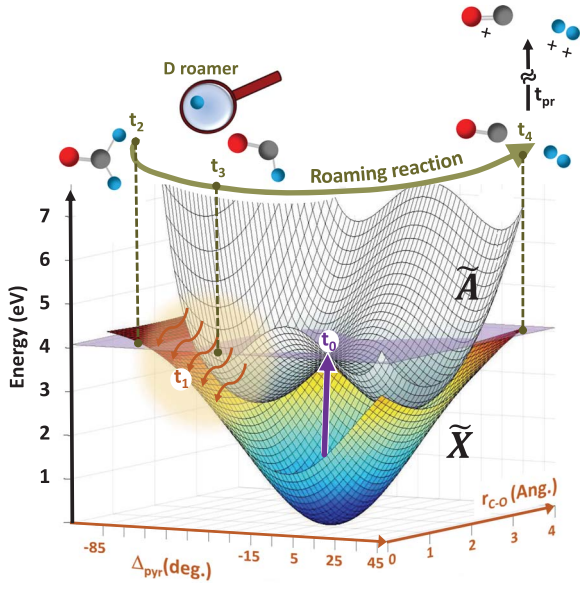


Fig. 1. Scheme of roaming dynamics in D_2CO visualized for two coordinate systems. The orange axis shows relevant vibrational modes: A pump pulse (purple arrow, 304 nm) excites the $\tilde{A}$ state at time t_0 to the purple energy level. Because of vibronic coupling of the out-of-plane bending (pyramidization) and the C–O stretching mode with hot vibrational states of the ground state, a nonradiative population transfer is initiated at t_1. Once on $\tilde{X}$, three reaction types can occur: D_2 + CO molecular dissociation, D + DCO radical dissociation, and roaming, each following its own reaction potential that is not explicitly shown here. Corresponding exemplary molecular structures for three subsequent times are given along the roaming reaction coordinate (green curved line). Excited molecules experience D_2 vibrations (t_2), which lead to one deuteron roaming around the remaining DCO fragment at large interatomic distances (t_3). Eventually, the roamer D abstracts the bound deuteron and together they leave the CO fragment, which leads to D_2 formation and abstraction (t_4). Note that t_1 to t_4 do not correspond to specific absolute times but rather indicate the order of these steps within a statistical process. A strong NIR pulse (800 nm, ~8 fs duration, black arrow) probes the molecular structure at variable time delays (t_{pr}).

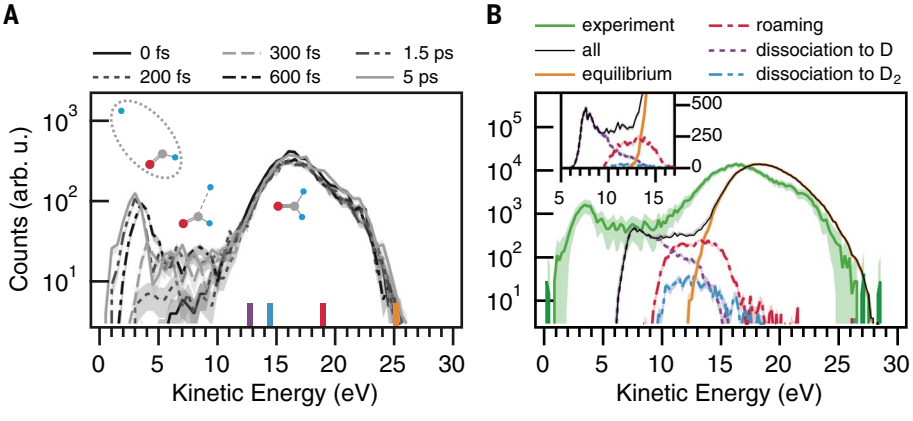


Fig. 2. KER spectra. (**A**) Time-resolved experimental KER spectra from the three-body breakup (D^+, D^+, CO^+). The shaded range indicates statistical errors. Sticks correspond to the simulation results using pure-Coulomb potentials and typical geometries (Fig. 3) following the color coding for different geometries: roaming (red), dissociation to D_2 + CO (blue), dissociation to D + DCO (purple), and equilibrium geometry (orange). (**B**) Comparison of the KER distributions between experiment (green, time integrated for 0 to 1.5 ps) and MD simulations (black, time-integrated for 0 to 2 ps) in logarithm scale. The inset shows the simulated KER spectra in the 5 to 17 eV range in linear scale. The errors of the simulated spectra are comparable to the line width.

relative KER expected for the molecular fragments (D^+, D^+, CO^+) arising from different channels based on pure-Coulomb potentials and classical mechanics (colored sticks). Fragments from a chemically bound structure (e.g., in equilibrium geometric configuration) resulted in higher KER than fragments undergoing roaming or even radical dissociation involving large interatomic distances at the moment of explosion. On the basis of this estimation, we formulated the working hypothesis that the sought-after roaming dynamics could be extracted within this low-energy KER region.

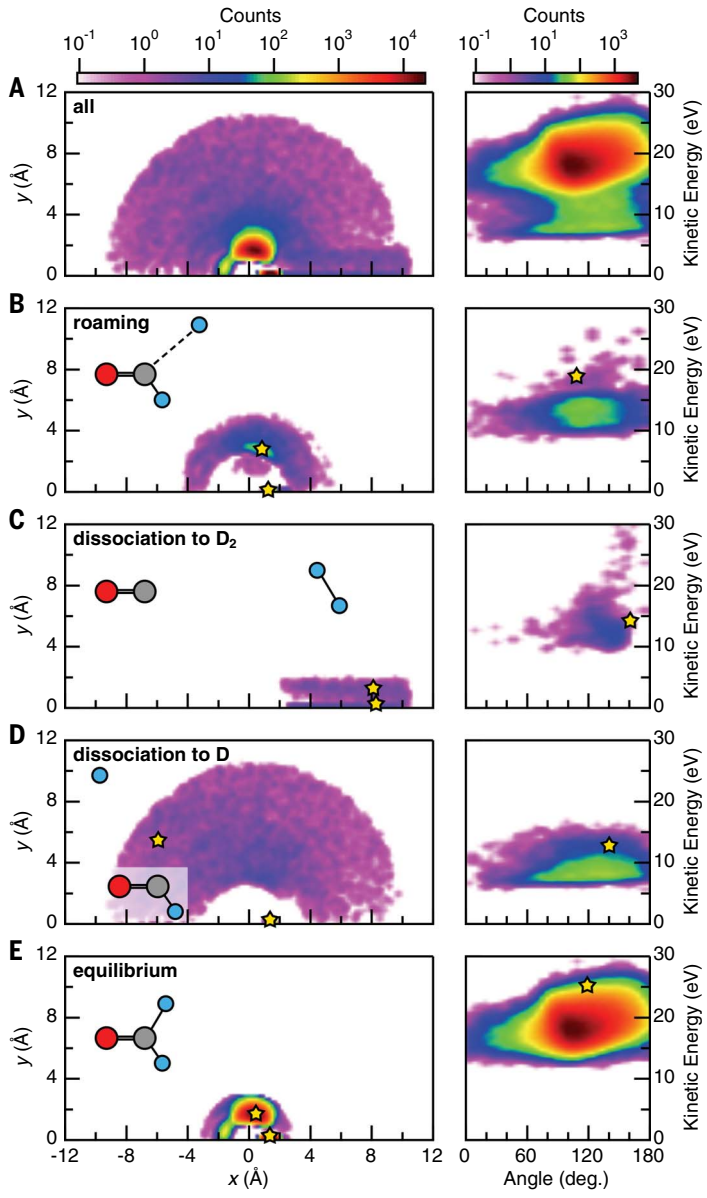


Fig. 3. Trajectory simulations. Shown are trajectory simulations for 37370 cm^{-1} corresponding to 304-nm photon energy (time integrated for 0 to 2 ps, 1500 trajectories). The left column contains the real-space representation before Coulomb explosion and maps the position of the free deuteron with respect to the other fragments. Typical geometries for each pathway are represented on relative scale. The CO fragment is fixed at the $x = 0 \mid y = 0$ coordinate, and its nearby deuteron is confined to the x-axis. In the right column, the calculated corresponding KER versus deuteron momentum-vector angle distribution is given. (**A**) Sum of all trajectories without any gate applied. (**B**) Only roaming results, according to the gate definition in section 2.4 of the supplementary materials. (**C**) Dissociation to D_2. (**D**) Channel of the radical dissociation, with one D atom being close to the CO, the other one being far away. (**E**) Equilibrium geometry with R_1, R_2 < 2.8 Å. Deuteron positions and simulation results using pure-Coulomb potentials are shown as the yellow stars [e.g., in the right panel of (B) at 19 eV and 108°]. The energy offset between theoretical and experimental values is discussed in section 3 of the supplementary materials.

For a more accurate analysis, we used high-level simulations to interpret the experimental results, as detailed in the supplementary materials, including (i) modeling the time scale for nonradiative decay from $\tilde{A}$ to $\tilde{X}$ (section 2.1), (ii) evolution on the vibrationally hot ground state with molecular dynamics (MD) trajectory simulations (section 2.2), and (iii) subsequent probing with an ionizing laser pulse that results in prompt "explosion" of the triply charged ion (section 2.3). Briefly, in step (ii), the motion on the hot ground state was simulated using established quasi-classical trajectory calculations on D_2CO in a manner similar to previous studies of H_2CO (*22*–*24*) but using a recent ab initio potential energy surface (PES) (*25*) that has been shown to produce excellent agreement with experimental results (*9*). Step (iii) simulated the experimental probe step by projecting geometries for trajectory outcomes onto the triply charged state. Although the subsequent separation of the ions was principally controlled by the Coulomb repulsion, it was not sufficiently precise to assume a purely repulsive triply charged state (Fig. 2A). We have therefore calculated a realistic D_2CO^{3+} PES representative of the present Coulomb explosion imaging experiment using the CASSCF(7,7) method with GAMESS package and fitted it adopting both energies and gradients using permutational invariant polynomials.

The KER distribution obtained from theory steps (ii) and (iii) was compared with the time-integrated experimental observation ranging from 0 to 1.5 ps in Fig. 2B. The shape of the time-integrated theory spectrum ranging from 0 to 2 ps (solid black line) was consistent with experimental results (green line), showing the main peak below 20 eV, a small peak at a lower KER region, and a plateau between these peaks. Deviations of 2 to 4 eV in absolute energy were attributed to the finite duration of the probe pulse and bond-softening effects (*26*, *27*), as discussed in section 3 of the supplementary materials. This overall theory spectrum was decomposed into contributions from roaming, dissociation to D_2, dissociation to D, and the equilibrium geometry, applying the trajectory assignment discussed in the following. As the KER of different photodissociation pathways overlapped, additional means were required to separate their individual contributions. From geometric considerations, we assumed that the angles between deuteron momentum vectors should exhibit a broad distribution for the case of radical dissociation, where one deuteron encircles the remaining DCO. By contrast, one would expect a confined angular distribution close to 180° for the case of molecular dissociation because the D_2 fragment would explode

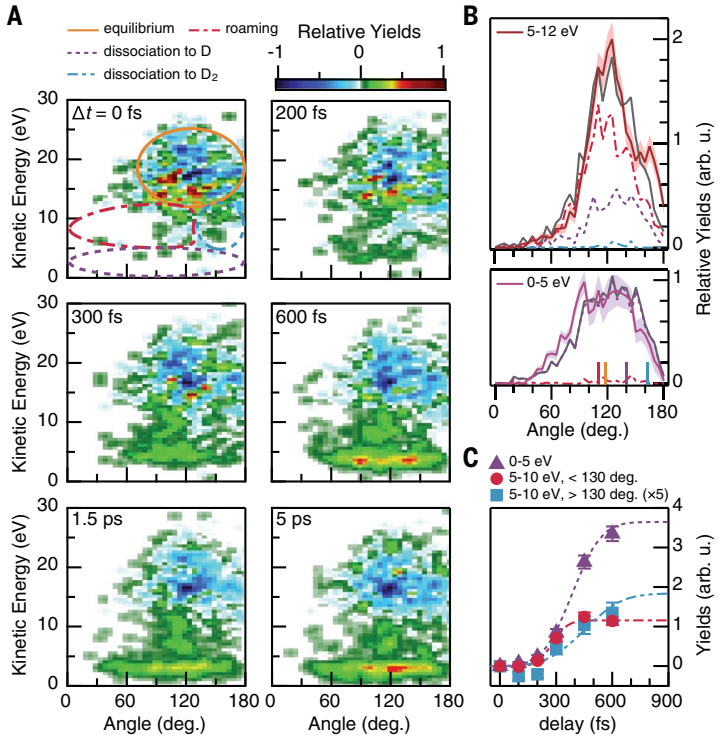


Fig. 4. Experimental results for 304-nm pump wavelength of the three-body breakup (D^+, D^+, CO^+) channel. (**A**) Time evolution of the experimental KER versus deuteron momentum angle representation. The pump-probe delay Δt is indicated for each subplot. The distribution at Δt = −300 fs was subtracted as background from each pump-probe delay. Indicated in the upper left subplot are the areas corresponding to the dissociation channels as identified in Fig. 3. The blue circle and purple ellipse represent dissociation to D_2 and D, respectively. The red ellipse shows the area relevant for roaming and the orange ellipse the one of the equilibrium configurations. Each image is smoothed with a 3 × 3 binomial filter. (**B**) Time-integrated angular distribution of deuteron momenta for experiment (red and purple solid lines) and MD simulations. Effective contributions of each trajectory are shown individually for roaming (red), dissociation to D (purple), and dissociation to D_2 (blue). The solid gray line corresponds to the superposition of all contributions in the low- and high-KER range (bottom and top panel, respectively). For determining the effective contribution, see section 5.1 of the supplementary materials. The shaded range indicates statistical errors of the experimental data. Colored sticks represent classical calculations as in Fig. 2. (**C**) Yields of selected KER and angular regions as a function of Δt (0 to 900 fs). The full range plot is given in fig. S16. Statistical errors are shown as error bars.

almost like a free diatomic molecule. The angular distribution of the roaming channel was assumed to be in between these two because both deuterons are separated, but the repulsive force on the triply charged state between D and DCO is stronger than for radical dissociation. This was kept in mind for the following analysis.

In search for an experimental observable to separate different reaction channels, we analyzed how the related simulated trajectories of (ii) behaved in real space (Fig. 3, left). A geometric representation of all trajectories, regardless of their paths, is given in Fig. 3A in a time-integrated manner up to pump-probe delays of 2 ps. For comparison, diagrams of typical molecular structures are given in the left column and relate to the classically calculated sticks in Figs. 2A and 4B, as well as to the yellow stars in Fig. 3.

Following the criteria discussed for H_2CO (*28*), the characteristics of each path were identified as specific distances and angles between the molecular fragments (see section 2.2 of the supplementary materials). On the basis of this assignment, gates to isolate roaming and dissociation in D_2CO were created, which do not focus on the outcome of the propagation, as do previous methods, but are sensitive to transient features. Details are given in section 2.4 of the supplementary materials. These gates are defined based on the statistical characteristics of each channel from the set of in total 1500 trajectories. Because roaming (Fig. 3B) can be understood as a hindered dissociation, the deuteron never fully detaches from the DCO—it even returns to it. Such definition is not restricted to trajectories that have completed roaming as, e.g., in (*3*) and (*28*), but can extract the actual transient roaming behavior of all trajectories whether they complete the roaming path or end up in a conventional dissociation channel. Such information is only accessible in a time-resolved experiment. By contrast, dissociation to produce D_2 was confined to a rectangular region because both deuterons stayed close together and departed from the central CO fragment in Fig. 3C. The radical dissociation contained a fully departing deuteron that manifested itself in an extended semicircle (Fig. 3D). The residual of Fig. 3E corresponds to the equilibrium geometry, leading to the highest KER distribution peak around 16 eV.

To predict an experimental observable, these calculated geometries were projected onto and propagated on the present D_2CO^{3+} PES. The simulated KER versus deuteron momentum-angle distributions of the (D^+, D^+, CO^+) channel are shown in the right column of Fig. 3. Applying the gates discussed above to simulate the experimental observable provides a tool to identify the contribution of each path. For roaming, a center region from 11 to 15 eV extending over a limited range of angles around 115° was identified (Fig. 3B). The molecular dissociation in Fig. 3C showed up at localized angles over a broad energy range because of the different bond lengths in D_2 involved. By contrast, the radical dissociation was discernible as a sharp stripe in the energy region from 6 to 10 eV and spread over all angles (Fig. 3D). The equilibrium distribution is shown in Fig. 3E.

Even though the KER versus angle distributions after gating overlapped in certain regions, we were able to identify for each pathway an area in the angle-KER plane of clear dominance, as detailed in section 2.4 of the supplementary materials. The KER (Fig. 2B) and angle (Fig. 4B) projections revealed, e.g., that roaming was clearly dominant from 10 to 13 eV and from 80° to 115°.

We prepared the experimental data accordingly to separate the sought-after roaming dynamics from the other dissociation pathways in Fig. 4A. In the KER region <12 eV, the onset of dynamics occurred already before 100 fs, followed by a constant increase up to ~1.5 ps. The corresponding angle-integrated data for selected energy regions (Fig. 4B) and for additional time delays and excitation wavelengths are presented in the supplementary materials. Building on previously discussed theory results, the four different regions were identified, as outlined in the upper left plot: roaming (red), dissociation to D_2 (molecular dissociation, blue), dissociation to D (radical dissociation, purple), and the equilibrium contribution (orange). Because of the energy offset between theory and experiment discussed previously, the four regions appeared at lower energies compared with theory. The initial distribution at equilibrium (orange), where most of the counts remained during the time of the

experiment, was centered at 20 eV and depleted with time. By far most of the radical dissociation (purple) covered kinetic energies between 0 and 5 eV and extended over a broad angle area between deuteron momenta. At kinetic energies between 5 and 12 eV, it was possible to discern two separated angular distributions in the experiment that we attributed to roaming (<130°) and molecular dissociation (>130°). The angular distribution of the roaming channel (red) was centered around 120°; the weak contribution of molecular dissociation (blue) was centered around 150° and, as expected, was localized. Thus, even though both roaming and dissociation to D_2 shared the same KER region of 5 to 12 eV, we were able to clearly separate them in their distributions of deuteron angular momentum. This fact allowed us to identify for each reaction path an area of dominant detection probability.

The branching ratios from trajectory simulations on the vibrationally hot ground state predicted 9% roaming, 57% radical dissociation, and 34% molecular dissociation for excitation at 304 nm. However, as discussed in section 2.4 of the supplementary materials, this result did not reproduce the actual experimental situation. Because of the effect of enhanced ionization (*27*), the experiment favored the observation of roaming over dissociation.

To derive the branching ratios of the different pathways in our experiment, we decomposed the time-integrated angular distributions of the two KER regions of interest (red and purple solid lines in Fig. 4B) into the effective contributions of each trajectory estimated by the time-resolved fitting (see section 5.1 of the supplementary materials). The sum of individual weights well reproduced the experimental distribution.

After the previous analysis, we expected almost exclusively contributions from the radical dissociation channel in the lower KER range. In both theory and experiment, we observed that these counts were broadly spread over all momentum vector angles (dashed and solid purple lines in the lower panel of Fig. 4B). In the higher KER range (red solid line in the upper panel), we identified substantial contributions from roaming and radical dissociation, as well as weak molecular dissociation contributions. From the MD simulations, we concluded a clear domination of roaming, especially in the angular region around 120°.

From this fit, we concluded that in the KER region of 5 to 12 eV, 70% of all counts originated from roaming and 29% from dissociation to D. Only 1% resulted from dissociation to D_2, which coincided well with the ratio from the molecular two-body breakup channel compared with the D + DCO channel presented in section 5.2 of the supplementary materials and related to the chosen pump energy, as predicted by theory (*28*, *29*). We thus observed evidence that the experimental signal located in the angle-selected red ellipse of Fig. 4A arose predominantly from the transient roaming signal.

A degree of overlap between roaming and the radical dissociation channel cannot be ruled out because of the nature of involved PESs. Because the molecule itself does not seem to "know" whether it proceeds toward roaming or dissociation in this range, the experiment was not able to distinguish this at this point either. However, the detection method applied here, together with the theory used, provided a means to identify the leading contribution of roaming in the selected area.

From trajectory simulations, it is known that roaming dynamics can take different forms in terms of when the proton starts to move, how many times it encircles the remaining HCO fragment, and when it finally abstracts the second H atom (*28*). In the experiment with 304-nm excitation, these individual paths merged to a distribution with an ensemble average value. A signal rise time of 306 ± 39 fs for the roaming channel and 385 ± 9 fs for the radical dissociation channel was deduced from Fig. 4C and fig. S16.

This rise time represented a superposition of the population transfer time from $\tilde{A}$ to $\tilde{X}$, any transient dissociation time, as well as the actual roaming time. The roaming signal appeared already at a time delay of 100 fs and constantly increased up to 1.5 ps, being persistent up to several nanoseconds. This time dependence is within the times identified in trajectory simulations for H_2CO: 78 fs for roaming and 1.7 ps for dissociation once the population had reached the vibrationally excited $\tilde{X}$ state.

The previously mentioned ultrafast population transfer from the $\tilde{A}$ state to high-lying vibrational states of $\tilde{X}$ occurred through radiationless vibronic coupling, as discussed in section 2.1 of the supplementary materials, and is calculated to correspond to ≈3% of the $\tilde{A}$-state population within the first picosecond after excitation. This degree of population transfer was adequate to obtain sufficient counts in a quasi-background-free technique such as CEI that is capable of separating such dynamics from an otherwise overwhelming background stemming from conventional reaction pathways such as dissociation. Our ultrafast detection method with single-molecule sensitivity allowed us to observe the onset of dynamics orders of magnitude faster than previously assumed. This is the core prerequisite for the roaming measurements reported herein.

In conclusion, we have directly observed transient roaming reactions in real time that involved both completed and uncompleted roaming paths, based on a generalized definition of roaming. The onset of roaming in D_2CO occurred several orders of magnitude earlier than expected because the onset of the $\tilde{A} \rightarrow \tilde{X}$ transition in these experiments occurred on the hundreds of femtoseconds rather than the nanoseconds time scale as commonly assumed. This observation was possible because an essentially background-free detection was able to successfully isolate the selected channels. The present results open a new perspective for detecting weak statistic dynamics hidden in an overwhelming background and could be extended to a vast variety of molecular systems.

REFERENCES AND NOTES

1. J. M. Bowman, A. G. Suits, *Phys. Today* **64**, 33–37 (2011).
2. J. M. Bowman, P. L. Houston, *Chem. Soc. Rev.* **46**, 7615–7624 (2017).
3. D. Townsend *et al*., *Science* **306**, 1158–1161 (2004).
4. F. A. L. Mauguière *et al*., *Annu. Rev. Phys. Chem.* **68**, 499–524 (2017).
5. J. M. Bowman, *Mol. Phys.* **112**, 2516–2528 (2014).
6. A. G. Suits, *Annu. Rev. Phys. Chem.* **71**, 77–100 (2020).
7. R. Fernando *et al*., *J. Phys. Chem. A* **119**, 7163–7168 (2015).
8. M. S. Quinn *et al*., *Science* **369**, 1592–1596 (2020).
9. P. L. Houston *et al*., *J. Chem. Phys.* **147**, 013936 (2017).
10. A. S. Mereshchenko, E. V. Butaeva, V. A. Borin, A. Eyzips, A. N. Tarnovsky, *Nat. Chem.* **7**, 562–568 (2015).
11. N. Ekanayake *et al*., *Sci. Rep.* **7**, 4703 (2017).
12. N. Ekanayake *et al*., *Nat. Commun.* **9**, 5186 (2018).
13. E. Livshits, I. Luzon, K. Gope, R. Baer, D. Strasser, *Commun. Chem.* **3**, 49 (2020).
14. L. Poisson *et al*., *Phys. Chem. Chem. Phys.* **16**, 581–587 (2014).
15. C.-M. Tseng, M. Fushitani, A. Matsuda, A. Hishikawa, *J. Electron Spectrosc. Relat. Phenom.* **228**, 25–30 (2018).
16. M. J. Rosker, M. Dantus, A. H. Zewail, *J. Chem. Phys.* **89**, 6113–6127 (1988).
17. H. Ibrahim, M. Héjjas, N. Schwentner, *Phys. Rev. Lett.* **102**, 088301 (2009).
18. R. Fung *et al*., *Nature* **532**, 471–475 (2016).
19. X. Ding *et al*., *J. Chem. Phys.* **151**, 174301 (2019).
20. E. S. Yeung, C. B. Moore, *J. Chem. Phys.* **58**, 3988–3998 (1973).
21. B. Fu, B. C. Shepler, J. M. Bowman, *J. Am. Chem. Soc.* **133**, 7957–7968 (2011).
22. S. A. Lahankar *et al*., *J. Chem. Phys.* **125**, 44303 (2006).
23. S. A. Lahankar, S. D. Chambreau, X. Zhang, J. M. Bowman, A. G. Suits, *J. Chem. Phys.* **126**, 044314 (2007).
24. S. A. Lahankar *et al*., *Chem. Phys.* **347**, 288–299 (2008).
25. X. Wang, P. L. Houston, J. M. Bowman, *Philos. Trans. A Math Phys. Eng. Sci.* **375**, 20160194 (2017).
26. A. D. Bandrauk, M. L. Sink, *J. Chem. Phys.* **74**, 1110–1117 (1981).
27. H. Ibrahim, C. Lefebvre, A. D. Bandrauk, A. Staudte, F. Légaré, *J. Phys. At. Mol. Opt. Phys.* **51**, 042002 (2018).
28. P. L. Houston, R. Conte, J. M. Bowman, *J. Phys. Chem. A* **120**, 5103–5114 (2016).
29. J. D. Farnum, X. Zhang, J. M. Bowman, *J. Chem. Phys.* **126**, 134305 (2007).
30. T. Endo *et al*., Dataset for: Capturing roaming molecular fragments in real-time, ion momentum data, photoelectron data, quasiclassical simulation data with five different wavelengths, potential energies, and gradients of triply charged state. Zenodo (2020); https://doi.org/10.5281/zenodo.4089116.

ACKNOWLEDGMENTS

We thank A. Laramée and G. Lebrun for technical support; B. Wales and J. Sanderson for providing the analysis software for the CEI measurement files; A. Stolow for discussions; S. Gräfe for simulations on the ionization potential; and the reviewers for comments that truly helped to improve the reliability of our analysis. **Funding:** This work was supported by the Canada Foundation for Innovation, NSERC, FRQNT, the JSPS Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers (grant no. S2601 to T.E. and A.H.), and the World Research Unit (B-1) of Reaction Infography (R-ing) at Nagoya University, Japan (A.H. and F.L.). V.W. and S.B. acknowledge financial support from the NSERC-Vanier fellowships program. J.M.B acknowledges financial support from NASA (grant no. 80NSSC20K0360). **Author contributions:** T.E. performed all experiments and data analysis. V.W., S.B., J.D., P.L., and B.E.S. assisted with CEI experiments in ALLS. F.L. was responsible for the ALLS facility at INRS in Canada. H.F., M.F., and A.H. assisted with the photoelectron experiments and analysis

at Nagoya University, Japan. S.P.N. and M.S.S. provided the quantum dynamics simulations. P.L.H. supervised and performed trajectory calculations and analysis. J.M.B. directed the construction of PESs and trajectory code. C.Q. and T.E. calculated the propagation on the D_2CO^{3+} PES. A.H. and F.L. discussed results and provided classical interpretation. H.I. conceived and directed the project, supervised experiments, and analyzed data. T.E., P.L.H., J.M.B., and H.I. wrote the manuscript with assistance from all other authors. **Competing interests:** The authors declare no competing 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. All data have been uploaded to Zenodo (*30*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1072/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S10
Equations S1 to S20
References (*31*–*62*)
MCTDH Operator File and PES Cuts

17 April 2020; resubmitted 16 September 2020
Accepted 23 October 2020
10.1126/science.abc2960

TOPOLOGICAL SYSTEMS

Exceptional nexus with a hybrid topological invariant

Weiyuan Tang[1], Xue Jiang[2,1], Kun Ding[3,4]*, Yi-Xin Xiao[4,5], Zhao-Qing Zhang[4], C. T. Chan[4], Guancong Ma[1]*

Branch-point singularities known as exceptional points (EPs), which carry a nonzero topological charge, can emerge in non-Hermitian systems. We demonstrate with both theory and acoustic experiments an "exceptional nexus" (EX), which is not only a higher-order EP but also the cusp singularity of multiple exceptional arcs (EAs). Because the parameter space is segmented by the EAs, the EX possesses a hybrid topological invariant (HTI), which consists of distinct winding numbers associated with Berry phases accumulated by cyclic paths on different complex planes. The HTI is experimentally characterized by measuring the critical behaviors of the wave functions. Our findings constitute a major advance in the fundamental understanding of non-Hermitian systems and their topology, possibly opening new avenues for applications.

Realistic physical systems are predominantly described by physical laws that are continuous and differentiable. The occasional occurrence of singularities usually entails exotic physics. In many cases, a singularity can be characterized by a topological invariant. For example, Weyl points are singularities of Berry potential, and two Weyl points of opposite charges are connected by a spectral excitation known as a Fermi arc (*1*). Also, "nexus fermions" have recently been found as high-degeneracy points that typically emerge as kissing points of nodal lines (*2*, *3*). The recent advance of non-Hermitian physics (*4*–*6*) is underlain by a type of singularity called an exceptional point (EP) (*7*, *8*). Mode switching can occur when an EP is encircled (*9*), resulting in the possibility of fractional topological charge. The discovery of EP structures such as EP rings (*10*, *11*) and EP surfaces (*12*, *13*) has further enriched the potential topological classes in non-Hermitian systems. It was also found that order-2 EPs (EP2s) can possess a 1/2 topological charge of opposite sign, leading to the unexpected bulk Fermi arcs (*14*).

These findings show that topological notions can lead to more profound understandings of non-Hermitian systems (*15*–*19*). However, thus far all the examples are EP2s that only involve two coalescing states. Lately, the realization of an order-3 EP (EP3) in photonics has brought new possibilities for fascinating applications (*20*). Here, we study a three-state non-Hermitian system and demonstrate with both theory and experiments the topological properties of an exceptional nexus (EX). Besides being an EP3 (*20*), the EX is a cusp singularity at the intersection of multiple exceptional arcs (EAs) consisting entirely of EP2s. The EX entails exotic geometry in its vicinity, which manifests as different winding numbers associated with distinct Berry phases obtained in different complex planes. We characterize the multiplicity of winding numbers by a hybrid topological invariant (HTI), which is a topological notation that links to the behaviors of eigenvectors, making it different from "discriminant numbers" that associate with eigenvalues (*21*). Using acoustic experiments, we not only observed the EAs and the EX but also confirmed the HTI by leveraging the relation between HTI and eigenvector phase rigidity.

[1]Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China. [2]Department of Electronic Engineering, Fudan University, Shanghai 200433, China. [3]Blackett Laboratory, Department of Physics, Imperial College London, London SW7 2AZ, UK. [4]Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. [5]School of Physical Science and Technology, Soochow University, Suzhou 215006, China.
*Corresponding author. Email: k.ding@imperial.ac.uk (K.D.); phgcma@hkbu.edu.hk (G.M.)

We begin by considering a three-level non-Hermitian Hamiltonian

$$H = (\omega_0 - i\gamma_0)\mathbf{I} + \kappa \begin{pmatrix} \sqrt{2}i(1+\Lambda) & -1 & 0 \\ -1 & i\Xi & -1 \\ 0 & -1 & -\sqrt{2}i(1+\Lambda) \end{pmatrix} \quad (1)$$

where ω_0 is the onsite eigenfrequency, $-i\gamma_0$ represents intrinsic onsite loss ($\gamma_0 > 0$), $\mathbf{I}$ is an identity matrix, and κ is the magnitude of the hopping among the three states. The Hamiltonian depends on two complex parameters: $\Xi = \delta_f + i\delta_A$, $\Lambda = \delta_g + i\delta_B$, where δ_B, δ_g, δ_A, $\delta_f \in \mathbb{R}$. Specifically, δ_B and δ_A are the detuning in the respective sites, and δ_g and δ_f are loss or gain. The non-Hermiticity coming from loss and gain is normalized by the parameter κ so that the parameters Ξ and Λ can become dimensionless for the sake of simplicity. Equation 1 has two EP3s at $(\Xi, \Lambda) = (0, 0)$ and $(0, -2)$, at which all three states coalesce. Because the system has mirror symmetry about $\Lambda = -1$, we focus only on the EP3 at (0, 0), marked by a red star in Fig. 1, A and B.

An EP3 is the coalescence of two or more EP2s. The parameter space defined by Ξ and Λ here provides sufficient degrees of freedom to realize EP2 structures. In Fig. 1, A and B, we can identify several trajectories of EP2s, which we call EAs. To understand their origins, we begin from the red pair of EAs on the $\delta_A\delta_g$ plane (Fig. 1A, red plane). Their existence is protected by parity-time symmetry, which holds on the $\delta_A\delta_g$ plane. The red EAs can be viewed on the

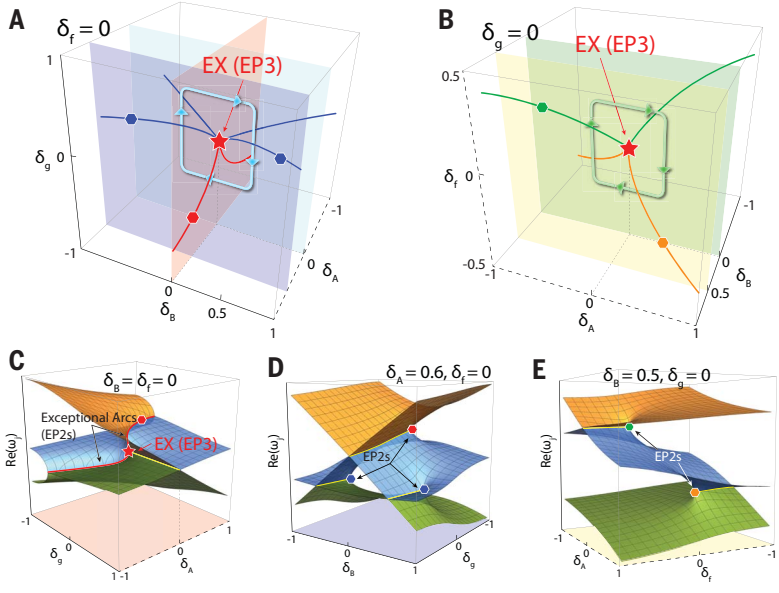


Fig. 1. The EAs joining at the EX (EP3). (**A** to **C**) Six EAs and the EX (red star) in the $\delta_A\Lambda$ space (A); four EAs and the EX in the $\delta_B\Xi$ space (B). The red EAs can be seen on the ERSs in (C). (**D** and **E**) ERSs on the blue and yellow planes, respectively. The solid and dashed axes belong to the complex Λ plane and Ξ plane, respectively. The yellow curves in (C) to (E) are nodal curves for real eigenfrequencies. The cyan and green planes in (A) and (B) are respectively Λ and Ξ planes, on which the EX is encircled to obtain the HTI.

eigenvalue Riemann surfaces (ERSs) in Fig. 1C, which reveals how they emerge. For example, negative δ_A reduces the eigenfrequency of the middle state, bringing it closer to the lower state, so that the two states coalesce at a specific $\delta_g \in (-1, 0)$ and produce an EP2, whose position is continuously tunable by the variation of (δ_A, δ_g), leading to the formation of a continuous and smooth arc. Another EA at a positive-valued δ_A must exist, because the sign change of δ_A indicates the coalescence of the upper and middle states. It follows that the two EAs involving three different states must converge at $(\delta_A, \delta_g) = (0, 0)$ to form the EX; that is, the EX is an EP3. Near the EX, the EAs follow

$$\delta_A^2 + \left(\frac{4}{3}\delta_g\right)^3 = 0 \qquad (2)$$

which describes a semicubical parabola producing a cusp at $(\delta_A, \delta_g) = (0, 0)$. Also, the EAs approach $\delta_A \to \pm\infty$ when the system becomes Hermitian with $\delta_g \to -1$, which means that they are open in the parameter space.

Although parity-time symmetry no longer holds away from the $\delta_A\delta_g$ plane, the emergence of EAs in the (Ξ, Λ) space is deterministic without requiring any symmetry (*22*). Figure 1, A and B, shows four additional pairs of EAs (blue, green, and orange curves). Because these EAs are determined by three parameters, we can only observe their intersections with a parameter plane. Two examples are shown in Fig. 1, D and E, in which the EP2s are clearly identified.

To observe the EAs and the EX, we realized Eq. 1 using a coupled acoustic cavity system (*16*). As shown in Fig. 2A, three identical cuboid cavities are connected by small tubes to implement the tight-binding hopping because the on-site acoustic mode resonates along the z direction, and the waves decay exponentially inside the tube. Here, we chose the second longitudinal mode with eigenfrequency $\omega_0 = 19{,}613$ rad/s. The hopping was set at $\kappa = 49.9$ rad/s by properly choosing the cross section of the tubes (*22*). We tuned specific parameters and measured the pressure response spectra at the central position of each cavity, from which the eigenfrequencies can be retrieved using a Green's function method. In Fig. 2C, the real eigenfrequencies on the $\delta_A\delta_g$ plane are shown as functions of δ_A at several different values of δ_g. EP2s can clearly be identified in the results (Fig. 1C). The EP2s from the experiments are marked in Fig. 2B, which falls on the EAs computed from our model. The EX can be observed by the convergence of two EP2s when $\delta_g \approx 0$, signifying the coalescence of all three states.

One pair of the blue EAs at $\delta_B, \delta_g > 0$ in Fig. 1A was also observed (Fig. 2, D and E).

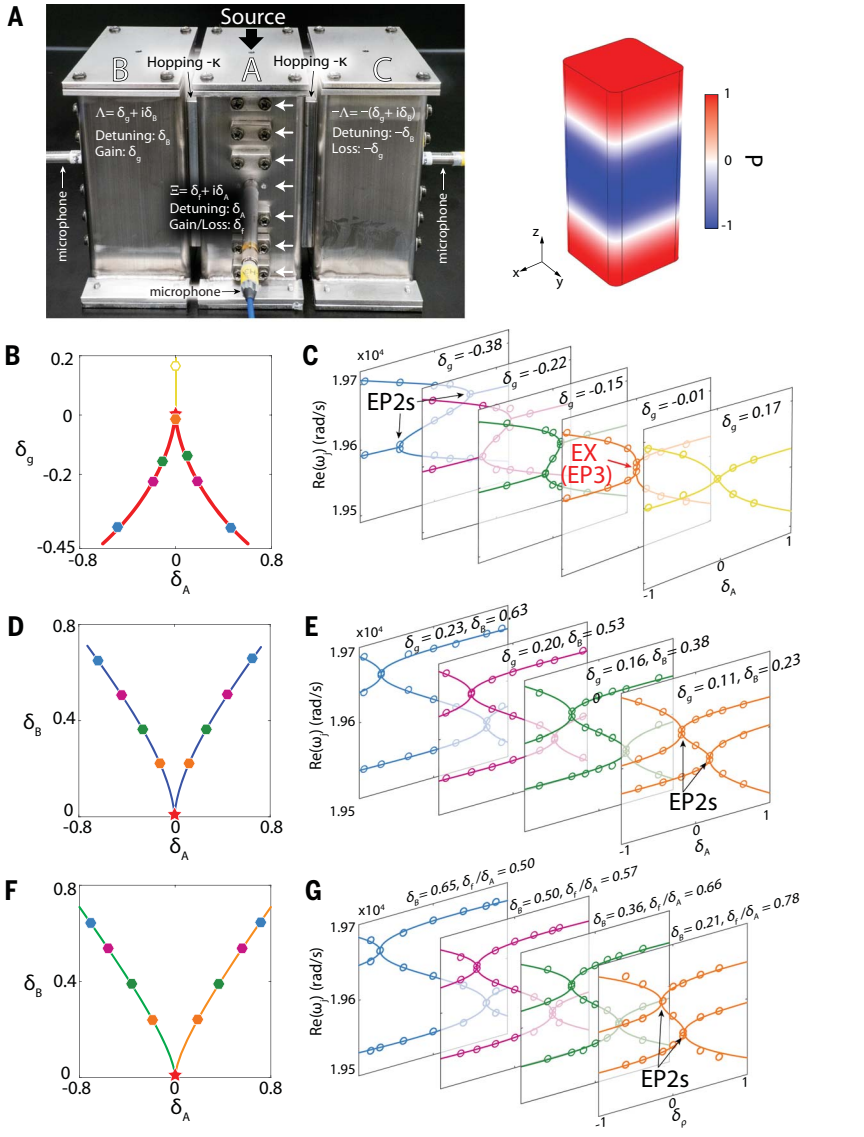


Fig. 2. Observation of EAs and the EX. (**A**) A ternary acoustic cavity system realizing Eq. 1. The pressure profile P of the onsite mode is shown at the right. (**B**) The EAs (red) on the $\delta_A\delta_g$ plane. The solid markers are experimentally identified EP2s. The open marker corresponds to a point in the broken phase. (**C**) The measured real parts of eigenfrequencies on the $\delta_A\delta_g$ plane (markers). (**D** and **F**) The EAs projected to $\delta_A\delta_B$ plane. (**E** and **G**) The measured real parts of the respective eigenfrequencies. In (G), the horizontal axes of each panel are along a different parametric direction, denoted $\delta_\rho = \delta_f/\sin\alpha$ with $\alpha = \arctan(\delta_f/\delta_A)$. In (B) to (G), estimated errors are no larger than the size of the markers.

Likewise, one green and one orange EA in Fig. 1B are shown in Fig. 2, F and G. Note that the observation of these EAs requires the tuning of three parameters; consequently, the projection of the EAs on the $\delta_A\delta_B$ plane is shown in Fig. 2, D and F.

The EAs impose intriguing geometry near the EX. Although the eigenvalue anisotropy of the EX has been experimentally confirmed (*22*), more profound insights can be obtained by considering its topological property embedded in the wave functions. Note that a topological charge cannot be computed by enclosing the EX with a surface, because any enclosing surface must encounter the EAs. Nonetheless, cyclic variations within the Ξ and Λ planes can each produce an encircling path without crossing any EAs. When encircled on the Ξ plane (Fig. 1B, green loop), all eigenstates are recovered after three complete cycles, as seen on the ERSs as shown in Fig. 3, A and B. We used a parallel transport method (*23*) to drive the three eigenvectors to evolve on the same loop. The resultant eigenvectors picked up a gauge-invariant quantized Berry phase of -2π after three cycles (Fig. 3C), indicating a fractional winding number $\mathcal{W}_\Xi = -2/3$. By contrast, encircling on the Λ plane (Fig. 1A, cyan loop)

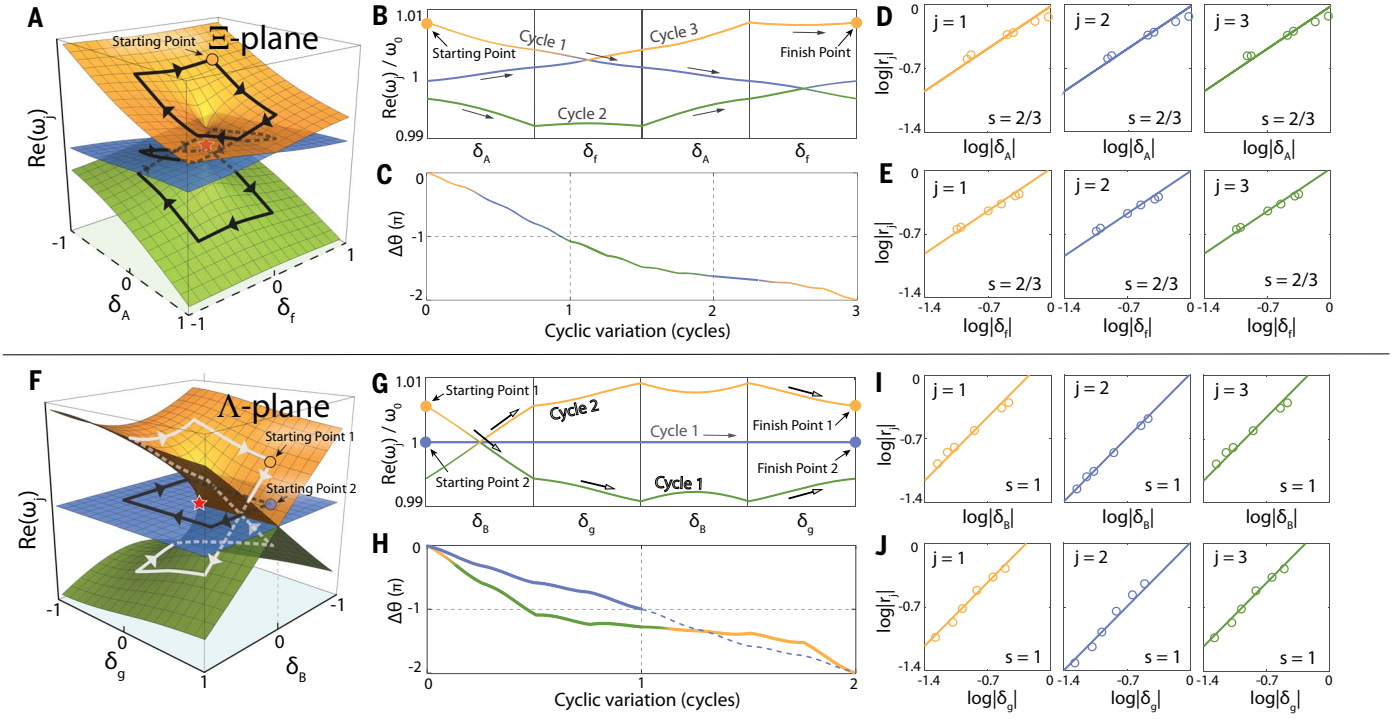


Fig. 3. HTI of the EX. (**A** to **C**) Encircling of the EX on the Ξ plane (with Λ = 0) shown on the real part of the ERSs. The path traverses all three states and returns to the starting point after three full cycles [(A) and (B)], accumulating a total Berry phase of Δθ = −2π (C). (**D** and **E**) The measured phase rigidities (markers) have a critical exponent of 2/3 when the EX is approached from δ_A and from δ_f, respectively. (**F** to **H**) Encircling of the EX on the Λ plane (with Ξ = 0) [(F) and (G)] yields a total Berry phase of Δθ = −2π (H). The middle state restores in only one cycle with Δθ = −π, and the second cycle repeats the first cycle, as shown by the dashed blue line. (**I** and **J**) The measured phase rigidities of the three eigenvectors, with a critical exponent of 1 for both δ_B and δ_g. The data in (D), (E), (I), and (J) are on a log-log scale with base 10. The curves in (B), (C), (G), and (H) are manually colored according to the ERSs in (A) and (F), respectively.

led to an entirely different winding process. The upper and lower states exchanged after one cycle without traversing the middle state and were fully restored upon two complete cycles (Fig. 3, F and G). The corresponding Berry phase is −2π after two cycles. The middle state is restored in one cycle with a Berry phase of −π. Hence, the winding number is $\mathcal{W}_\Lambda = -2/2 = -1$. Consequently, the EX possesses a combination of two distinct winding numbers in the two complex planes. We denote such a unique property as an HTI.

Next, we experimentally verified the existence of HTIs. Emerging through cyclic evolutions of the eigenstates, an HTI is also tied to the splitting of eigenfunctions near the EX (*16*, *24*); therefore, it can be characterized using phase rigidity, defined as

$$|r_j| = \left| \frac{\langle \tilde{\phi}_j^{\mathrm{L}} | \tilde{\phi}_j^{\mathrm{R}} \rangle}{\langle \tilde{\phi}_j^{\mathrm{R}} | \tilde{\phi}_j^{\mathrm{R}} \rangle} \right| \qquad (3)$$

where $\langle \tilde{\phi}_j^{\mathrm{L}} |, |\tilde{\phi}_j^{\mathrm{R}} \rangle$ are the normalized biorthogonal left and right eigenvectors, and the index $j = 1, 2, 3$ labels the three eigenvectors. At the EX (EP3), three eigenvectors are aligned so that $|r_j|$ vanishes. Deviation from the EX along a complex parameter results in the splitting of the eigenvectors, which is revealed by the phase rigidity's critical exponent s. Consequently, the EX's winding numbers are embedded within the phase rigidity, which can be experimentally obtained by measuring the acoustic field distribution in all three cavities (*22*).

We experimentally obtained the phase rigidity of all three states by approaching the EX along both Λ and Ξ. For each complex parameter, the EX was approached separately along the real and imaginary axes of the same parameter. The outcomes were four sets of phase rigidities probed by approaching from four directions. In Fig. 3, D, E, I, and J, where the data are shown on a log-log scale, linear dependence is seen for all states and all four parameters. In Fig. 3, D and E, the slopes for the δ_A and δ_f directions are both 2/3, indicating critical exponents of $s_\Xi = 2/3$, which conform well to the winding number $|\mathcal{W}_\Xi| = 2/3$. In the δ_B and δ_g directions, the values of $|r_j|$ follow a linear fit with a unity slope (Fig. 3, I and J), implying a critical exponent of $s_\Lambda = 1$, which agrees with $|\mathcal{W}_\Lambda| = 1$. Also, these experimental results in Fig. 3 align well with the two possible cases of phase rigidities for an EP3 (*25*), which are $s = (N-1)/N$ and $s = (N-1)/2$, wherein $N = 3$ is the EP's order. The mirror symmetry about Λ = −1 implies that an additional EX at (Ξ, Λ) = (0, −2) has an identical HTI. The system's chirality inverts when κ changes sign, producing another two EXs with an opposite HTI.

Our discovery of the HTI shows that a higher-order EP can possess rich topological properties beyond those found in Hermitian systems. By traversing both Λ and Ξ, even more winding possibilities can emerge. How to realize these new topological physics in momentum space and what phenomena they can generate are intriguing questions. Meanwhile, leveraging these singular behaviors of higher-order EPs may open new avenues for applications.

REFERENCES AND NOTES

1. N. P. Armitage, E. J. Mele, A. Vishwanath, *Rev. Mod. Phys.* **90**, 015001 (2018).
2. T. T. Heikkilä, G. E. Volovik, *New J. Phys.* **17**, 093019 (2015).
3. G. Chang *et al.*, *Sci. Rep.* **7**, 1688 (2017).
4. R. El-Ganainy *et al.*, *Nat. Phys.* **14**, 11–19 (2018).
5. L. Feng, R. El-Ganainy, L. Ge, *Nat. Photonics* **11**, 752–762 (2017).
6. M.-A. Miri, A. Alù, *Science* **363**, eaar7709 (2019).
7. W. D. Heiss, *J. Phys. A* **45**, 444016 (2012).
8. I. Rotter, *J. Phys. A* **42**, 153001 (2009).
9. Q. Zhong, M. Khajavikhan, D. N. Christodoulides, R. El-Ganainy, *Nat. Commun.* **9**, 4808 (2018).
10. B. Zhen *et al.*, *Nature* **525**, 354–358 (2015).
11. A. Cerjan *et al.*, *Nat. Photonics* **13**, 623–628 (2019).

12. X. Zhang, K. Ding, X. Zhou, J. Xu, D. Jin, *Phys. Rev. Lett.* **123**, 237202 (2019).
13. H. Zhou, J. Y. Lee, S. Liu, B. Zhen, *Optica* **6**, 190 (2019).
14. H. Zhou *et al*., *Science* **359**, 1009–1012 (2018).
15. C. Dembowski *et al*., *Phys. Rev. Lett.* **86**, 787–790 (2001).
16. K. Ding, G. Ma, M. Xiao, Z. Q. Zhang, C. T. Chan, *Phys. Rev. X* **6**, 021007 (2016).
17. H. Shen, B. Zhen, L. Fu, *Phys. Rev. Lett.* **120**, 146402 (2018).
18. K. Kawabata, K. Shiozaki, M. Ueda, M. Sato, *Phys. Rev. X* **9**, 041015 (2019).
19. K. Kawabata, T. Bessho, M. Sato, *Phys. Rev. Lett.* **123**, 066405 (2019).
20. H. Hodaei *et al*., *Nature* **548**, 187–191 (2017).
21. Z. Yang, A. P. Schnyder, J. Hu, C.-K. Chiu, arXiv 1912.02788 [cond-mat.mes-hall] (31 May 2020).
22. See supplementary materials.
23. A. A. Soluyanov, D. Vanderbilt, *Phys. Rev. B* **85**, 115415 (2012).
24. E. N. Bulgakov, I. Rotter, A. F. Sadreev, *Phys. Rev. E* **74**, 056204 (2006).
25. Y.-X. Xiao, Z.-Q. Zhang, Z. H. Hang, C. T. Chan, *Phys. Rev. B* **99**, 241403 (2019).

ACKNOWLEDGMENTS

G.M. thanks Ruoyang Zhang for discussions. **Funding:** Supported by Hong Kong Research Grants Council awards GRF 12300419, 12302420, 16303119, ECS 22302718, C6013-18G, and AoE/P-02/12; National Science Foundation of China (NSFC) grants 11922416 and 11802256; NSFC grant 11904055 and the CAST Young Elite Scientists Sponsorship Program (X.J.); the Gordon and Betty Moore Foundation (K.D.); and NSFC grant 11847205 and China Postdoctoral Science Foundation grant 2018M630597 (Y.-X.X.). **Author contributions:** K.D. and G.M. conceived the research; W.T., K.D., Z.-Q.Z., and G.M. performed theoretical analysis; W.T. and X.J. carried out numerical calculations and experiments; all authors analyzed and discussed the results and contributed to the manuscript; and G.M. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1077/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Reference (*26*)

17 July 2020; accepted 13 October 2020
10.1126/science.abd8872

3D PRINTING

Critical instability at moving keyhole tip generates porosity in laser melting

Cang Zhao[1,2]*, Niranjan D. Parab[3]†, Xuxiao Li[4], Kamel Fezzaa[3], Wenda Tan[4], Anthony D. Rollett[5,6]*, Tao Sun[7]*

Laser powder bed fusion is a dominant metal 3D printing technology. However, porosity defects remain a challenge for fatigue-sensitive applications. Some porosity is associated with deep and narrow vapor depressions called keyholes, which occur under high-power, low–scan speed laser melting conditions. High-speed x-ray imaging enables operando observation of the detailed formation process of pores in Ti-6Al-4V caused by a critical instability at the keyhole tip. We found that the boundary of the keyhole porosity regime in power-velocity space is sharp and smooth, varying only slightly between the bare plate and powder bed. The critical keyhole instability generates acoustic waves in the melt pool that provide additional yet vital driving force for the pores near the keyhole tip to move away from the keyhole and become trapped as defects.

Laser powder bed fusion (LPBF), currently the most extensively used metal additive manufacturing process, is revolutionizing the manufacturing industry (*1*–*3*). In a typical LPBF process, a high-power-density laser is scanned across a thin powder layer, under the guidance of a digital computer-aided design and drafting (CADD) model, to locally melt the metal powder and fuse it to the previous layer. Despite LPBF's unrivaled capability for directly fabricating parts with complex geometries, the associated porosity remains its Achilles heel because it impairs the performance—for example, corrosion resistance and fatigue life—of the product (*4*). Intensive research efforts have advanced the understanding and control of the porosity (*5*). Of the various mechanisms that cause porosity, the keyhole mode of melting has recently attracted broad attention because of its importance in laser welding (*6*) and most especially in LPBF, as revealed with high-speed synchrotron x-ray imaging (*7*).

In keyhole-mode melting, the strong recoil pressure from the rapid evaporation of metal pushes the surrounding melt liquid downward, creating a deep and narrow cavity called a keyhole. Inside the keyhole, the laser beam undergoes multiple reflections, which largely enhances laser absorption and improves energy efficiency (*7*–*10*). However, under some laser conditions, the keyhole walls constantly fluctuate and collapse. This process is generally attributed to the complex interplays among the thermocapillary force, Marangoni convection, recoil pressure, and appearance of gas plasma (*11*–*13*). Such keyhole instability generates pores, which become detrimental structural defects in the build if trapped by the solidification front.

Power-velocity (*P-V*) mapping (*14*–*16*), which directly relates product quality (density or porosity) to the process parameters (sometimes with hatch spacing), is an effective but heuristic tool for LPBF users. One of the missing pieces in the physical foundation of *P-V* maps is whether the so-called "keyhole porosity boundary," which separates processing conditions that produce nominally fully dense parts from those with pores, is well defined and reproducible. Moreover, in LPBF, the presence of powder has been linked to larger keyhole fluctuations and greater pore formation when compared with those of laser melting of bare plate (*7*). However, quantification of the boundary and any powder-induced shift is lacking. The relevant spatiotemporal scales of keyhole pore formation require subnanosecond temporal resolution, micrometer spatial resolution, and megahertz frame rates. These are precisely the domain of synchrotron x-ray imaging, allowing us to investigate keyhole pore formation.

We used operando high-speed x-ray imaging at a third-generation high-energy light source (*17*–*21*) for the completion of this missing piece in the *P-V* map by quantitatively outlining the keyhole porosity boundary, elucidating the powder effect, and revealing the fundamental origin of keyhole pores and their initial motions during the LPBF process. The sample was a thin Ti-6Al-4V plate (~400 μm in thickness) (fig. S1A) with, for certain experiments, a layer of powder on top and sandwiched together between two glassy carbon plates. A laser beam with a Gaussian profile and in continuous wave mode scanned the sample along a single straight line at various powers (*P*) and speeds (*V*). Meanwhile, we recorded full-field x-ray images at high spatial (2.0 to 3.0 μm/pixel) and temporal (0.1 to 7500 ns) resolutions at high frame rates (50 to 5000 kHz) (*22*).

The keyhole porosity boundary separating the keyhole porosity regime from the stable melting regime was unexpectedly smooth and sharp, with or without the presence of powder (Fig. 1A). Across the boundary toward higher energy density, discernable pores were released into the liquid from the keyhole bottom (Fig. 1B). The maximum pore size (among all observed in any given experiment) trended up with decreasing scan speed at a constant laser power. A few cases deviated from this trend, most likely because of variations in powder bed thickness in different samples and from the short time window available for x-ray imaging relative to the pore-formation events. With decreasing laser power, the keyhole porosity boundary becomes increasingly sensitive to scan speed, particularly in the presence of powder, and appeared to asymptote to a threshold in power.

Pore formation, regardless of laser power and the presence or absence of powder, occurred when the bottom tip of a letter "J"-shaped keyhole was pinched off, released at high speed toward the wake of the laser beam, and eventually trapped by the solidification front (Fig. 1B and figs. S3, A, C, and D, and S4). This process suggests a dominant mechanism for keyhole pore formation around the boundary, in which acoustic waves push the pinched-off pore(s) away from the keyhole. Meanwhile, the "role of powder" region, bounded by the two boundaries (with and without powder) (Fig. 1A), joins two subregions. In the high-power-velocity (high-*PV*) region, for a constant scan speed, lower power was needed for a powder bed sample to generate keyhole pores, whereas in the low-power-velocity (low-*PV*) region, higher power was needed. This interesting result implies a secondary keyhole pore formation mechanism (fig. S3, B and E). Near the low-*PV* boundary, the large and rapid fluctuation of a dramatically narrowed straight keyhole of high aspect ratio induces small pores directly beneath the laser beam, which resembles the stationary laser melting situation when the laser beam is turned off or shaded (*17*, *23*). This area in the *P-V* space also largely overlaps with the transition zone we defined (*7*), in which the keyhole is prone to fluctuate in the depth direction. Compared with bare plate samples, in which such pores tend to merge quickly into the subsequent growing keyhole and disappear, the larger keyhole fluctuations in powder bed samples create sufficient waiting time for the pores. These pores are nearly stationary after pinching off the keyhole and

[1]Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China. [2]Key Laboratory for Advanced Materials Processing Technology, Ministry of Education, Beijing 100084, China. [3]X-ray Science Division, Argonne National Laboratory, Lemont, IL 60439, USA. [4]Department of Mechanical Engineering, University of Utah, Salt Lake City, UT 84112, USA. [5]Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [6]NextManufacturing Center, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [7]Department of Materials Science and Engineering, University of Virginia, Charlottesville, VA 22904, USA.
*Corresponding author. Email: cangzhao@tsinghua.edu.cn (C.Z.); rollett@andrew.cmu.edu (A.D.R.); ts7qw@virginia.edu (T.S.)
†Present address: Intel Corporation, 2501 NE Century Boulevard, Hillsboro, OR 97124, USA.

suspended inside the melt pool (or potentially accelerated by the local melt flow through forces such as drag away from the keyhole) (*22*) to be pinned by the advancing solidification front.

The keyhole porosity boundary is closely related to the keyhole depth, d_l (and the melt pool depth, d_m) (fig. S2B). Typically, at a constant laser power, keyhole pores form only when the keyhole depth is larger than a threshold, d_c (fig. S6, A and G). This threshold increases, statistically, with scan speed (Fig. 2A) and laser power (Fig. 2B) in distinct power-law forms, $d_c \sim V^6$ and $d_c \sim P$, respectively. In general, the relationships between the keyhole depth (d_l) away from the boundary and the two processing parameters (P and V) are complicated and highly coupled, $d_l \sim (AP + B)/V$ ($A > 0$ and $B < 0$ are constants) (figs. S5 to S7) (*22*). The smooth and sharp keyhole porosity boundary may imply that the condition for keyhole pore formation is material-constant–dependent or keyhole-geometry–dependent. An examination of the keyhole depth threshold (Fig. 2G) reveals a simple linear relationship with the energy density (E) in the high-PV segment, $d_c \sim E$, with the slope being the same with or without powder (50 cm^3/MJ) and the difference between their vertical intercepts being ~100 μm (close to the powder bed thickness).

However, this linearity in the $d_c \sim E$ plot does not hold for the data in the low-PV range. An energy density threshold $E_c \sim 3.7$ MJ/m^2 appears, below which no keyhole pores form. This also implies the secondary mechanism for the keyhole pore formation in this region, which we attributed to the keyhole depth fluctuations.

The relative fluctuation R (the coefficient of variation of keyhole depth) increased strongly with decreasing scan speed and laser power in the low-PV range of the keyhole porosity boundary (Fig. 2, D and E). Especially in the presence of powder, $R > 1$. This means that the laser sometimes drills very deep, far beyond the average and across the sustainable depth, and leaves a pore at the keyhole bottom (movie S7). Additionally, the powder-induced relative fluctuation ΔR_p (the difference between presence versus absence of powder) changed dramatically with the laser condition (Fig. 2E, inset). This indicates that the powder motion has substantial influence on the keyhole pore generation. We speculate that powder particles occasionally enter the laser beam and shade the keyhole momentarily from full direct laser illumination before flying away (*24*). This abruptly but temporarily lowers the local temperature on the surface layer of the keyhole walls and the recoil pressure on the liquid below the surface layer and thereby promotes the keyhole depth fluctuation. Our observations (Fig. 2, C and F) collectively confirm the intrinsic connection between keyhole depth and keyhole porosity and complements the keyhole porosity boundary in the P-V space (Fig. 1A).

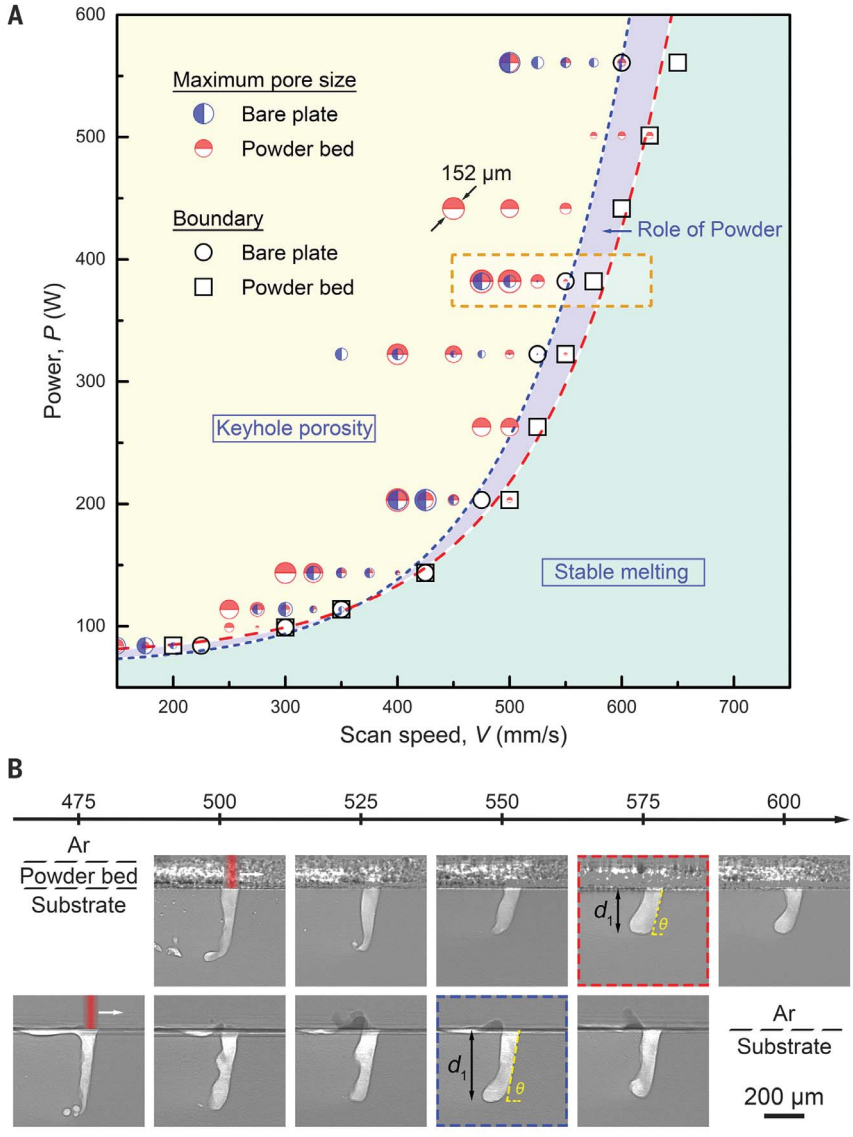


Fig. 1. Keyhole porosity boundary and role of powder in laser melting. (**A**) P-V space, which is divided into a stable melting regime (light green shaded area, consisting of stable keyhole, transition, and conduction regimes) and a keyhole porosity regime (light yellow area), indicated by open circles and a blue short-dashed line for the case of bare plate and by open squares and a red long-dashed line for samples with a ~105-μm powder layer on top. The extent to which adding powder increases the instability and widens the porosity regime is indicated by the light purple shaded regions and a blue arrow identifying the role of powder. In the keyhole porosity regime, the diameters of the blue and red half solid circles indicate the measured maximum equivalent pore size for each specific P-V combination. The dashed lines are to guide the eye. (**B**) Tableau of representative x-ray images at constant P = 382 W and varying scan speed across the keyhole porosity boundary indicated by the orange dashed rectangle in (A). (Top) Powder bed samples. (Bottom) Bare plate samples. The red and blue dashed framed images correspond to the respective boundaries in scan speed, below which pores are trapped in the resolidified metal. The overall keyhole depth, d_l, and front keyhole wall angle, θ, are labeled. All images were background corrected by the images collected before laser melting. The laser spot size is ~100 μm.

During a laser scan, most of the incident beam impinges on the front keyhole wall (*8*, *9*). In the stable keyhole regime, the projection of the front keyhole wall on the sample surface is the effective laser spot D_0. In previous studies, D_0 was constant across the P-V space, regressed from the nearly linear d_l versus tanθ relationship, where θ is the front keyhole wall angle (*6*, *7*). This, however, fails when the

P-V combination crosses the keyhole porosity boundary (Fig. 2H). As the laser scan speed decreased, the projected length of the front keyhole wall *s* (defined as $d_1/\tan\theta$, along the laser scan direction) fluctuated at first around D_0 and then decreased with *V*. Two possible reasons could explain this trend. The Gaussian profile of the laser concentrates the intensity at the center, and the edge of the beam at low power does not provide sufficient energy to vaporize the metal. The low scan speed tends to increase the front keyhole wall angle (beyond a threshold called Brewster's angle), largely reducing the laser absorption of the low-power Gaussian beam and further cutting off the edge of the front keyhole wall (*22*).

The depth-to-length aspect ratio of a keyhole appears to be a good indicator for the keyhole pore formation (Fig. 1B) as well as the maximum keyhole pore size (Fig. 1A). Regardless of the broadening of the keyhole

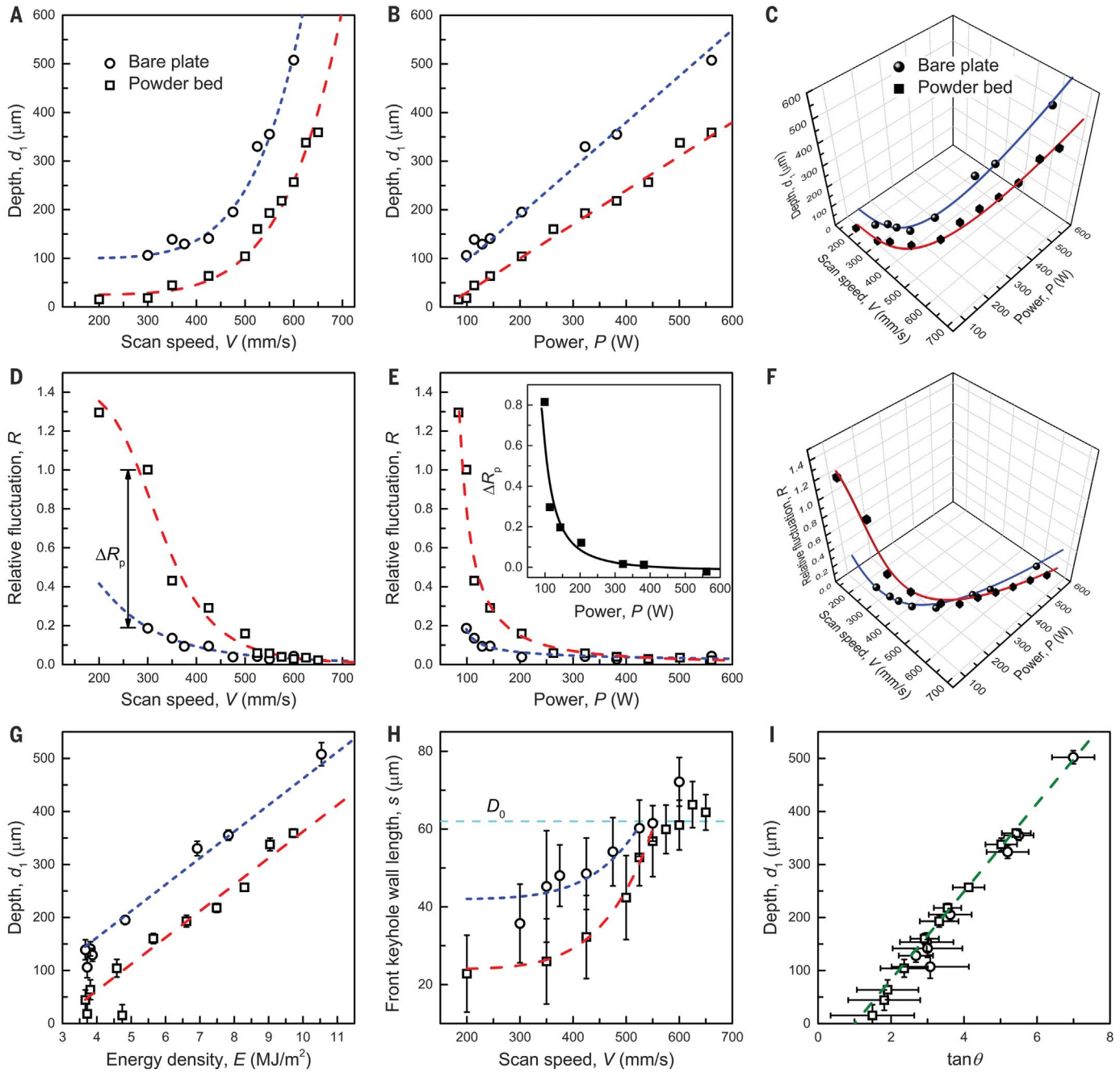


Fig. 2. Keyhole scaling and fluctuation on the porosity boundary. (**A** and **B**) Keyhole depth, d_1, as a function of (A) scan speed *V* and (B) laser power *P*. (**C**) Keyhole depth in the *P-V* space. (**D** and **E**) Relative fluctuation of keyhole depth, $R = \Delta d_1/d_1$, as a function of (D) scan speed and (E) laser power. (**F**) Relative fluctuation of keyhole depth in the *P-V* space. (**G**) Keyhole depth as a function of energy density *E*. (**H**) Front keyhole wall length, *s*, as a function of scan speed. The length is defined as $d_1/\tan\theta$, where θ is the front keyhole wall angle. The horizontal dashed line indicates the effective laser spot size D_0 in the stable keyhole regime. (**I**) Keyhole depth as a function of tangent of front keyhole wall angle. The green dashed line in (I) is linearly regressed through all data points. The circles with blue lines are from bare plate samples, and the squares with red lines are powder bed samples. The error bar of each datum point represents the standard deviation.

by multiple reflections and Fresnel absorption (fig. S7E), the tangent of front keyhole wall angle is closely related to the aspect ratio (fig. S7F). On the keyhole porosity boundary, the keyhole depth threshold clearly follows a linear relationship with tanθ (Fig. 2I). Data from both bare plate and powder bed samples fall on the same line as $d_c = \alpha(\tan\theta - 1)$, where α is ~83 μm, which is evidently larger than the effective laser spot D_0 (~62 μm). We call α the keyhole pore constant. Another characteristic of our plot is that the horizontal intercept of the line is around tanθ = 1, which suggests that keyhole pores are only generated when the keyhole depth is larger than its length on the boundary. The linearity in the relationship (Fig. 2, B and I) is well maintained over the entire keyhole porosity boundary, despite that the mechanisms for generating a pore at high- and low-*PV* segments could be different. This suggests a deep connection between the dominant and secondary mechanisms.

A dominant mechanism exists for pore formation near the keyhole porosity boundary in which instability forms pores and acoustic waves drive them away from the keyhole tip, regardless of laser power and the presence or absence of powder. We used megahertz x-ray imaging to measure the keyhole and pore evolutions with high spatial and temporal resolutions (Fig. 3A; figs. S8, S13, S17, and S19; and movies S11 and S12). In general, four phases of the process can be defined: (i) formation of a mini-keyhole, (ii) formation of an instant keyhole pore (P_0), (iii) formation of a needle-like keyhole bottom (NKB), and (iv) rapid pore motion away from the keyhole and pinning.

The sequence of events leading to the formation of a keyhole pore starts from a letter "J"-shaped keyhole and a mini-keyhole on top of a protrusion on the front keyhole rim (Fig. 3A) (*9*, *25*). The mini-keyhole creates a strong recoil on the side surface of the protrusion, driving it to swing down and impact on the rim of the rear keyhole wall. During this process, the keyhole wall underneath the protrusion, particularly the region close to the front wall, receives much-reduced laser energy. This results in a drop of the local recoil pressure, the subsequent rapid liquid filling driven by the surface tension (*9*, *26*), and eventually, the impact of the liquid on the side surface of the rear wall. The collapse of this keyhole region occurred at a horizontal speed of ~13 m/s (fig. S9) (*22*), giving rise to a "newborn" keyhole, a pore P_0, and a high-speed microjet (fig. S10).

Pore P_0 then undergoes an asymmetric collapse (Fig. 3, B to D and F, and figs. S8 and S10). In the first few microseconds after pore P_0 pinches off the keyhole, the original keyhole tip is almost stationary. This phenomenon occurs because the high overheating on the free side of the surface layer (approximately hundreds of kelvin above the boiling point) and the limited thermal diffusion of the heat toward the surrounding liquid (a few micrometers in a few microseconds) maintain the pressure balance through continuous vaporization for a short period. By contrast, the top-right edge of pore P_0 moves very fast, driven by the rapidly piercing microjet (Fig. 3D and figs. S10 and S15C). When the microjet tip P_{jet} impinges on the opposite wall of the pore, the impact speed is ~16 m/s, and we estimated the "water hammer" pressure (*27*, *28*) to be ~140 MPa (*22*). This process results in rapid pore collapse, splitting, and rebound (Fig. 3F, fig. S11, and movie S11) and an acoustic wave (*29*, *30*). This impact mechanism is analogous to the shock wave–pore interaction in liquids by laser or ultrasound (*29*, *31*).

Unlike the overall keyhole depth d_1, the mini- and newborn keyhole depth d_2 surprisingly exhibited a repetitive pattern (Fig. 3E). The process begins from a protrusion on the front keyhole rim, grows to the overall keyhole depth, and ends with the creation of a successor keyhole. With the keyhole collapse time t_1 being the reference, the shifted keyhole depth had a global drill rate of ~17 m/s before entering the plateau stage (fig. S15A), which allowed us to estimate the average temperature of the melt at the keyhole bottom as 3900 K (*22*).

Despite this creation process, the keyhole pore survives only under a certain set of conditions. A protrusion structure may exist at the bottom of the front keyhole wall (for example, Fig. 3A, time 8.28 μs, and fig. S14, 47.84 μs), which caused the reduction of the recoil pressure applied to the liquid beneath the keyhole tip surface. This keyhole tip was reshaped into a narrow NKB (for example, Fig. 3J, time 9.20 μs, and Fig. 3K, 48.76 μs) by the acoustic wave induced by the pore rebound, with facilitation by the surface tension effect (Fig. 3H). We found evidence of this from the directional reshaping, mainly on the side facing the pore (Fig. 3K), and from the nearly synchronized response of the NKB to the pore. Because of these two aforementioned effects, the newborn keyhole depth d_2 experiences a rapid drop at t_2 (Fig. 3E). The NKB emits another strong acoustic wave (Fig. 3I), driving the pore to migrate away, before which the pore is almost stationary (for example, Fig. 3J, pore P_1).

A keyhole with a needle-shaped bottom is unstable. So once the protrusion disappears (Fig. 3J and fig. S12) or the laser drills through (Fig. 3K and fig. S14), the keyhole can then focus the laser intensity inside the NKB through multiple reflections (fig. S15D) (*22*), creating the conditions for rapid drilling and keyhole expansion. An acoustic wave was generated after the formation of the NKB, as shown by the almost instantaneous and synchronized responses of pores P_1 and P_2 (Fig. 3I and movie S11); the pore P_1 was accelerated from nearly stationary to ~10 m/s in less than 1 μs and then decelerated (fig. S16); and the initial deformation and even collapse of the pore was largely on the side facing the NKB. We found traces of microjet penetration, whereas the far side was unperturbed (Fig. 3, J and K, and figs. S11 and S14).

We speculate that the explosive behavior of the NKB was due to phase explosion. This defines an extreme material behavior caused by rapid heating; when the local temperature approaches the critical point, a metastable liquid decomposes instantly into a mixture of gas and fine droplets (*32*–*34*). In addition to the strongly enhanced laser absorption on the NKB walls, several observations support this idea (Fig. 3, J and K, and figs. S12, S14, and S22). (i) The keyhole bottom expansion starts roughly from the inside of the NKB, for which we found evidence from features around the center of the expanded bottom. (ii) The expanding keyhole walls were roughened and wrinkled, which suggested an impact by the gas-droplet mixture. (iii) The primary pore may collapse first, before the expansion (not widening). The explosion of the surface-layer liquid at the keyhole bottom emits an acoustic wave directly into the liquid below, which travels at the sound speed (C). The wave amplitude was indicated by the rapid pore collapse (for example, through microjetting), with the collapse speed of ~18 m/s (Fig. 3K and fig. S14) and the water hammer pressure of ~150 MPa (*22*). In the wake of the wave, the liquid pressure dropped, causing a sudden widening of the keyhole bottom (Fig. 3K, time 49.68 μs, and fig. S14).

Limited by the short interaction time (D_p/C ~ a few nanoseconds, where D_p is the equivalent pore diameter), the pore can gain sufficient kinetic energy and escape from the large thermal gradient field around the keyhole only when the amplitude of the acoustic wave from the NKB is high enough. The amplitude appears higher along the depth direction because the pores tend to migrate downward first. In addition to the acoustic wave and the inertial force of the liquid jet it causes (*35*), the translational motion of a pore around the keyhole bottom is collectively controlled by the thermocapillary force (*36*) and the viscous drag force (*35*). These are determined by the local thermal gradient and the Marangoni flow, respectively. The thermocapillary force always attracts the pore to the keyhole, whereas the viscous drag force, because the local melt tends to flow away from the keyhole, does so only when the pore moves faster than the local melt along the same direction. This confirms the importance of the acoustic wave in porosity formation in combination with keyhole instability [more discussion is provided in (*22*), Generation and role of acoustic waves].

With insufficient kinetic energy from the acoustic wave, the pore remains in the region

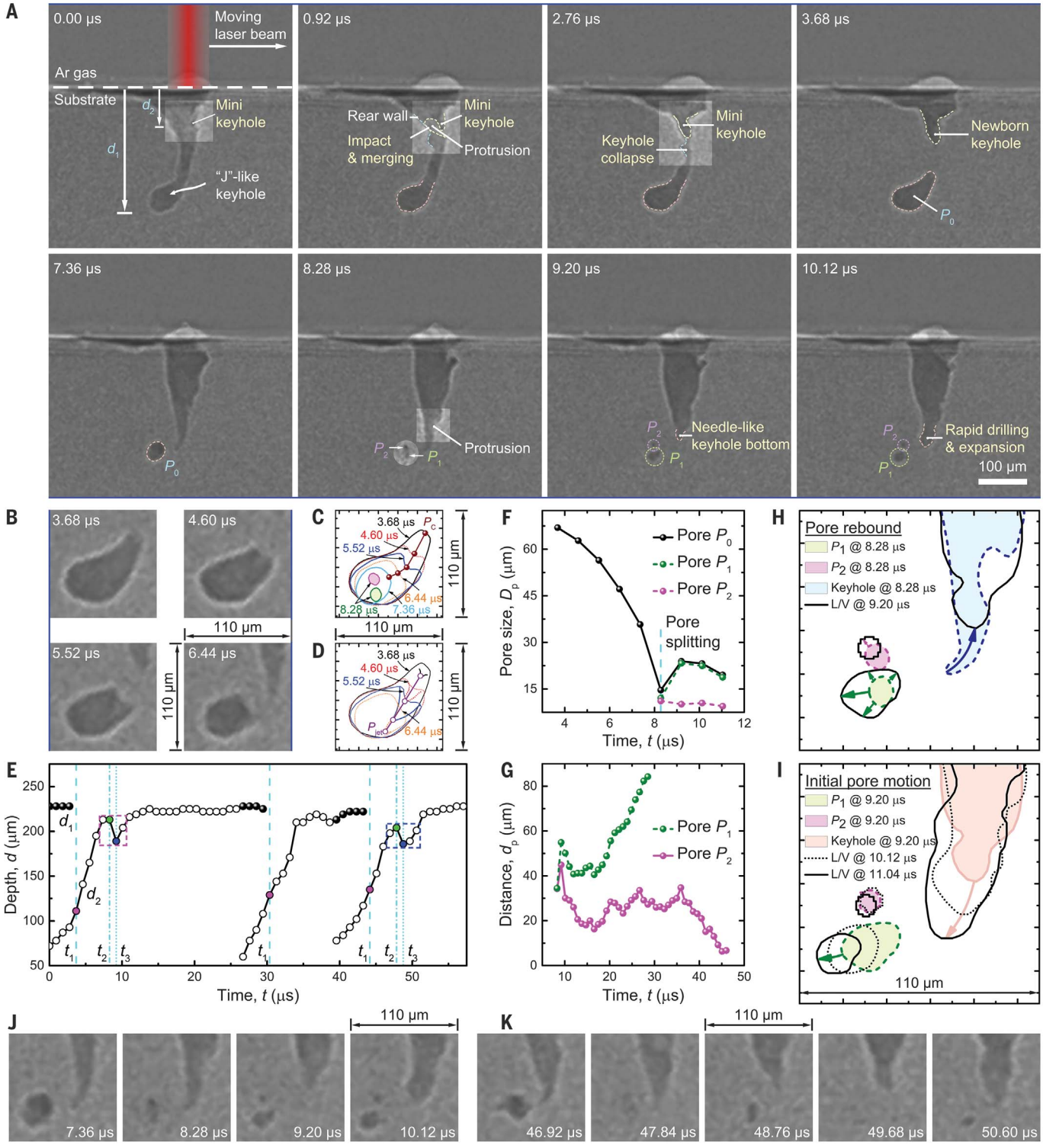


Fig. 3. Keyhole pore formation and motion driven by acoustic waves from keyhole instability. (**A**) Megahertz x-ray images of a keyhole pore-formation process. (**B**) X-ray images showing the nonuniform collapse of pore P_0, driven by a microjet. (**C** and **D**) Contours of the pore P_0 and microjet morphologies. (**E**) Keyhole depths, d_1 and d_2. Three time nodes are defined: At t_1 (magenta circle), with the keyhole closure, pore P_0 forms; at t_2 (green circle), the primary pore starts to rebound; and at t_3 (blue circle), the keyhole depth reaches a local minimum, and the pores start to migrate away from the keyhole. (**F**) Equivalent pore diameter, D_p, estimated from (C). (**G**) Distances of pores away from the nearest keyhole wall. (**H**) Formation of a NKB, attributed to the coupling between the existence of a protrusion and the rebound of pore P_1. (**I**) Initial pore motions caused by the acoustic wave emitted from the NKB. L-V in (H) and (I) is the liquid-vapor interface of a keyhole or a pore. (**J** and **K**) X-ray images of keyhole pore collapse, rebound, and motion, corresponding to the two events indicated with the magenta and blue dashed rectangles in (E), respectively. All images were background corrected followed by contrast reversal. Frame-by-frame images are documented in movie S11 and fig. S8.

where there is a large thermal gradient field, and its motion is dominated by the competition between the thermocapillary and viscous drag forces. After the postulated phase explosion, pore P_2, unlike pore P_1, was not pushed away but "dances" with the keyhole bottom wall, until an abrupt morphology change of the keyhole bottom (because of the pressure drop inside) (fig. S18) successfully lures the pore in deep (Fig. 3G and fig. S17). The keyhole then grows rapidly owing to the increased laser absorption inside the forcedly developed bottom and eventually captures pore P_2. The evolving distances of these two pores from the nearest keyhole wall (Fig. 3G) suggests two thresholds (plateaus in the plots) under this particular laser condition. One is the edge of the strong thermal gradient field, 42 ± 2 μm from the keyhole wall, beyond which the thermocapillary force is negligible. The other is the point at which the two forces balance, 28 ± 3 μm, below which the thermocapillary force dominates because of the high thermal gradient of ~10^8 K/m (*22*) and draws the pore violently into the keyhole.

The thermocapillary-viscous drag force boundary remains nearly constant as the keyhole bottom evolves. When the keyhole drilling reaches its limit, the pore may become nearly stationary; for example, pore P_2 stops moving since time 18.40 μs for ~8 μs (movie S11). The force balance could be then used to estimate the drag coefficient C_d (*35*, *36*) [~6.0 in this case (*22*)], which has long been a controversial problem (*35*, *37*). However, when the keyhole shrinks and expands, the pore dances along. The sharp and narrow thermocapillary-viscous drag force boundary implies that if the pore is around the boundary, it always manages to accommodate its motion to the local melt flow pattern. But if trapped inside, the pore is unable to escape because of the increasing attraction from the keyhole wall, as hinted at by the rapid decrease in the distance away from the nearest keyhole wall (Fig. 3G).

With operando high-speed synchrotron x-ray imaging, we found that the boundary between the keyhole porosity and stable melting regimes in the *P-V* space was unexpectedly smooth and sharp, varying only slightly with the presence of powder. Close to this porosity boundary, we discovered a new mechanism that is responsible for the formation of small and often spherical pores: Acoustic waves generated during the highly dynamic laser-metal interaction play a key role in driving pores near the keyhole tip far enough away from the large thermal gradient field around the keyhole that they are trapped through solidification. A more detailed investigation of the mechanisms of acoustic waves is merited. In addition to this dominant mechanism, at low power levels a secondary mechanism results in pore trapping as a fluctuating keyhole collapses and retracts; drag forces may also pull the pore away from the retracting keyhole. At a broader scale, not only does the existence of a well-defined keyhole porosity boundary offer a more secure basis for predicting process windows, but the new understanding of how small pores are generated close to the boundary explains some of the confusion in the LPBF literature about the origins of porosity. In other words, small spherical pores may be inherited from the powder, or they may originate from the use of marginally stable melting conditions.

REFERENCES AND NOTES

1. S. A. Khairallah, A. T. Anderson, A. Rubenchik, W. E. King, *Acta Mater.* **108**, 36–45 (2016).
2. W. J. Sames, F. List, S. Pannala, R. R. Dehoff, S. S. Babu, *Int. Mater. Rev.* **61**, 315–360 (2016).
3. J. H. Martin *et al*., *Nature* **549**, 365–369 (2017).
4. R. Cunningham, S. P. Narra, C. Montgomery, J. Beuth, A. D. Rollett, *JOM* **69**, 479–484 (2017).
5. T. DebRoy *et al*., *Prog. Mater. Sci.* **92**, 112–224 (2018).
6. R. Fabbro, *J. Phys. D Appl. Phys.* **43**, 445501 (2010).
7. R. Cunningham *et al*., *Science* **363**, 849–852 (2019).
8. A. Kaplan, *J. Phys. D Appl. Phys.* **27**, 1805–1814 (1994).
9. C. Zhao *et al*., *Phys. Rev. X* **9**, 021052 (2019).
10. W. E. King *et al*., *J. Mater. Process. Technol.* **214**, 2915–2925 (2014).
11. A. Matsunawa, J.-D. Kim, N. Seto, M. Mizutani, S. Katayama, *J. Laser Appl.* **10**, 247–254 (1998).
12. N. Seto, S. Katayama, A. Matsunawa, *J. Laser Appl.* **12**, 245–250 (2000).
13. A. A. Martin *et al*., *Nat. Commun.* **10**, 1987 (2019).
14. J. C. Ion, H. R. Shercliff, M. F. Ashby, *Acta Metall. Mater.* **40**, 1539–1551 (1992).
15. J. L. Beuth *et al*., in *Solid Freeform Fabrication Proceedings*, (Univ. Texas at Austin, 2013), pp. 655–665.
16. H. Gong, K. Rafi, H. Gu, T. Starr, B. Stucker, *Additive Manufacturing* **1-4**, 87–98 (2014).
17. C. Zhao *et al*., *Sci. Rep.* **7**, 3602 (2017).
18. N. D. Parab *et al*., *J. Synchrotron Radiat.* **25**, 1467–1477 (2018).
19. N. P. Calta *et al*., *Rev. Sci. Instrum.* **89**, 055101 (2018).
20. C. L. A. Leung *et al*., *Nat. Commun.* **9**, 1355 (2018).
21. A. A. Martin *et al*., *Mat, Today Adv,* **1**, 100002 (2019).
22. Materials and methods are available as supplementary materials.
23. J. Y. Lee, S. H. Ko, D. F. Farson, C. D. Yoo, *J. Phys. D Appl. Phys.* **35**, 1570–1576 (2002).
24. S. A. Khairallah *et al*., *Science* **368**, 660–665 (2020).
25. R. Fabbro *et al*., *J. Laser Appl.* **30**, 032410 (2018).
26. M. Bayat *et al*., *Additive Manufacturing* **30**, 100835 (2019).
27. J. Dear, J. Field, A. J. Walton, *Nature* **332**, 505–508 (1988).
28. M. B. Lesser, J. E. Field, *Annu. Rev. Fluid Mech.* **15**, 97–122 (1983).
29. G. N. Sankin, W. N. Simmons, S. L. Zhu, P. Zhong, *Phys. Rev. Lett.* **95**, 034501 (2005).
30. S. Fujikawa, T. Akamatsu, *J. Fluid Mech.* **97**, 481–512 (2006).
31. C. J. Todaro *et al*., *Nat. Commun.* **11**, 142 (2020).
32. A. Miotello, R. Kelly, *Appl. Phys., A Mater. Sci. Process.* **69**, S67–S73 (1999).
33. P. Lorazo, L. J. Lewis, M. Meunier, *Phys. Rev. Lett.* **91**, 225502 (2003).
34. J. C. Ramirez-San-Juan *et al*., *Opt. Express* **18**, 8735–8742 (2010).
35. T. Watanabe, Y. Kukita, *Phys. Fluids A Fluid Dyn.* **5**, 2682–2688 (1993).
36. N. O. Young, J. S. Goldstein, M. J. Block, *J. Fluid Mech.* **6**, 350–356 (2006).
37. J. Magnaudet, I. Eames, *Annu. Rev. Fluid Mech.* **32**, 659–708 (2000).

ACKNOWLEDGMENTS

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357. We acknowledge A. Deriy and F. D. Carlo at the APS and Z. Guo and W. Chen at Purdue University for their technical support of the beamline experiments. We also acknowledge R. Cunningham at Messer Americas, W. King at The Barnes Group Advisors, and D. Du at Tsinghua University for the fruitful discussions. C.Z., T.S., and N.D.P. thank Argonne National Laboratory for supporting efforts before their relocations on developing beamline capabilities and the operando laser system that eventually led to this work. C.Z. thanks D. Du, F. Lin, G. Liu, and X. Liu at Tsinghua University for the financial support and assistance during the coronavirus disease 2019 (COVID-19) crisis. **Funding:** The work was supported by the start-up fund from the Department of Mechanical Engineering at Tsinghua University, the start-up fund from the University of Virginia, the Department of Defense Office of Economic Adjustment under award ST1605-17-02, the National Aeronautics and Space Administration (NASA) University Leadership Initiative program under grant 80NSSC19M0123, and the National Science Foundation under grant 1752218. **Author contributions:** T.S., A.D.R., and C.Z. conceived the idea and designed the project. C.Z., T.S., and N.D.P. designed and performed the experiments with assistance from K.F. and A.D.R.; C.Z., T.S., K.F., and N.D.P. designed and built the LPBF simulator. C.Z. carried out the image processing and data analysis and interpretation and proposed and proved the acoustic wave mechanism, with help from all authors. W.T. and X.L. conducted simulations, with input from C.Z. The manuscript was written by C.Z., T.S., and A.D.R. All authors commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate our conclusions are provided in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1080/suppl/DC1
Materials and Methods
Figs. S1 to S22
References (*38*–*52*)
Movies S1 to S12

4 June 2020; accepted 20 October 2020
10.1126/science.abd1587

CONSERVATION GENETICS

Individual heterozygosity predicts translocation success in threatened desert tortoises

Peter A. Scott[1,2]*, Linda J. Allison[3], Kimberleigh J. Field[3], Roy C. Averill-Murray[3], H. Bradley Shaffer[1,4]

Anthropogenic environmental modification is placing as many as 1 million species at risk of extinction. One management action for reducing extinction risk is translocation of individuals to locations from which they have disappeared or to new locations where biologists hypothesize they have a good chance of surviving. To maximize this survival probability, the standard practice is to move animals from the closest possible populations that contain presumably related individuals. In an empirical test of this conventional wisdom, we analyzed a genomic dataset for 166 translocated desert tortoises (*Gopherus agassizii*) that either survived or died over a period of two decades. We used genomic data to infer the geographic origin of translocated tortoises and found that individual heterozygosity predicted tortoise survival, whereas translocation distance or geographic unit of origin did not. Our results suggest a relatively simple indicator of the likelihood of a translocated individual's survival: heterozygosity.

In a world of rapid environmental change, habitat loss, and species endangerment, translocation of individual plants and animals is becoming increasingly common as a conservation strategy of last resort. For both animals and plants, the long-term success of translocations is often dismal (*1*–*3*). There are many possible explanations for this low success, and conservation biologists have proposed several guidelines for improving translocation outcomes, including limiting translocation distances and only exchanging individuals from within the same genetic units to minimize outbreeding depression (*4*). Although levels of inbreeding and heterozygosity have long been linked to individual survival and fitness (*5*–*7*), the relationship between these and translocation success has received little attention. Here, we exploited a long-term dataset for threatened Mojave desert tortoises to generate genomic profiles for 166 tortoises and link those data to individual post-translocation survival.

Mojave desert tortoises (*Gopherus agassizii*) are widely distributed members of the Mojave and Sonoran desert ecological communities west and north of the Colorado River in California, Nevada, Utah, and Arizona, USA (Fig. 1). Although the tortoise is a ubiquitous member of the relatively intact desert ecosystem, decreasing population trends led to the early listing of the species as threatened under the U.S. Endangered Species Act (*8*). Part of the species' recovery plan includes translocating tortoises salvaged from harmful anthropogenic activity and habitat destruction to new sites to augment declining populations (*9*, *10*).

Since the establishment of a recovery plan (*9*), genetic (*11*, *12*) and genomic (*4*, *13*) studies have quantified native population structure within Mojave desert tortoises and have consistently shown that the greatest axis of variation separates the Upper Virgin River and Northeastern Mojave recovery units (hereafter "northern Mojave") from the rest of the species' distribution (hereafter "southern Mojave"; Fig. 2). Additional fine-scale population structure has been documented within these two regions, leading some researchers to recommend translocations only within these genetically defined populations (*4*, *11*, *14*, *15*). Others have recommended limiting translocations to specified distances (200 to 276 km) on the basis of spatially distributed genetic structure (*16*). As is generally the case, these recommendations have assumed that moving animals within, but not between, genetic units (major clades or genetically related metapopulations) should be the guiding principle.

Following the accumulation of hundreds of displaced tortoises at the Desert Tortoise Conservation Center (DTCC) in 1996, the 100-km^2 Large-Scale Translocation Site (LSTS) was established. The LSTS is located in the Ivanpah Valley near Jean, Nevada, within the natural range of the tortoise, and is surrounded by either a tortoise-barrier fence or relatively inhospitable mountains (Fig. 1 and fig. S1). Because the majority of the tortoises received at the DTCC were captives (many from Nevada's free pet tortoise pickup program), most individuals lacked reliable information on their native site of origin. Between 1997 and 2014, ~9105 tortoises (~50.2% of which were adults) of unknown provenance were translocated to the LSTS, where they intermingled with an estimated 1450 adult local tortoises that were natural residents at the site (*17*). Most native and translocated tortoises in the LSTS have since died, consistent with steep declines in neighboring populations and likely furthered by high post-translocation densities and less comprehensive health screening during the first decade of the translocation program. However, roughly 350 adults were estimated by line-distance surveys to be alive in 2015 (*18*).

By 2016, there were three classes of LSTS tortoises: known-living and known-dead translocated individuals, and unmarked individuals presumed to be pre-translocation residents. For simplicity we refer to these as living, dead, and resident, respectively. Because no information is available on the origins of translocated tortoises, we generated restriction site–associated DNA sequencing (RADseq) genomic data and used these data to infer the geographic origins of a set of living and dead tortoises by comparing them to 270 low-coverage Mojave desert tortoise genomes that were field-collected from across the species' range (*13*) (Fig. 1).

By comparing living and dead tortoises from the LSTS, we explicitly address three questions central to assisted migration and genetic rescue efforts: (i) Do tortoises from more distant localities have lower survival fitness than those from nearby sites of origin? (ii) Do within–genetic unit (northern or southern Mojave) translocations enjoy greater survival than those that cross this primary genetic boundary? (iii) Are tortoises with higher overall heterozygosity, measured at deeply sequenced RAD loci, more likely to survive than less genetically variable individuals?

Mapped RADseq reads from the LSTS tortoises contained 6,711,580 of the 36,138,619 single-nucleotide polymorphisms (SNPs) found among 270 low-coverage Mojave desert tortoise genomes (*13*). We empirically evaluated several approaches to infer the place of origin of LSTS tortoises (*18*). Placing 12 known-origin calibration samples at the location of their genetic nearest neighbors resulted in a mean error of 61.7 km (SD = 60.2) from their true origin. A multi-individual, centroid-based placement approach using the eight closest genetic relatives reduced the mean placement error to 41.7 km (SD = 25.0). Finally, the optimal combined approach resulted when individuals with heterozygosity (π) values of <0.0020 were placed with their closest genomic match (presumably their closest relative) and individuals with π values of >0.0020 were placed at the centroid of their closest eight genetic relatives. This combined approach resulted in a mean error of 35.6 km (SD = 27.7). The combined method is thus more accurate, but that accuracy may result from overfitting a complex model with only 12 calibration animals. Countering this concern, we note that 87% of all LSTS tortoises had π values of <0.0020 and hence were geolocated only on the basis of their genetic nearest neighbors; moreover, given the very

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA. [2]Department of Life, Earth, and Environmental Sciences, West Texas A&M University, Canyon, TX 79016, USA. [3]U.S. Fish and Wildlife Service, Desert Tortoise Recovery Office, Reno, NV 89502, USA. [4]La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA.
*Corresponding author. Email: pete.a.scott@gmail.com

low coverage (~1.5×) of our dataset of 270 reference tortoises, we expected that more heterozygous LSTS tortoises would be difficult to associate with a true closest relative because heterozygosity is underestimated at ~1.5× coverage. Thus, averaging across a set of close matches should outperform a single match for those relatively heterozygous individuals. Because the dual method had the lowest combined error for the 12 calibration samples, we used it to determine the geographic and genetic provenance of all LSTS tortoises.

We calculated probable geographic origins for 166 living and dead LSTS tortoises that were matched for release year and sex (Fig. 2). Sixty-eight of the 79 living tortoises and 78 of the 87 dead tortoises were genomically placed in the geographically proximal northern Mojave genetic unit, and the remainder in the more distant southern Mojave unit (*13*). We found no difference in the proportion of northern versus southern Mojave desert tortoises that died or survived after translocation [χ^2 (df = 1, N = 166) = 1.18, P = 0.28]. We also detected no effect of geographic distance between the site of origin and the LSTS for individuals that died or survived after translocation (Fig. 3B; P = 0.83).

In contrast, we found that LSTS-translocated survivors had much higher individual heterozygosity when they were compared to those that died (Fig. 3A; mean π of living tortoises = 0.00180, mean π of dead tortoises = 0.00146; P = 0.00000005), indicating that individual genetic diversity predicted translocation success after accounting for release year and sex. The mean heterozygosity of the survivors was 23.09% greater than that of a matched set of tortoises that died over the same period. Although the importance of genetic diversity (or its presumed proxy, population size) of stock populations for translocation has been the subject of a few recent studies, this small body of work has yielded contradictory results on the role of population-level variation in translocation success, with some evidence for negligible importance (*19*–*21*) and other evidence for substantial importance (*22*, *23*). This discrepancy may stem from the assumption that population-level diversity is an accurate proxy for individual heterozygosity, which has not been tested. Our results demonstrate that individual heterozygosity, rather than population size or overall diversity, is a key, easily measured metric for predicting translocation success.

To explore the possibility that our observed relationship between post-translocation survival and heterozygosity is an artifact of sample age or condition, we confirmed that read depth, sample age, influence of outlier loci, and library complexity are not correlated with heterozygosity (*18*). Hence, we conclude that heterozygosity itself is a strong indicator of post-translocation

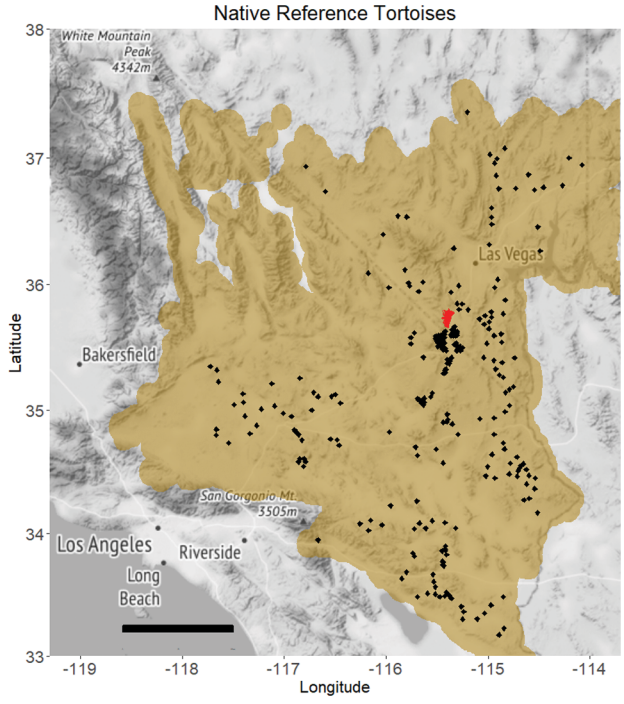


Fig. 1. Map showing the approximate historical distribution of the Mojave desert tortoise (*Gopherus agassizii*). The area in tan represents the distribution; black diamonds denote locations of 270 native low-coverage genome samples. The LSTS is shown as a red polygon. We lack samples for inference only from the northwest portion of the historical distribution for the species. Scale bar, 100 km.

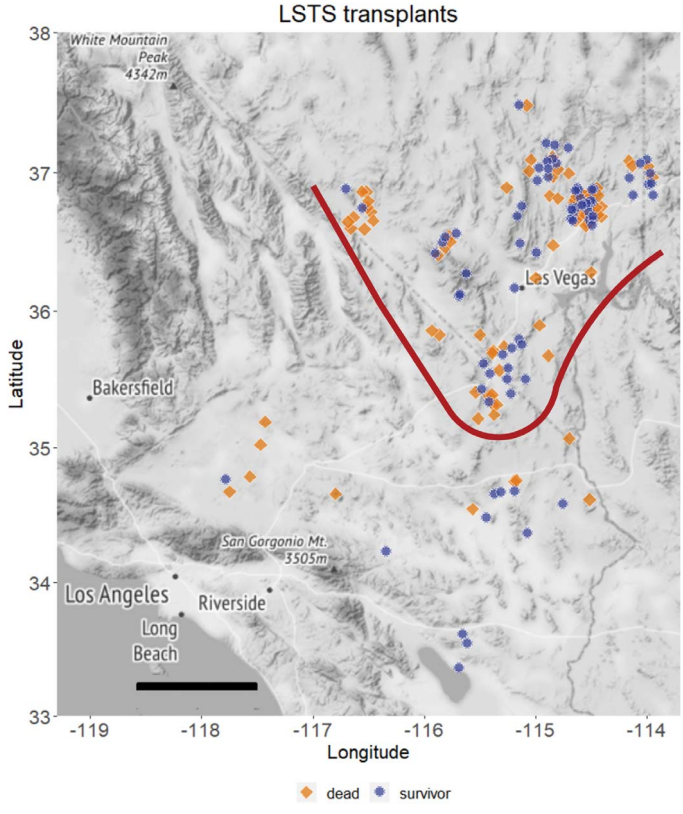


Fig. 2. Map showing the inferred origins of 79 translocation survivors and 87 sex- and release year–matched translocated dead tortoises. Points are slightly jittered for visual clarity. The red line shows the boundary between northern and southern Mojave tortoise genetic units. The inferred points of origin of tortoises that died and survived are shown as orange and blue diamonds, respectively. Scale bar, 100 km.

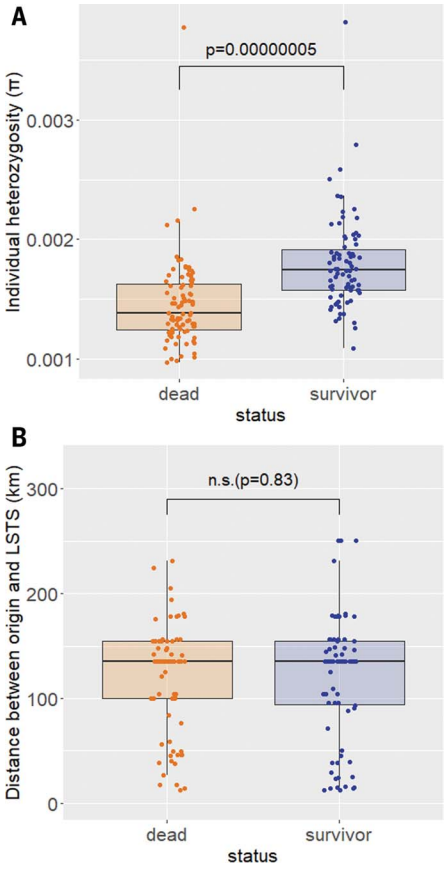


Fig. 3. Box-and-whisker plots showing the mean and distribution of individual heterozygosity. (**A** and **B**) Heterozygosity π (A) and straight-line translocation distances (B) of LSTS mortalities and survivors. Reported *P* values are based on *t* tests (n.s., not significant).

survival. We are not suggesting that individual heterozygosity should be the only criterion for deciding which individuals to translocate: Local ecology, disease exposure history, and individual condition are some of the other factors that are often critical, and we stress the importance of verifying these results in other systems. We also emphasize that our RAD data, although extensive, are only a proxy for the entire genome, and that additional studies with high-coverage whole-genome resequencing could help to determine whether survivorship is linked to runs of homo- or heterozygosity, level of individual inbreeding, or specific loci under strong selection. Regardless, one of the advantages of individual heterozygosity is that it can now be easily and economically measured, with reduced representation approaches or at the whole-genome level for most organisms, making it a particularly attractive tool for man1agers and decision-makers.

Our analysis of LSTS survivors and mortalities, combined with detailed landscape genomic data for the entire species, indicates that matching the geographical provenance of translocated tortoises to their new site had virtually no predictive power in determining survival fitness over a decadal time scale. However, individual heterozygosity did, with more heterozygous individuals outsurviving less heterozygous ones. The overall benefits of genome-level variability have long been suggested as a key fitness component (*7*, *24*), and we were able to use long-term field data to directly show this benefit under natural conditions. Although field observations have confirmed living status for only 3.8% of translocated tortoises, our data indicate that even under these severe conditions, the more variable tortoises outsurvived their less variable conspecifics, which suggests that these genetically variable individuals were better able to survive after translocation. To confirm that individual heterozygosity is responsible for the increase in survival requires understanding the proximate reasons for the massive mortality seen in both translocated and resident tortoises, and we currently lack that knowledge. Long-term monitoring found that the annual mean declines of 7.4% and 9.2% at the individual sites closest to the LSTS occurred between 2004 and 2014 (*25*), corresponding to population reductions of roughly 57% and 65% over 11 years. These declines, although not fully understood, are at least partially attributed to severe regional drought (*26*), which has been associated with sharp increases in mortality of Mojave desert tortoises (*27*–*29*). Given our current understanding, we can only speculate that drought, combined with high post-release densities, disease, and/or the ecological disruption associated with translocation, may be contributing to the high LSTS mortalities.

However, even without a proximate mechanism, our results suggest that an optimal strategy of assisted migration could be to prioritize moving the most genomically variable individuals, rather than current practice based solely on geographic or genetic similarity. Given the future climate and anthropogenic changes anticipated across the region, assisted migration will likely be a key component of management of desert tortoises and many other declining or endangered species, and our data indicate that targeting the most genomically variable individuals can enhance the success of this work. Future research aimed at understanding the proximate reasons for this increased survival at the genetic and physiological level constitutes an important next step for more efficient conservation-based translocation success.

REFERENCES AND NOTES

1. B. B. Beck, L. G. Rappaport, M. S. Price, A. Wilson, in *Creative Conservation: Interactive Management of Wild and Captive Animals*, P. J. S. Olney, G. M. Mace, A. T. C. Feistner, Eds. (Chapman and Hall, 1994), pp. 265–284.
2. J. Fischer, D. B. Lindenmayer, *Biol. Conserv.* **96**, 1–11 (2000).
3. S. Godefroid *et al.*, *Biol. Conserv.* **144**, 672–682 (2011).
4. S. Sánchez-Ramírez *et al.*, *Sci. Rep.* **8**, 11231 (2018).
5. D. H. Reed, R. Frankham, *Conserv. Biol.* **17**, 230–237 (2003).
6. J. R. Chapman, S. Nakagawa, D. W. Coltman, J. Slate, B. C. Sheldon, *Mol. Ecol.* **18**, 2746–2765 (2009).
7. M. Szulkin, N. Bierne, P. David, *Evolution* **64**, 1202–1217 (2010).
8. U.S. Fish and Wildlife Service, *Fed. Regist.* **55**, 12178–12191 (1990).
9. U.S. Fish and Wildlife Service, *Desert Tortoise (Mojave Population): Recovery Plan* (1994).
10. U.S. Fish and Wildlife Service, *Revised Recovery Plan for the Mojave Population of the Desert Tortoise (Gopherus agassizii)* (2011).
11. R. W. Murphy, K. H. Berry, T. Edwards, A. M. McLuckie, *Chelonian Conserv. Biol.* **6**, 229–251 (2007).
12. B. E. Hagerty, C. R. Tracy, *Conserv. Genet.* **11**, 1795–1807 (2010).
13. H. B. Shaffer *et al.*, bioRxiv 195743 [preprint]. 29 September 2017.
14. T. Edwards, C. R. Schwalbe, D. E. Swann, C. S. Goldberg, *Conserv. Genet.* **5**, 485–499 (2004).
15. T. Edwards, K. Berry, *Conserv. Genet.* **14**, 649–659 (2013).
16. R. C. Averill-Murray, B. E. Hagerty, *Chelonian Conserv. Biol.* **13**, 35–41 (2014).
17. R. Tracy, P. Medica, P. S. Corn, *Translocation Long-Term Monitoring, Tortoise Density Evaluation, and Establishment of New LSTSs* (Clark County, Nevada, Desert Conservation Program, 2006).
18. See supplementary materials.
19. B. A. Lawrence, T. N. Kaye, *Restor. Ecol.* **19**, 166–176 (2011).
20. M. C. Yates, D. J. Fraser, *Evol. Appl.* **7**, 871–882 (2014).
21. M. C. Yates, E. Bowles, D. J. Fraser, *Proc. R. Soc. B* **286**, 20191989 (2019).
22. G. Bowman, C. Perret, S. Hoehn, D. J. Galeuchet, M. Fischer, *J. Ecol.* **96**, 1056–1064 (2008).
23. C. G. Oakley, *Evol. Appl.* **6**, 823–831 (2013).
24. B. Hansson, L. Westerberg, *Mol. Ecol.* **11**, 2467–2474 (2008).
25. L. J. Allison, A. M. McLuckie, *Herpetol. Conserv. Biol.* **13**, 433–452 (2018).
26. U.S. Drought Monitor; www.drought.gov/drought/states/nevada.
27. C. C. Peterson, *Biol. Conserv.* **70**, 101–108 (1994).
28. K. M. Longshore, J. R. Jaeger, J. M. Sappington, *J. Herpetol.* **37**, 169–177 (2003).
29. J. E. Lovich *et al.*, *Biol. Conserv.* **169**, 214–224 (2014).
30. J. Sambrook, D. W. Russell, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, 2001).
31. N. J. Bayona-Vásquez *et al.*, *PeerJ* **7**, e7724 (2019).
32. B. K. Peterson, J. N. Weber, E. H. Kay, H. S. Fisher, H. E. Hoekstra, *PLOS ONE* **7**, e37135 (2012).
33. A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **30**, 2114–2120 (2014).
34. E. Aronesty, ea-utils: Command-line tools for processing biological sequencing data (2011); https://github.com/ExpressionAnalysis/ea-utils.
35. M. Tollis *et al.*, *PLOS ONE* **12**, e0177708 (2017).
36. H. Li, arXiv 1303.3997 [q-bio.GN] (16 March 2013).
37. H. Li *et al.*, *Bioinformatics* **25**, 2078–2079 (2009).
38. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, *BMC Bioinformatics* **15**, 356 (2014).
39. D. A. Eaton, *Bioinformatics* **30**, 1844–1849 (2014).
40. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
41. M. Foll, O. Gaggiotti, *Genetics* **180**, 977–993 (2008).
42. M. Foll, M. C. Fischer, G. Heckel, L. Excoffier, *Mol. Ecol.* **19**, 4638–4647 (2010).
43. M. C. Fischer, M. Foll, L. Excoffier, G. Heckel, *Mol. Ecol.* **20**, 1450–1462 (2011).
44. M. Plummer, N. Best, K. Cowles, K. Vines, *R News* **6**, 7–11 (2006).
45. A. Raj, M. Stephens, J. K. Pritchard, *Genetics* **197**, 573–589 (2014).

ACKNOWLEDGMENTS

The Great Basin Institute conducted surveys and collected blood samples from the LSTS. We thank E. McCartney-Melstad, P. Ralph, C. R. Tracy, and members of the Shaffer lab at UCLA for input and advice, and P. Kareiva and D. Simberloff for thoughtful comments on earlier drafts of this manuscript. The findings and conclusions

provided in this article are those of the authors and do not necessarily represent the views of their affiliated organizations. All field work and sample collection was carried out under USFWS recovery permit TE-108507-2 and with written permission from the Bureau of Land Management to L.J.A. **Funding:** Partial funding was provided by the Bureau of Land Management Las Vegas Field Office and the NSF. **Author contributions:** All authors contributed to conceptualizing the main experiment based on sequencing LSTS tortoises. L.J.A. and K.J.F. coordinated collection of LSTS samples and identification of tortoise histories. P.A.S. conducted all laboratory preparation, bioinformatics processing, and analyses. P.A.S. and H.B.S. wrote the manuscript with input from R.C.A.-M., L.J.A., and K.J.F. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All sequence data are available in the NCBI Short Read Archive (SRA accession: PRJNA638160). Data on yearly tortoise releases and resightings are available in the supplementary text: "Population density estimates in the LSTS and surrounding regions". These data are accompanied by additional information regarding pre-LSTS population estimates, regional tortoise declines, and our understanding that long-term drought is a causal force in these declines.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1086/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 and S2
References (*30*–*45*)

4 February 2020; resubmitted 11 September 2020
Accepted 20 October 2020
10.1126/science.abb0421

CORONAVIRUS

Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate

Sandhya Bangaru[1], Gabriel Ozorowski[1], Hannah L. Turner[1], Aleksandar Antanasijevic[1], Deli Huang[2], Xiaoning Wang[3], Jonathan L. Torres[1], Jolene K. Diedrich[3], Jing-Hui Tian[4], Alyse D. Portnoff[4], Nita Patel[4], Michael J. Massare[4], John R. Yates III[3], David Nemazee[2], James C. Paulson[2,3], Greg Glenn[4], Gale Smith[4], Andrew B. Ward[1]*

Vaccine efforts to combat the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is responsible for the current coronavirus disease 2019 (COVID-19) pandemic, are focused on SARS-CoV-2 spike glycoprotein, the primary target for neutralizing antibodies. We performed cryo–electron microscopy and site-specific glycan analysis of one of the leading subunit vaccine candidates from Novavax, which is based on a full-length spike protein formulated in polysorbate 80 detergent. Our studies reveal a stable prefusion conformation of the spike immunogen with slight differences in the S1 subunit compared with published spike ectodomain structures. We also observed interactions between the spike trimers, allowing formation of higher-order spike complexes. This study confirms the structural integrity of the full-length spike protein immunogen and provides a basis for interpreting immune responses to this multivalent nanoparticle immunogen.

Severe acute respiratory syndrome coronavirus (SARS-CoV) caused a global outbreak from 2002 to 2003 (*1*). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), from the same lineage of the β-CoV genus as SARS-CoV, recently emerged in China and spread rapidly, infecting more than 28 million people worldwide by September 2020 (*2*). Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, was declared a pandemic by the World Health Organization (WHO). In response, several SARS-CoV-2 vaccine candidates are being developed and tested at various stages of clinical trials (*3*–*5*). The SARS-CoV-2 spike (S) trimeric glycoprotein is a focus of vaccine development because it is the primary target of host immune defenses (*5*, *6*).

Like other type 1 fusion proteins, the SARS-CoV-2 S prefusion trimer is metastable and undergoes structural rearrangement from a prefusion to a postfusion conformation upon S-protein receptor binding and cleavage (*7*, *8*). The structure of the stabilized SARS-CoV-2 spike ectodomain has been solved in its prefusion conformation and resembles the SARS-CoV spike (*9*–*11*). Here, we describe the structure of a leading SARS-CoV-2 S vaccine candidate (NVAX-CoV2373) based on a full-length (FL) S, residues 1 to 1273, which includes the transmembrane (TM) and the cytoplasmic tail (CT) (Fig. 1A). The final construct, SARS-CoV-2-3Q-2P, was also modified at the S1/S2 polybasic cleavage site from RRAR to QQAQ to render it protease resistant, along with two proline substitutions at residues K986 and V987 in the S2 fusion machinery core for enhanced stability (Fig. 1A). The FL spikes, expressed and purified from insect cells, were formulated in 0.01% (v/v) polysorbate 80 (PS 80) detergent. To characterize the structural integrity of the 3Q-2P-FL immunogen, we performed negative-stain electron microscopy of the FL spike constituted in PS 80 in the presence of Matrix-M adjuvant, recapitulating the vaccine formulation being tested in humans. Imaging revealed trimeric spike proteins present as free trimers or as multitrimer rosettes, containing as many as 14 trimers with their TM domains enclosed in micellar cores of PS 80 detergent (Fig. 1B). Tight clustering of the spikes in the NVAX-CoV2373 nanoparticle formulation may lead to stronger immune responses over soluble trimers alone, similar to other viral glycoprotein immunogens (hemagglutinin and respiratory syncytial virus F) (*12*, *13*).

We next performed single-particle cryo–electron microscopy (cryo-EM) on the spike formulated in PS 80 detergent (Fig. 2A). Initial two-dimensional (2D) classification revealed the presence of two distinct classes: free spike trimers and dimers of trimers (Fig. 2A). The threefold symmetric (C3) reconstruction of the free spike trimer resulted in a 3.6 Å–resolution map, whereas the asymmetric reconstruction (C1) was refined to 3.8-Å resolution (Fig. 2B and fig. S1, A and B). In previous structures, receptor binding domains (RBDs) exist in either a closed (RBD-down) or an open (RBD-up) conformation that can engage in ACE2 binding (*9*, *10*, *14*). By contrast, we observed that all three RBDs on the 3Q-2P-FL spike trimer were in the closed conformation in our reconstructions (Fig. 2B and fig. S1C). Despite the RBD-down conformation, binding analysis of the 3Q-2P-FL immunogen to ACE2 by both biolayer interferometry and enzyme-linked immunosorbent assay clearly shows binding to ACE2, indicating that the RBD is dynamic and the receptor binding site accessible (*15*). Another study on the prefusion structure of an FL spike protein reported similar findings with RBDs clamped down as a consequence of potential clashes between S2 residues 828 to 853 and subdomain 1 (SD1) when RBD is in open conformation (*16*). Recent reports by Henderson *et al.* have revealed that introducing mutations and removing N-linked glycosylation at certain positions can alter the propensity toward "up" and "down" states of the RBD (*17*, *18*).

Overall, our cryo-EM map was well resolved in both S1 and S2 subunits (fig. S1D), enabling us to model the full S1 N-terminal domain (NTD) and C-terminal domain (CTD) that were less resolved in previous structures (*9*, *10*). Our final atomic model contains residues 14 to 1146 with breaks only in the flexible loop (619 to 631) and the cleavage site (678 to 688) (Fig. 2C). Superimposition of the coordinate models of 3Q-2P-FL spike with published spike structures [Protein Data Bank (PDB) IDs: 6VXX and 6VSB] revealed substantial domain rearrangements in the S1 subunit of 3Q-2P-FL spike (Fig. 2D). The S1 NTD rotated ~14° relative to published models, whereas the CTD and subdomains showed minor local rearrangements (Fig. 2D). Another recent study also observed differences in NTD conformations at lower pH, although our cryo-EM studies were carried out at neutral pH (*19*). In our 3Q-2P-FL structure, we observed a shift in residues flanking the 615 to 635 loop, resulting in a salt bridge between residue D614 on one protomer and K854 on a neighboring protomer (Fig. 3A). This observation is particularly notable given the increased prevalence of D614→G (D614G) mutation in the emerging SARS-CoV-2 strains and its potential role in viral transmission and pathogenesis (*20*). The 615 to 635 loop that is generally disordered in spike trimer structures, including ours, was recently modeled as a helix (PDB ID: 6X6P) (Fig. 3B), although the cryo-EM density (EMD-22078) does not support this assignment (fig. S1E) (*11*).

We observed two additional densities in the S1 subunit that did not correspond to peptide or glycans within the spike (fig. S2A). The first density was buried within a hydrophobic pocket of the CTD (Fig. 3C). We have previously showed palmitoleic acid occupying a similar pocket in the structure of porcine epidemic diarrhea virus (*21*). This density in SARS-CoV-2 S corresponded to linoleic acid, a polyunsaturated fatty acid; the presence of this ligand was confirmed by mass spectrometry of 3Q-2P-FL spike (fig. S2, B and C). The main chain carboxyl group of linoleic acid interacts with the R408 and Q409 residues of the RBD from the adjacent

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [2]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA. [3]Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA. [4]Novavax, Inc., 21 Firstfield Road, Gaithersburg, MD 20878, USA.
*Corresponding author. Email: andrew@scripps.edu

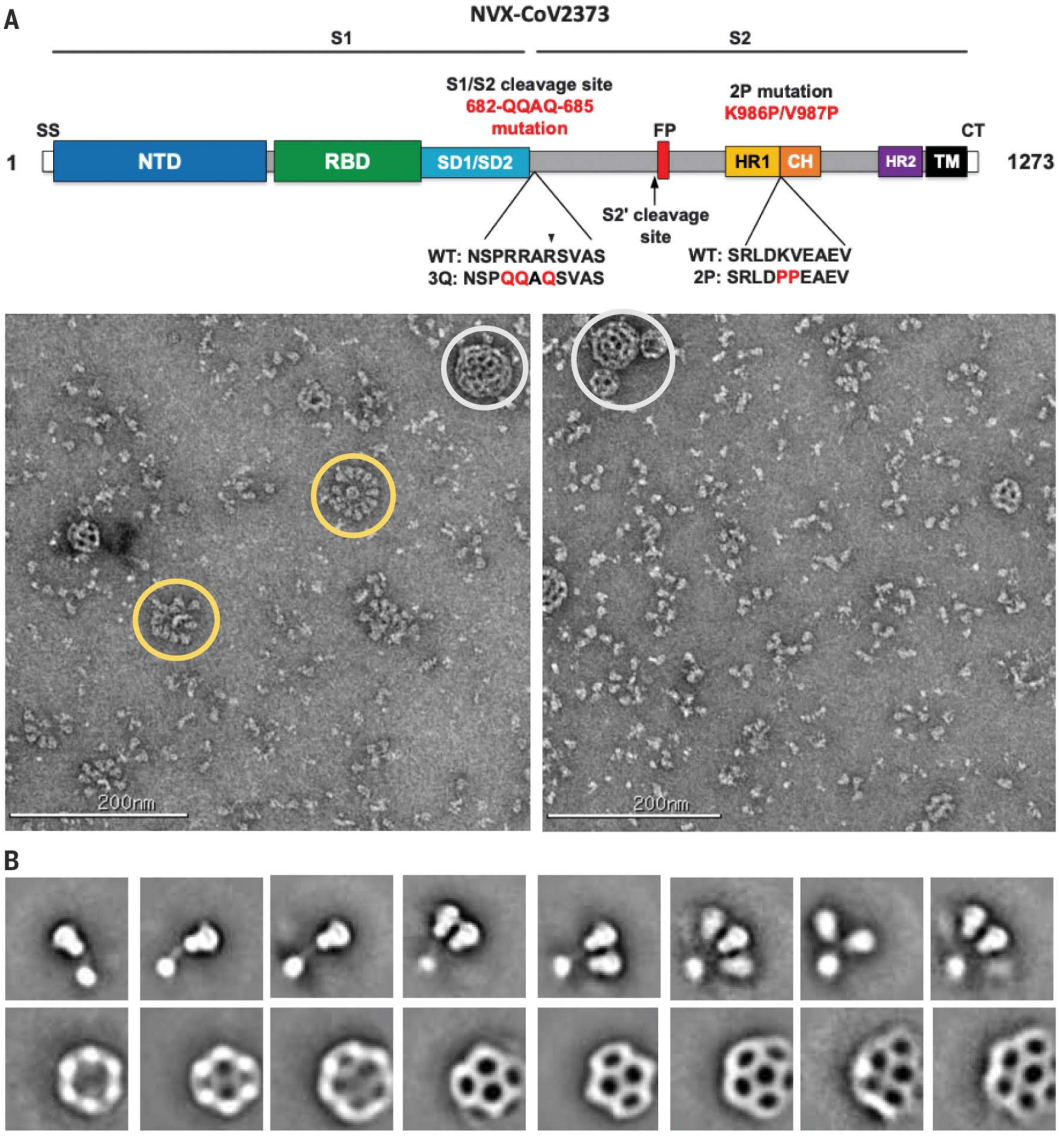


Fig. 1. Evaluation of SARS-CoV-2 3Q-2P-FL spike glycoprotein. (**A**) Linear diagram of the sequence and structure elements of the FL SARS-CoV-2 spike protein showing the S1 and S2 ectodomain. Structural elements include a cleavable signal sequence (SS, white), NTD (blue), RBD (green), SD1 and SD2 (light blue), protease cleavage site 2′ (S2′, arrow), fusion peptide (FP, red), heptad repeat 1 (HR1, yellow), central helix (CH, brown), heptad repeat 2 (HR2, purple), TM domain (black), and CT (white). The native furin cleavage site was mutated (RRAR→QQAQ) to be protease resistant and stabilized by introducing two proline (2P) substitutions at positions K986P and V987P to produce SARS-CoV-2 3Q-2P-FL spike. A, Ala; D, Asp; E, Glu; K, Lys; L, Leu; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; V, Val. (**B**) Representative negative-stain EM images and 2D classes of SARS-CoV-2 3Q-2P-FL, formulated in PS 80 detergent in the presence of Matrix-M adjuvant. In the raw micrograph, spike rosettes are circled in yellow and Matrix-M adjuvant cages are circled in white. 2D classes showing individual spikes, higher-order spike nanoparticles, and Matrix-M cages of different sizes. Matrix-M does not appear to interact with the spike nanoparticles.

protomer, potentially stabilizing the observed RBD-down state (Fig. 3C) and consistent with a recent report (*22*). The second unassigned density, present in the NTD, was larger and more surface exposed than the first (Fig. 3D and fig. S2D). The aliphatic tail of PS 80 fit well into this hydrophobic pocket, whereas the carbonyl and hydroxyl groups were in proximity to residues R190 and H207 with potential for multiple hydrogen bonds between them (Fig. 3D and fig. S2D). The location of the PS 80 ligand provides a possible explanation for the S1 shift seen in our FL trimer density. PS 80 is specific to the formulation of the Novavax 3Q-2P-FL immunogen, but other ligands may also bind this pocket and provide a potential target for drug design against SARS-CoV-2.

Classification of multimeric spike trimer particles yielded two separate classes: a dimer-of-trimers class that reconstructed to a final resolution of 4.5 Å with twofold symmetry and a trimer-of-trimers class that was resolved to 8.0-Å resolution (Fig. 4, A and B, and fig. S3A). In both reconstructions, the interaction between each pair of trimers involved the SD2 of one protomer from each trimer engaging with the NTD of the adjacent trimer (Fig. 4C), with trimer axes tilted 44.5° relative to each other. The dimer-of-trimer interaction was mainly coordinated by the 615 to 635 loop, which, in contrast to the free-trimer structure, was now fully resolved (Fig. 4D). The loop reaches into and induces subtle changes to a pocket on the adjacent NTD compared with the free-trimer model (Fig. 4D). Residues Y145 and H146 in the binding pocket appear to switch positions in the loop-bound state, resulting in a salt-bridge interaction between H146 and D627 and potential stacking between W152 and H146 (Fig. 4E). We also observed minor displacement of residues 68 to 75 and 248 to 250 surrounding the pocket. In the dimer-of-trimers, we also observed N282 glycans at the dimer interface (fig. S3B). As a control, we also performed cryo-EM studies of the SARS-CoV-2-3Q-FL (without 2P). Notably, the structures of the trimers were identical, and we also observed dimers of trimers (fig. S3, C to E)

Sequence alignment of residues in the 615 to 635 loop and corresponding NTD binding pocket across representative CoV strains belonging to lineage B of betacoronaviruses revealed residues 621-PVAIHADQ-628 are well conserved, but there are notable differences in the binding pocket residues (fig. S4A). Substantial gaps in the interacting NTD loops along with the absence of H146 at the corresponding site on SARS-CoV make it unlikely that SARS-CoV participates in similar intertrimeric interactions. Although the residues in the NTD pocket were almost identical between SARS-CoV-2 and its closely related bat strain Bat-SL-RatG13, we observed some residue differences and one to three amino acid

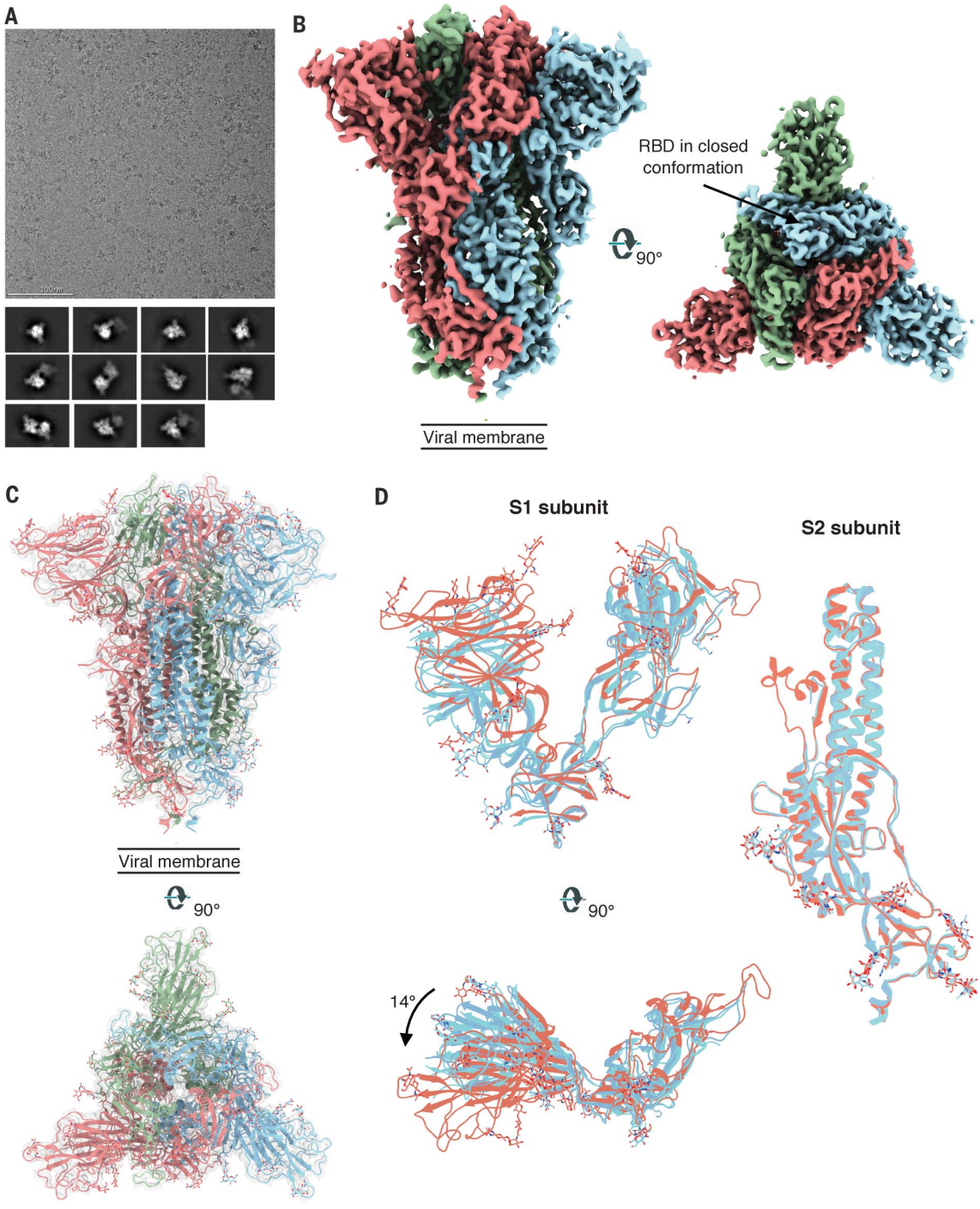


Fig. 2. Cryo-EM analysis of SARS-CoV-2 3Q-2P-FL spikes. (**A**) Representative electron micrograph and 2D class averages of 3Q-2P-FL spikes showing free trimers and complexes of trimers. (**B**) Side and top views of the B factor–sharpened cryo-EM map of 3Q-2P-FL free trimers showing the spike in prefusion state, with the RBDs in closed conformation. The protomers are colored in blue, green, and coral for clarity. (**C**) Side and top view of the atomic model of free trimer represented as a ribbon diagram fit into the map density. The protomers are colored in blue, green, and coral, and the map is shown as a transparent gray density. (**D**) Comparison of 3Q-2P-FL spike with published structures (PDB IDs 6VXX and 6VSB) on a subunit level. PDB 6VXX is shown in cyan, PDB 6VSB in blue, and 3Q-2P-FL spike in coral.

deletions in the loops comprising the NTD binding pocket of representative strains Bat-SL-CoVZC45, BetaCoV/pangolin/Guangdong/1/2019, and BetaCoV/pangolin/Guangxi/P4L/2007 (fig. S4A).

Some human CoVs, including OC43, exclusively use NTD–sialic acid (SA) interactions as their receptor engagement, whereas others such as Middle East respiratory syndrome (MERS) CoV that use the CTD-RBD for primary receptor binding have also been reported to bind SA receptors through their NTD to aid initial attachment to the host cells (*23*–*25*). Structural comparisons of the SARS-CoV-2 NTD dimerization pocket with that of the SA binding site on MERS spike revealed that they did not coincide with each other (PDB ID: 6Q04) (*25*) (fig. S4B). Computational and structural studies have proposed residues on SARS-CoV-2 spike that may be involved in SA binding (*26*, *27*). Structural comparison of this putative glycan binding site to the dimerization site revealed them situated adjacent to one another with residues in loop 70 contributing to both the binding pockets (fig. S4C).

We next performed cell surface expression and pseudovirus replication assays with SARS-CoV-2 wild-type (WT) spike and spikes containing mutations in the 615 to 635 loop and NTD pocket. Each residue in the loop 621-PVAIHADQ-628 and residue H146 in the binding pocket were individually mutated to either alanine or glycine. Additionally, we made a spike construct with all eight residues 621-PVAIHADQ-628 replaced with a glycine-serine

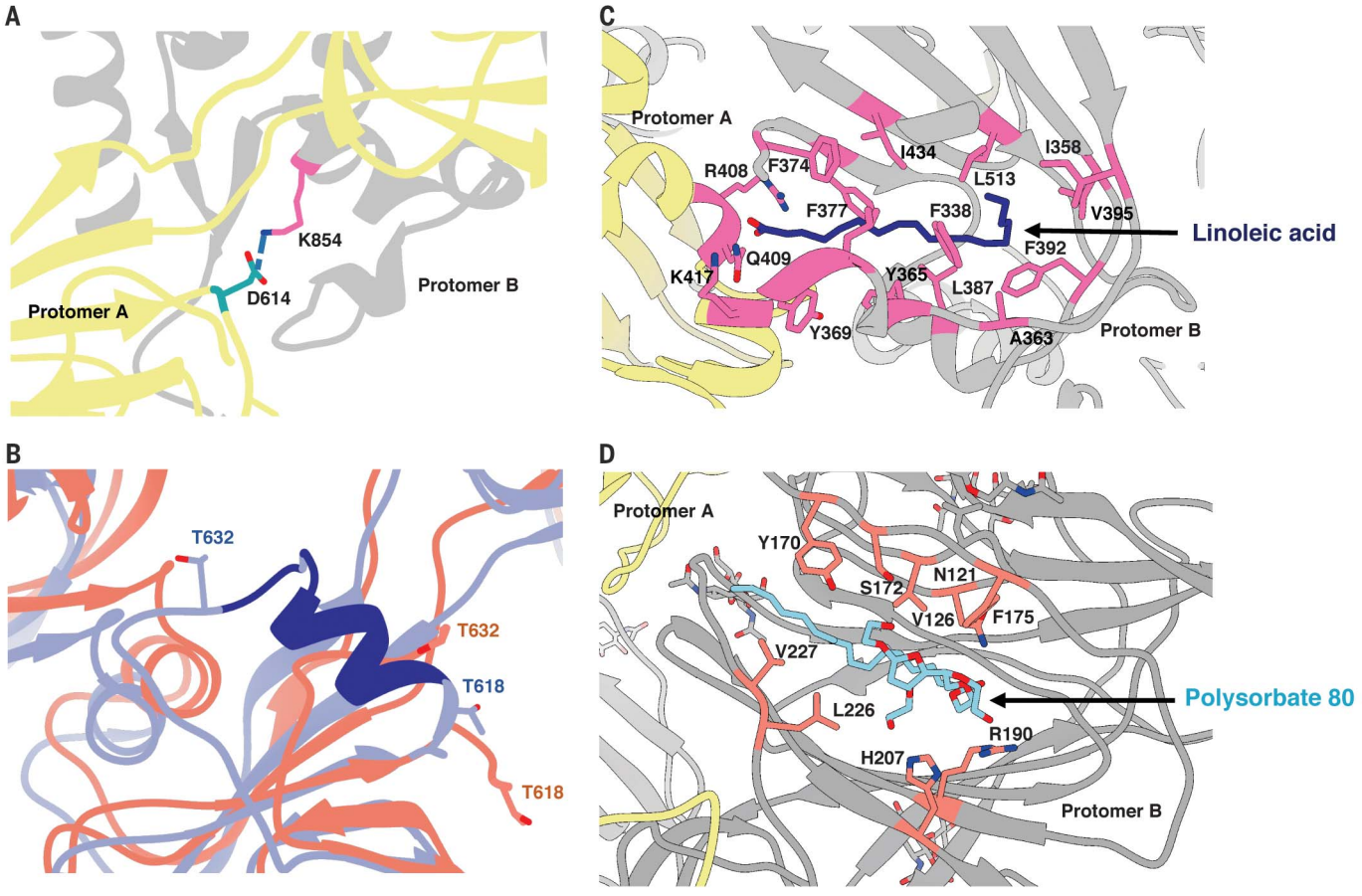


Fig. 3. Structural features of the SARS-CoV-2 3Q-2P-FL spike trimer. (**A**) Interprotomeric salt-bridge interaction between D614 and K854 in 3Q-2P-FL spike trimer. (**B**) Comparison of the 615 to 635 loop between 3Q-2P-FL spike shown in coral and PDB 6X6P shown in blue. The residues that were built in 6X6P model but not in our model are shown in dark blue. Threonines at positions 618 and 632 flanking the gap in the 3Q-2P-FL trimer model are shown on both models to highlight their relative positions. T, Thr. (**C**) Linoleic acid (dark blue) binding within a hydrophobic pocket of one RBD where the fatty acid head group reaches out to interact with the closed RBD of the adjacent protomer. The interacting residues are shown in pink. F, Phe; I, Ile; Y, Tyr. (**D**) PS 80 detergent (blue) binding within the NTD with potential hydrogen bonding with R190 and H207. The interacting residues are shown in orange. Adjacent protomers are shown in yellow and gray in (A), (C), and (D). H, His.

(GS) linker to completely abrogate binding. Compared with the WT, the mutants generally exhibited lower levels of infectivity (Fig. 4F). Cell surface expression of these mutants in 293T cells revealed that these mutations also disrupted surface expression of the spike protein, with linear correlation between surface expression and pseudovirus replication (Fig. 4G).

Glycans on viral glycoproteins play a wide role in protein folding, stability, and immune recognition and also in facilitating immune evasion. We therefore conducted site-specific glycosylation analysis of the SARS-CoV-2 prefusion spike protein produced in Sf9 insect cells as previously described (*28*) to assess the extent of glycosylation and the degree of glycan processing from high-mannose or hybrid type to complex type. The analysis detected glycosylation at all 22 N-linked glycan sequons present on SARS-CoV-2 spike (Fig. 4H). Overall, there was high glycan occupancy of >98%, with only two sites (603 and 657) >5% unoccupied. We did not see clear glycan density at either 603 or 657 in the cryo-EM reconstruction of the 3Q-2P-FL spike. Most sites showed extensive glycan processing to complex or paucimannose-type glycans, with only four sites exhibiting ≥40% oligomannose. The glycan analysis also confirmed the presence of glycans at sites 1158, 1173, and 1194 present in the membrane-proximal region of the spike not resolved by cryo-EM. By comparison with site-specific glycan processing of the spike protein produced in mammalian human embryonic kidney (HEK) 293F cells, both mammalian cells and insect cells exhibit extensive processing at most sites. In general, however processing of glycans on the 2019 CoV prefusion spike protein from insect cells was somewhat greater, particularly at sites 709 and 717, which were predominately oligomannose in spike from HEK293 cells but exclusively complex or paucimannose in spike from Sf9 cells (*29*).

Our structural work is consistent with the burgeoning body of spike structures, albeit with notable differences in the rearrangement of S1 domains and formation of intertrimer interactions (*9*, *10*). Both these findings were seen in the FL spike immunogens assembled into compact and dense nanoparticles. Cryo–electron tomographic reconstructions of intact SARS-CoV-2 virions showed a relatively dispersed distribution of spike protein trimers on the viral surface and no evidence of higher-order aggregates (*30*). However, another study showed that the D614G mutation present in close proximity to the dimerization loop results in a several-fold increase of spike numbers on the viral surface, resulting in higher spike protein density and a more infectious virion (*20*). The greater density may be aided by the ability to form such higher-order multimers. Alternatively, the loop that mediates interspike interactions may play a role in viral viability, consistent with our loop mutant data.

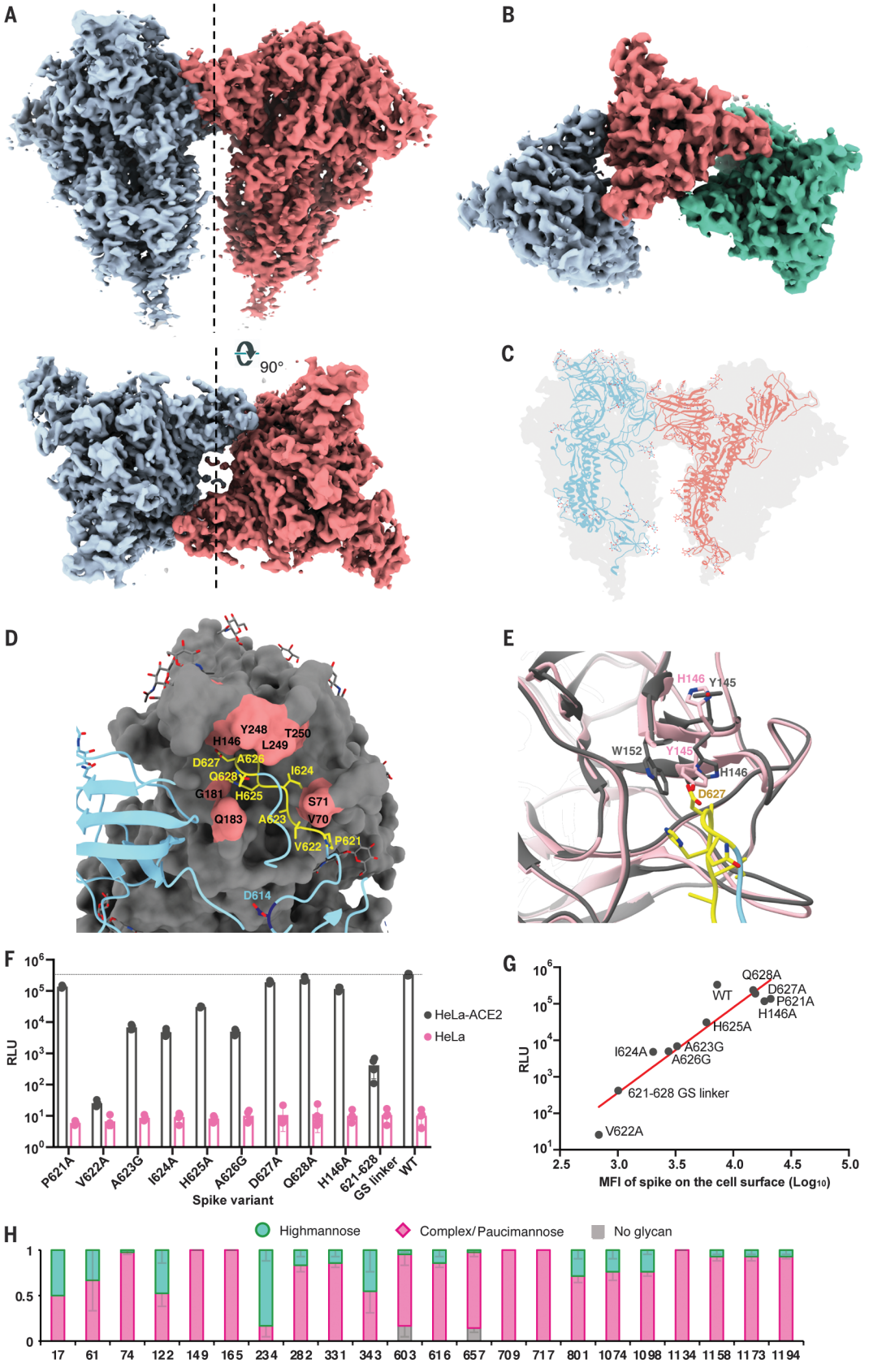


Fig. 4. Trimer-trimer interactions and glycan analysis. (**A**) Side and top views of the sharpened cryo-EM map of 3Q-2P-FL dimers of spike trimers. Individual spike trimers are shown in blue and coral along a twofold axis of symmetry (dotted line). (**B**) Top view of the B factor–sharpened cryo-EM map of trimer-of-trimers complex with individual trimers colored in blue, coral, and green. (**C**) Ribbon representation of a protomer from one trimer (blue) interacting with the protomer from the adjacent trimer (coral) docked into the dimers-of-trimers density. (**D**) A close-up view of the interaction between the protomers of adjacent trimers. One protomer is shown as a ribbon diagram in blue, and its binding partner is shown as surface in gray. Residues 621-PVAIHADQ-628 in the loop with potential interactions to the neighboring NTD are colored yellow, and the residues in the NTD binding pocket are highlighted in coral. Residue D614 at the start of the loop is highlighted in dark blue. Glycosylation at residue 616 is not shown for clarity. G, Gly. (**E**) Changes occurring in the binding pocket in the bound state (gray) versus the free trimer (pink). Y145 and H146 switch positions to accommodate the loop better, also resulting in salt-bridge formation between H146 and D627. It also results in stacking between W152 and H146. W, Trp. (**F**) Pseudoviruses expressing SARS-CoV-2 WT or mutant spikes were used to infect HeLa or HeLa-ACE2 cells for 42 to 48 hours. Infection was measured by luciferase intensity RLU (relative light unit) in the lysed cells after infection. (**G**) Correlation between pseudovirus infection (RLU) and surface expression of SARS-CoV-2 spike variants in 293T cells measured by MFI (mean fluorescence intensity). (**H**) Site-specific glycan analysis of 3Q-2P-FL spike protein expressed in Sf9 insect cell line. Proportions shown for no occupancy, oligomannose, and complex or paucimannose potential N-linked glycosylation sites (PNGS) are the average and SEM of 3 to 32 distinctive peptides for each glycosite except for sites 17, 709, and 717, where only a single peptide was observed.

Analysis of safety and immunogenicity of the Novavax SARS-CoV-2-3Q-2P-FL immunogen in mice and baboons revealed strong B and T cell responses to the vaccine with no evidence of vaccine-associated enhanced respiratory disease (*15*). Phase 1 and 2 clinical trial results showed that the vaccine induced immune responses exceeding levels seen in COVID-19 patients (*31*). Overall, we found that NVAX-CoV2372 is stable, homogeneous, and locked in the antigenically preferred prefusion conformation. With structural, biophysical, and antigenic characterization now complete, ongoing evaluation in humans will provide the true proof-of-principle for this vaccine concept.

REFERENCES AND NOTES

1. J. D. Cherry, P. Krogstad, *Pediatr. Res.* **56**, 1–5 (2004).
2. N. Zhu *et al.*, *N. Engl. J. Med.* **382**, 727–733 (2020).
3. F. Amanat, F. Krammer, *Immunity* **52**, 583–589 (2020).
4. N. Lurie, M. Saville, R. Hatchett, J. Halton, *N. Engl. J. Med.* **382**, 1969–1973 (2020).
5. G. Salvatori *et al.*, *J. Transl. Med.* **18**, 222 (2020).
6. S. Belouzard, J. K. Millet, B. N. Licitra, G. R. Whittaker, *Viruses* **4**, 1011–1033 (2012).
7. A. C. Walls *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11157–11162 (2017).
8. B. J. Bosch, R. van der Zee, C. A. de Haan, P. J. Rottier, *J. Virol.* **77**, 8801–8811 (2003).
9. A. C. Walls *et al.*, *Cell* **181**, 281–292.e6 (2020).
10. D. Wrapp *et al.*, *Science* **367**, 1260–1263 (2020).
11. N. G. Herrera *et al.*, bioRxiv 2020.06.14.150607 [Preprint]. 17 June 2020.
12. G. Smith *et al.*, *PLOS ONE* **7**, e50852 (2012).
13. V. Shinde *et al.*, *N. Engl. J. Med.* **378**, 2346–2348 (2018).
14. J. Shang *et al.*, *Nature* **581**, 221–224 (2020).
15. J.-H. Tian *et al.*, bioRxiv 2020.06.29.178509 [Preprint]. 30 June 2020.
16. Y. Cai *et al.*, *Science* **369**, 1586–1592 (2020).
17. R. Henderson *et al.*, bioRxiv 2020.06.26.173765 [Preprint]. 26 June 2020.
18. R. Henderson *et al.*, *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020).
19. T. Zhou *et al.*, bioRxiv 2020.07.04.187989 [Preprint]. 4 July 2020.
20. L. Zhang *et al.*, bioRxiv 2020.06.12.148726 [Preprint]. 12 June 2020.
21. R. N. Kirchdoerfer *et al.*, bioRxiv 2020.02.18.955195 [Preprint]. 19 February 2020.
22. C. Toelzer *et al.*, *Science* **370**, 725–730 (2020).
23. R. J. G. Hulswit *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2681–2690 (2019).
24. W. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8508–E8517 (2017).
25. Y. J. Park *et al.*, *Nat. Struct. Mol. Biol.* **26**, 1151–1157 (2019).
26. E. Milanetti *et al.*, bioRxiv 2020.03.24.006197 [Preprint]. 6 April 2020.
27. M. Awasthi *et al.*, *Viruses* **12**, 909 (2020).
28. L. Cao *et al.*, *Nat. Commun.* **9**, 3693 (2018).
29. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, *Science* **369**, 330–333 (2020).
30. Z. Ke *et al.*, *Nature* 10.1038/s41586-020-2665-2 (2020).
31. C. Keech *et al.*, *N. Engl. J. Med.* 10.1056/NEJMoa2026920 (2020).

ACKNOWLEDGMENTS

We thank B. Anderson, H. L. Turner, and C. A. Bowman for their help with electron microscopy, data acquisition, and data processing. We thank B. Webb and L. T. Hoang for their assistance with mass spectrometry and data processing. We thank L. Holden for her assistance with the manuscript. We also thank A. M. Greene at Novavax, Inc., for editing the manuscript. **Funding:** This work was supported by grants from the National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Development (UM1 AI144462 to J.C.P. and A.B.W., R01 AI113867 to J.C.P., R01 AI132317 to D.N., and P01 AI110657 to A.B.W.), the Bill and Melinda Gates Foundation (OPP1170236 to A.B.W.), and Novavax, Inc., Molecular graphics and analyses were performed with UCSF Chimera developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health (R01-GM129325 and P41-GM103311) and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. **Author contributions:** S.B. and A.B.W. conceived and designed the study. S.B., H.L.T., G.O., and A.A. performed cryo-EM data collection, data processing, and model building. X.W., J.K.D., J.R.Y., and J.C.P. performed site-specific glycan analysis and data interpretation. J.L.T., D.H., and D.N. performed mutagenesis and pseudovirus assays. S.B., G.O., and A.B.W. analyzed and interpreted data. S.B. and A.B.W. wrote the paper, and all authors reviewed and edited the paper. J.H.T., A.D.P., N.P., M.J.M., G.G., and G.S. contributed NVX-CoV2373 and Matrix-M adjuvant and provided advice for sample handling. J.H.T., A.D.P., N.P., M.J.M., G.G., and G.S. also contributed to drafting of the manuscript. **Competing interests:** Authors J.H.T., A.D.P., N.P., M.J.M., G.G., and G.S. are current employees of Novavax, Inc., a for-profit organization, and these authors own stock or hold stock options. These interests do not alter the authors' adherence to policies on sharing data and materials. Authors H.L.T. and A.B.W. are inventors on U.S. patent application no. 62/412,703 ("Prefusion Coronavirus Spike Proteins and Their Use"). All other authors have no competing interests to declare. **Data and materials availability:** The EM maps have been deposited at the Electron Microscopy Data Bank (EMDB) with accession codes EMD-22352 (SARS-CoV-2 3Q-2P-FL spike trimer with C3 symmetry), EMD-22353 (SARS-CoV-2 3Q-2P-FL spike trimer with C1 symmetry), EMD-22354 (SARS-CoV-2 3Q-2P-FL spike dimer-of-trimers with C2 symmetry), EMD-22355 (SARS-CoV-2 3Q-2P-FL spike trimer-of-trimers with C1 symmetry), and EMD-22356 (SARS-CoV-2 3Q-FL spike trimer with C3 symmetry). The atomic models have been deposited at the Protein Data Bank with PDB IDs 7JJI (SARS-CoV-2 3Q-2P-FL spike trimer with C3 symmetry) and 7JJJ (SARS-CoV-2 3Q-2P-FL spike dimer-of-trimers with C2 symmetry). The vaccine construct was provided to the Scripps Research Institute under a Material Transfer Agreement with Novavax, Requests for this material should be addressed to Gale Smith at Novavax. Other materials are available from A.B.W. under a Material Transfer Agreement with the Scripps Research Institute.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1089/suppl/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References (*32*–*52*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

3 August 2020; accepted 13 October 2020
Published online 20 October 2020
10.1126/science.abe1502

CLIMATOLOGY

Abrupt shift to hotter and drier climate over inner East Asia beyond the tipping point

Peng Zhang[1,2], Jee-Hoon Jeong[1]*, Jin-Ho Yoon[3], Hyungjun Kim[4], S.-Y. Simon Wang[5], Hans W. Linderholm[2], Keyan Fang[6,2], Xiuchen Wu[7,8], Deliang Chen[2]

Unprecedented heatwave-drought concurrences in the past two decades have been reported over inner East Asia. Tree-ring–based reconstructions of heatwaves and soil moisture for the past 260 years reveal an abrupt shift to hotter and drier climate over this region. Enhanced land-atmosphere coupling, associated with persistent soil moisture deficit, appears to intensify surface warming and anticyclonic circulation anomalies, fueling heatwaves that exacerbate soil drying. Our analysis demonstrates that the magnitude of the warm and dry anomalies compounding in the recent two decades is unprecedented over the quarter of a millennium, and this trend clearly exceeds the natural variability range. The "hockey stick"–like change warns that the warming and drying concurrence is potentially irreversible beyond a tipping point in the East Asian climate system.

Global warming has led to a shift in the probability distribution of summer temperatures (*1*), causing more frequent summer heatwaves in the Northern Hemisphere midlatitudes (*2*, *3*). Diverse and complex regional or global feedback mechanisms determine the magnitude of changes in heatwave frequency and regionality (*4*–*9*). In addition to the emergence of heatwave-prone atmospheric stagnation, soil moisture deficit before or during droughts has been identified as a key factor exacerbating heatwaves through the land-atmosphere coupling (*10*–*14*). This drought-heatwave interaction is particularly pronounced in semiarid regions such as southern Europe, western North America, and inner East Asia (Mongolia and northern China), where the land-atmosphere coupling is strong (*15*–*20*).

Although increasing concentration of greenhouse gases may enhance soil moisture deficits and heatwave occurrences in a warmer climate (*21*–*23*), the extent to which summer heatwaves are affected by the global warming–induced soil moisture reduction in a long-term context remains unclear. In this study, we used tree-ring data to reconstruct both heatwave frequency and soil moisture variability in inner East Asia, centered over Mongolia, for the past 260 years. We found a robust tendency toward a hotter and drier climate, with a stronger coupling of heatwave and drought in recent decades that was not observed before the 1990s. This trend, which was found in both observations and reconstructions, is likely associated with an enhanced land-atmosphere coupling associated with persistent soil moisture deficit.

Inner East Asia, including Mongolia and its surroundings (Fig. 1), features arid and semiarid climates where annual precipitation is <300 mm. It is one of the hotspots showing the strongest warming in the latter part of the 20th century (*24*). The frequency of summer heatwaves in this region has increased significantly during the past two decades (Fig. 1A). Concurrently, soil moisture content has shown a significant drying trend (Fig. 1B), which is consistent with previous studies based on observation and land surface modeling (*25*–*27*) (also see fig. S1).

To determine whether this modern-era drying falls within or outside the range of long-term natural variability, we reconstructed summer heatwave and soil moisture variations based on independent tree-ring chronologies sampled in inner East Asia (see the supplementary materials).

[1]Faculty of Earth and Environmental Sciences, Chonnam National University, Gwangju, Korea. [2]Regional Climate Group, Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden. [3]School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology, Gwangju, Korea. [4]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan. [5]Department of Plants, Soils and Climate/Utah Climate Center, Utah State University, Logan, UT, USA. [6]Key Laboratory of Humid Subtropical Eco-Geographical Process (Ministry of Education), College of Geographical Sciences, Fujian Normal University, Fuzhou, China. [7]State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China. [8]Faculty of Geographical Science, Beijing Normal University, Beijing, China.
*Corresponding author. Email: jjeehoon@jnu.ac.kr

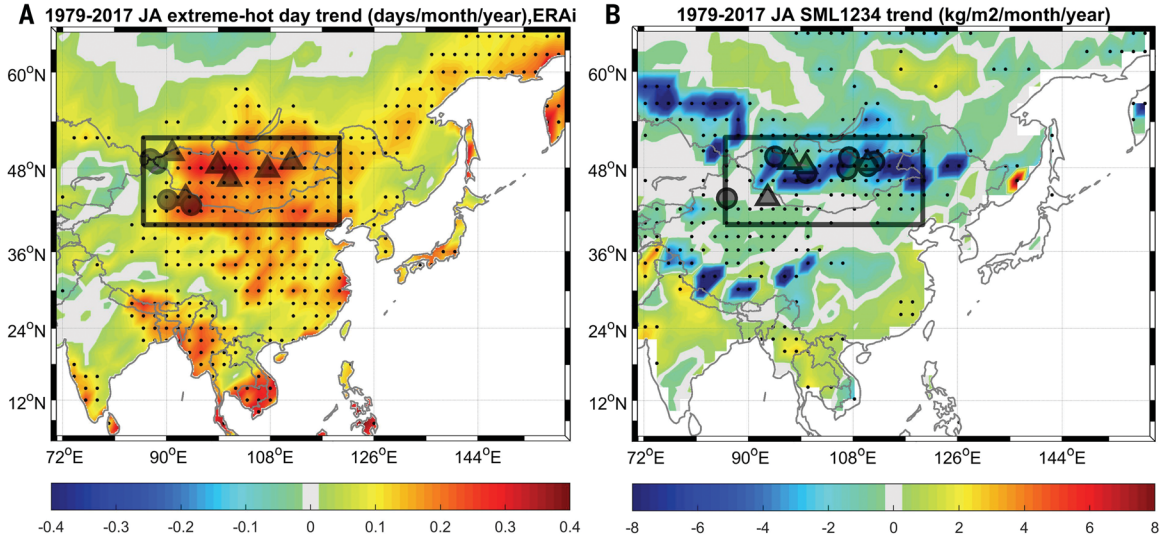


Fig. 1. Trends of hot day frequency and soil moisture content. (**A** and **B**) Trends of July-August mean extremely hot day frequency (A) and soil moisture content (B) in 1979–2017. Black frame marks the defined domain of inner East Asia. Black dots mark the grids where the trends are significant at the $P < 0.05$ level. Triangles and circles mark the locations where the tree-ring data were used to reconstruct the extremely hot day frequency and soil moisture variability on interannual and above-interannual time scale, respectively.

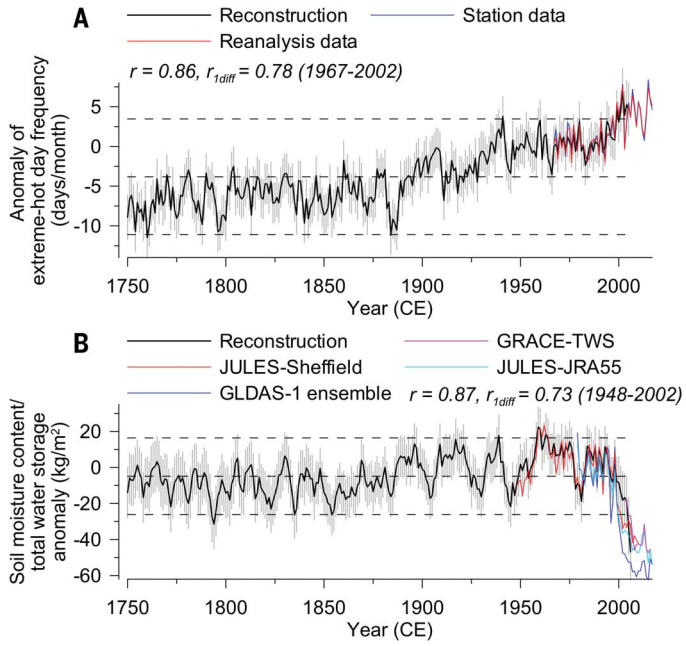


Fig. 2. Tree-ring width–based reconstructions. (**A** and **B**) Tree-ring width–based reconstructions of July-August mean extremely hot day frequency (A) and soil moisture anomaly (B) in 1750–2002 in inner East Asia (solid black curves) and their uncertainty (gray bars, defined as 2σ of reconstruction ensembles). The red and blue curves in (A) are the extremely hot day frequency variability derived from ERA interim and station datasets. The red, cyan, blue, and pink curves in (B) are the soil moisture anomaly derived from the JULES-Sheffield (1948–2010), JULES-JRA55 (1979–2017), and GLDAS-1 (ensemble mean, 1979–2017), and total water storage anomaly (relative to a value of −35kg/m^2) derived from the Gravity Recovery and Climate Experiment (GRACE) mission over 2002–2016 (for details, see the supplementary materials; note that there are data missing in August of 2013 and July of 2014), respectively. Dashed horizontal lines represent the long-term mean and ±2σ over the period 1750–2002. The correlation coefficients (*r*) are calculated between the reconstruction and the reanalysis data in (A) and between the reconstruction and JULES-Sheffield soil moisture in (B). r_{1diff} is the correlation coefficient of the first difference of time series.

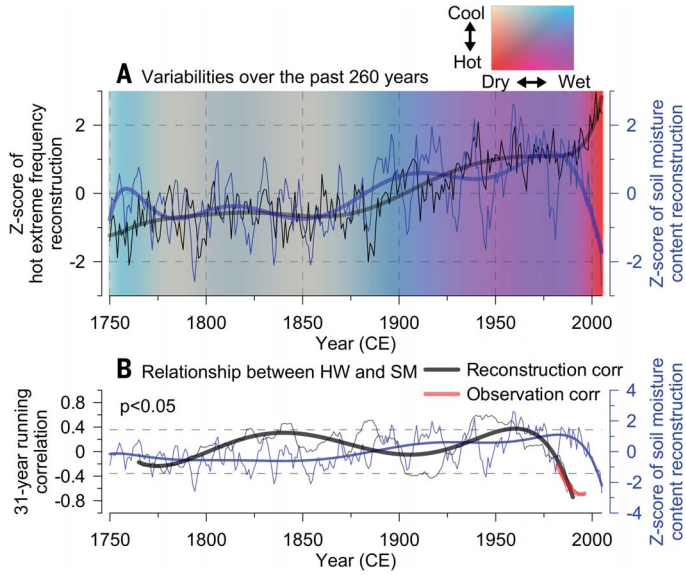


Fig. 3. Comparison of data over the past 260 years. Shown is a comparison between the July-August mean extremely hot day frequency variability and soil moisture variability in inner East Asia over the past 260 years. (**A**) Variability of the reconstructed extremely hot day frequency (black) and soil moisture content (blue) and their low-frequency variations over the past 260 years. (**B**) Thirty-one-year running correlation between hot extreme frequency and soil moisture over 1750–2005 based on the reconstructions (black curves) and over 1967–2011 based on reanalysis and JULES-Sheffield data (red curve), where the light dashed lines indicate the $P < 0.05$ significance level. The blue curves are the soil moisture reconstruction; bold curves are their low-frequency variations.

The heatwave reconstruction in Fig. 2A, explaining 74% of the variance in the instrumental heatwave record for the period 1967–2002, shows a weak long-term change before the 20th century. Subsequently, it displays a gradual increase until the 1950s, followed by a phase of pronounced increase since the 1970s. The elevated heatwave frequency in recent decades appears to exceed that of natural variability in the context of the past 260 years. This result echoes previous summer temperature reconstructions in the same region (*28*).

The soil moisture reconstruction in Fig. 2B, which explains 76% of the variance in observed soil moisture over the period 1948–2002, shows a negligible trend until the end of the 20th century, after which its accelerated decline becomes evident. It should be noted that the reconstructed and simulated soil moisture records reveal periods of reductions in the 1900s, 1920s, 1940s, and 1970s that are in good agreement with documented megadrought events in Mongolia (*29*–*31*) (also see fig. S2). However, the post–20th-century reduction of soil moisture reached a record low, plummeting below the −2 SD level. This is consistent with other soil moisture datasets and a satellite-based total water storage observation (Fig. 2B; also see the supplementary materials), suggesting an abrupt and unprecedented drying trend.

When the reconstructions of heatwave frequency and soil moisture are compared (Fig. 3A), a substantial shift toward a "drier-hotter" regime from the late 20th century onward is apparent. This latter regime is unusual because before the 1950s, the two variables shared a more or less coherent long-term change in which both were increasing. In summer, a dry soil condition can exert heat into the lower atmosphere through sensible heat flux while suppressing evaporative cooling, so positive feedback between droughts and near-surface warming can be triggered by the soil moisture deficit reaching certain thresholds. Over multidecadal time scales, heatwave frequency and soil moisture exhibit a weak anticorrelation ($r = -0.05$, $P > 0.1$) until the mid-20th century (thick lines in Fig. 3A). In the past 260 years, only the recent decades show a significant anticorrelation between heatwave frequency and soil moisture, alongside a radical decline in soil moisture (Fig. 3B). The observation of the compound warming and drying is unprecedented in inner East Asia, implying a regime change in the regional land-atmosphere coupling. The upshot of this result is that the regional near-surface air may have increased its sensitivity to soil moisture fluctuation. A series of recent heatwave events in Europe (*32*) and North America (*18*) have revealed this phenomenon, suggesting that the semiarid climate of this region has

entered a new regime in which soil moisture no longer mitigates anomalously high air temperature.

To attribute the apparent increase in the heatwave-drought concurrence to regional climate variability, we investigated the changes in the large-scale atmospheric circulation and the impact of the land feedback. Figure 4, A and B, show composite maps of geopotential height (GPH) and wind anomalies at 500 hPa for the years of unusually high heatwave frequency for two periods, 1979–1998 and 1999–2017. In both periods, a large-scale wave train spanning northern Eurasian continent is revealed, with prominent high-pressure anomalies over the study area, which has been noted previously (*33*). In the latter period, the pattern of amplified high-pressure anomalies is consistent with the observed trends in the summer upper-air circulations (*34*).

Figure 4, C and D, show the vertical sections of the high-pressure ridge across inner East Asia in terms of GPH and air temperature, depicting the warm-core feature of the drought-heatwave accompanying the high-pressure anomaly that has intensified in the latter period. Even though that amplified upper-tropospheric wave train (*34*) has contributed to this feature, the fact that increased land surface warming also extends upward into the middle troposphere in Fig. 4D suggests an intensified land-atmosphere coupling. This claim is linked to the aforementioned soil moisture deficit trend causing the drier land surface to produce stronger sensible heat flux into the lower atmosphere, which thickens the planetary boundary layer and therefore amplifies the upper-air high pressure (*35*). As part of the positive feedback, the strengthened anticyclone aloft further increases surface warming through enhanced downward shortwave radiation. Recent megaheatwaves in Europe (e.g., those in 2003 and 2010) have been attributed to similar land-atmospheric couplings and atmospheric feedbacks that exacerbate the heatwaves (*11–13*, *36–39*).

The recent change in land-atmosphere coupling can be quantified using the π diagnostic of soil moisture–temperature coupling (*13*) (see the supplementary materials). This metric quantifies the relation between near–surface temperature anomalies and soil moisture deficits in the land surface energy balance. During the past two decades, the strength of the land-atmosphere coupling has been profoundly enhanced over a large part of Asia, including southeastern Asia, China, and Mongolia (Fig. 5, A and B). The most extensive change occurred over Mongolia and northern China (Fig. 5C),

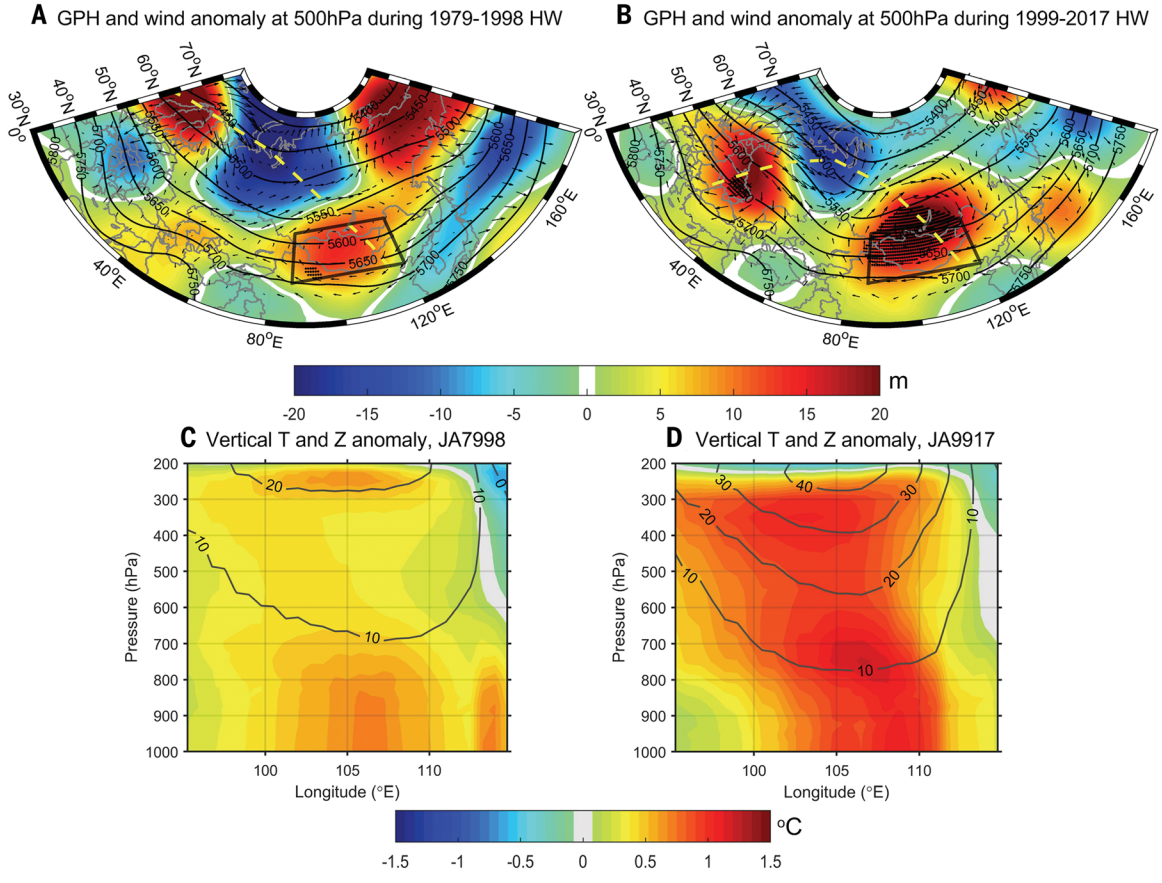


Fig. 4. Composite map. (**A** and **B**) Composite map of July-August mean geopotential height and wind anomalies at 500 hPa for the years when extremely hot day frequency was anomalously high (>0.4σ; the results are not sensitive to the threshold) over the period 1979–1998 [the selected years are 1980, 1987, 1989, 1991, 1996, and 1997 (A)] and 1999–2017 [the selected years are 1999, 2001, 2002, 2007, 2010, 2015, and 2016 (B)] and their corresponding vertical structures of temperature and geopotential height. Contour lines show climatology of the geopotential height, and black stripes mark the areas where pressure change is significant at the $P < 0.05$ level calculated based on the Monte Carlo method with 1000 iterations. Yellow dashed line denotes the track of the wave train pattern. Black frame marks the defined domain of inner East Asia. (**C** and **D**) Lower panels show corresponding vertical structures of temperature (shading) and geopotential height anomalies (contour lines in meters) along the wave train tracks in inner East Asia over the period of 1979–1998 (C) and 1999–2017 (D). The geopotential height, wind, and temperature data are derived from ERA interim monthly dataset with 0.75° latitude × 0.75° longitude resolution and 23 pressure levels from 200 up to 1000 hPa.

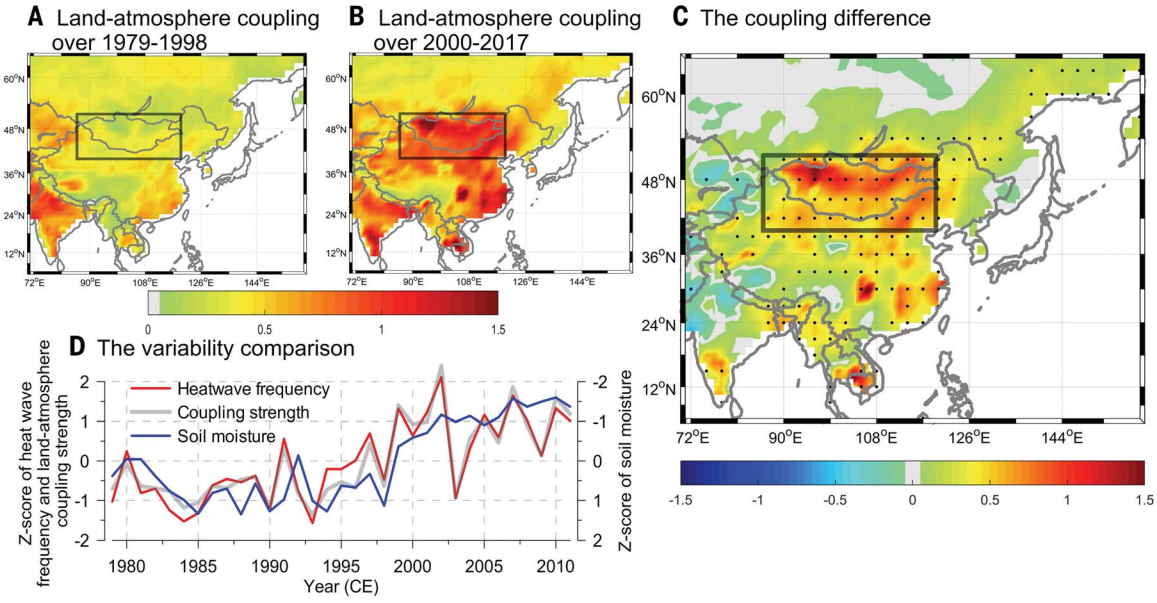


Fig. 5. Land-atmosphere coupling strength. (**A** to **C**) July-August land-atmosphere coupling strength over the period 1979–1998 (A) and 2000–2017 (B) and their differences [(B) minus (A)] (C) and comparison with soil moisture and heatwave frequency variabilities. Black dots mark the grids where differences are significant at the $P < 0.05$ level. The coupling strength is calculated based on the ERA interim daily dataset with 1.5° latitude × 1.5° longitude spatial resolution. (**D**) Interannual variability of the coupling strength (gray), hot extreme frequency (red), and soil moisture (blue, reversed value in *y*-axis) over the domain.

where a robust trend of drier-hotter air under the strengthened anticyclone was observed. The temporal variation of the coupling strength is almost identical to the heatwave frequency and also matches the continuous drying trend of soil moisture (Fig. 5D). The uptrend of the association between heatwave frequency and land-atmosphere coupling follows the persistent change in soil moisture deficit. This feature is not as robust over southern and coastal East Asia, where the local climate is largely influenced by the East Asia summer monsoon (*40*).

Tree-ring and observational evidence show that the increased frequency of summer heatwaves and the concurrent increase in severe droughts over inner East Asia in recent decades are unique in a 260-year context. Given the enhanced land-atmosphere interaction associated with the continuously drying soil, the land-atmosphere coupling over inner East Asia might have entered a different regime in which the soil moisture variability has become too small and the sensitivity of evaporation to more dryness has weakened (fig. S3). Such a regime shift would explain why heatwaves and droughts have become tightly coupled and can quickly enhance each other, leading to an increase in concurrent events of heatwave and drought. Although our analysis does not include a future projection, it can be expected that the "regime shift," which is supported by multiple lines of evidence, gives way to a new state in which there are stronger negative associations between heatwaves and soil moisture that are expected to persist under a warmer climate. The implication of these findings is that compound extremes of summer heatwaves and heatwave-droughts may occur more frequently and potentially become more severe in inner East Asia, and that corresponding mitigation efforts should be undertaken to cope with the associated increase in the risk of livestock and rangeland crisis in Mongolia.

REFERENCES AND NOTES

1. K. A. McKinnon, A. Rhines, M. P. Tingley, P. Huybers, *J. Geophys. Res. Atmos.* **121**, 8849–8868 (2016).
2. N. Christidis, G. S. Jones, P. A. Stott, *Nat. Clim. Chang.* **5**, 46–50 (2015).
3. D. Coumou, S. Rahmstorf, *Nat. Clim. Chang.* **2**, 491–496 (2012).
4. R. García-Herrera, J. Díaz, R. M. Trigo, J. Luterbacher, E. M. Fischer, *Crit. Rev. Environ. Sci. Technol.* **40**, 267–306 (2010).
5. J. Cohen *et al.*, *Nat. Geosci.* **7**, 627–637 (2014).
6. Q. Tang, X. Zhang, J. A. Francis, *Nat. Clim. Chang.* **4**, 45–50 (2014).
7. N. Pepin *et al.*, *Nat. Clim. Chang.* **5**, 424–430 (2015).
8. D. Coumou, J. Lehmann, J. Beckmann, *Science* **348**, 324–327 (2015).
9. Z. Wu, P. Zhang, H. Chen, Y. Li, *Clim. Dyn.* **46**, 3405–3417 (2016).
10. R. Vautard *et al.*, *Geophys. Res. Lett.* **34**, L07711 (2007).
11. E. M. Fischer, S. I. Seneviratne, P. L. Vidale, D. Lüthi, C. Schär, *J. Clim.* **20**, 5081–5099 (2007).
12. E. M. Fischer, S. I. Seneviratne, D. Lüthi, C. Schär, *Geophys. Res. Lett.* **34**, L06707 (2007).
13. D. G. Miralles, A. J. Teuling, C. C. Van Heerwaarden, J. Vilà-Guerau de Arellano, *Nat. Geosci.* **7**, 345–349 (2014).
14. J. Zscheischler, S. I. Seneviratne, *Sci. Adv.* **3**, e1700263 (2017).
15. R. D. Koster *et al.*, *Science* **305**, 1138–1140 (2004).
16. M. Hirschi *et al.*, *Nat. Geosci.* **4**, 17–21 (2011).
17. S. D. Schubert, H. Wang, R. D. Koster, M. J. Suarez, P. Y. Groisman, *J. Clim.* **27**, 3169–3207 (2014).
18. R. Saini, G. Wang, J. S. Pal, *J. Hydrometeorol.* **17**, 2191–2207 (2016).
19. T. Cowan *et al.*, *J. Clim.* **30**, 2437–2461 (2017).
20. E. Erdenebat, T. Sato, *Atmos. Sci. Lett.* **17**, 135–140 (2016).
21. G. Zittis, P. Hadjinicolaou, M. Fnais, J. Lelieveld, *Reg. Environ. Change* **16**, 1863–1876 (2016).
22. J. Lelieveld *et al.*, *Clim. Change* **137**, 245–260 (2016).
23. V. Humphrey *et al.*, *Nature* **560**, 628–631 (2018).
24. D. L. Hartmann, A. Klein Tank, M. Rusticucci, L. Alexander, S. Broennimann, Y. A.-R. Charabi, F. Dentener, E. Dlugokencky, D. Easterling, A. Kaplan, B. Soden, P. Thorne, M. Wild, P. Zhai, J. Hurrell, J. A. Marengo Orseni, F. Tangang, P. Viterbo, "Observations: atmosphere and surface" in *Climate Change 2013 – The Physical Science Basis: Contributions of Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, D. Qin, G.-K. Plattner, M. M. B. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley, Eds. (Cambridge Univ. Press, 2013), pp. 187–201.
25. M. Rodell *et al.*, *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
26. M. Rodell *et al.*, *Nature* **557**, 651–659 (2018).
27. B. R. Scanlon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1080–E1089 (2018).
28. E. R. Cook *et al.*, *Clim. Dyn.* **41**, 2957–2972 (2013).
29. E. R. Cook *et al.*, *Science* **328**, 486–489 (2010).
30. N. Pederson, A. E. Hessl, N. Baatarbileg, K. J. Anchukaitis, N. Di Cosmo, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4375–4379 (2014).
31. A. E. Hessl *et al.*, *Sci. Adv.* **4**, e1701832 (2018).
32. S. I. Seneviratne, D. Lüthi, M. Litschi, C. Schär, *Nature* **443**, 205–209 (2006).
33. M. P. Rao *et al.*, *Environ. Res. Lett.* **10**, 074012 (2015).
34. S. Y. Wang, R. E. Davies, R. R. Gillies, *J. Geophys. Res. Atmos.* **118**, 11059–11074 (2013).
35. S. Y. S. Wang *et al.*, *J. Geophys. Res. Atmos.* **120**, 8804–8816 (2015).
36. R. Lorenz, E. B. Jaeger, S. I. Seneviratne, *Geophys. Res. Lett.* **37**, L09703 (2010).
37. M. Stéfanon, P. Drobinski, F. D'Andrea, C. Lebeaupin-Brossier, S. Bastin, *Clim. Dyn.* **42**, 1309–1324 (2014).
38. M. Hauser, R. Orth, S. I. Seneviratne, *Geophys. Res. Lett.* **43**, 2819–2826 (2016).
39. E. Seo *et al.*, *Clim. Dyn.* **52**, 1695–1709 (2019).
40. D. Yihui, J. C. L. Chan, *Meteorol. Atmos. Phys.* **89**, 117–142 (2005).

41. P. Zhang, JULES-Sheffield soil moisture 1948-2010, figshare (2020); https://figshare.com/articles/dataset/JULES-Sheffield_soil_moisture_1948-2010/13040534.
42. P. Zhang, JULES-JRA55 soil moisture 1979-2017, figshare (2020); https://figshare.com/articles/dataset/JULES-JRA55_soil_moisture_1979-2017/13040540.

ACKNOWLEDGMENTS

We thank M.-S. Kim for collection and preprocessing of tree ring data; K.-M. Kim for providing JULES-Sheffield and JULES-JRA55 soil moisture data; colleagues and friends for valuable discussions on the results of this study; and the scientists who shared their valuable tree ring width data from ITRDB used in this study. **Funding:** This work was supported by the Korea Meteorological Administration Research and Development Program under grant no. KMI2018-07010; the Strategic Priority Research Program of the Chinese Academy of Sciences under grant no. XDA20060400; and by the Swedish Research Council for Sustainable Development (FORMAS) under grant no. 2018-02858. H.W.L. was supported by the Swedish Research Council (VR; grant nos. 2012-05246 and 2015-04031). This is a contribution to the Swedish BECC and MERGE. H.K. acknowledges Grant-in-Aid for Specially Promoted Research 16H06291 and 18KK0117 from the Japan Society for the Promotion of Science. S.-Y.S.W. is supported by NSF p2c2 award no. 1903721. **Author contributions:** P.Z. and J.-H.J. designed the research, conducted analyses, and wrote most of the manuscript content. S.-Y.S.W., J.-H.Y., and H.K contributed to analysis and did most of the report writing. H.W.L, K.F., X.W., and D.C. contributed to analysis and report writing. All authors discussed the study results and reviewed the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data used in this study are freely accessible. Tree-ring data can be downloaded from International Tree-Ring Data Bank at https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring (see the materials and methods for details about the tree-ring data used in this study). GLDAS soil moisture data are freely available online at https://ldas.gsfc.nasa.gov/gldas or through an interactive online application Giovanni (https://giovanni.gsfc.nasa.gov/giovanni/#). JULES-Sheffield (*41*) and JULES-JRA55 (*42*) soil moisture data can be found at figshare. ERA-interim and ERA-40 daily reanalysis datasets are available at https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets. GRACE-TWS data are available at https://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons/. The Global Historical Climate Network Daily (GHCN-D) dataset can be downloaded from https://www.ncdc.noaa.gov/ghcn-daily-description.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1095/suppl/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References (*43*–*73*)

17 February 2020; accepted 12 October 2020
10.1126/science.abb3368

SYNTHETIC BIOLOGY

Precise T cell recognition programs designed by transcriptionally linking multiple receptors

Jasper Z. Williams[1,2,3], Greg M. Allen[1,2,3,4], Devan Shah[1,2,3], Igal S. Sterin[1,2,3], Ki H. Kim[1,2,3], Vivian P. Garcia[1,2,3], Gavin E. Shavey[1,2,3], Wei Yu[1,2,3], Cristina Puig-Saus[5], Jennifer Tsoi[5], Antoni Ribas[5], Kole T. Roybal[1,2,3]*, Wendell A. Lim[1,2,3,6]†

Living cells often identify their correct partner or target cells by integrating information from multiple receptors, achieving levels of recognition that are difficult to obtain with individual molecular interactions. In this study, we engineered a diverse library of multireceptor cell-cell recognition circuits by using synthetic Notch receptors to transcriptionally interconnect multiple molecular recognition events. These synthetic circuits allow engineered T cells to integrate extra- and intracellular antigen recognition, are robust to heterogeneity, and achieve precise recognition by integrating up to three different antigens with positive or negative logic. A three-antigen AND gate composed of three sequentially linked receptors shows selectivity in vivo, clearing three-antigen tumors while ignoring related two-antigen tumors. Daisy-chaining multiple molecular recognition events together in synthetic circuits provides a powerful way to engineer cellular-level recognition.

Precise cell-cell recognition is critical throughout biology. Developing cells, neurons, and immune cells must identify their correct partner or target cells with extraordinary specificity. Such precise cellular recognition is mediated by multiple surface receptors that function in an integrated, combinatorial manner (*1*). Precise cellular recognition is also a central goal in treating diseases such as cancer, but current therapeutic approaches [antibodies, bispecific antibodies, or chimeric antigen receptors (CARs)] primarily rely on identifying malignant cells by interaction with a single cancer-associated antigen (Fig. 1A, left). These approaches, although powerful in some cases, are limited, because most cancer cells lack a single antigen whose expression cleanly distinguishes them from healthy cells (*2*). CAR T cells—which are engineered to redirect cytotoxic activity against a targeted extracellular antigen—have caused toxic on-target off-tumor cross-reactions with normal tissues in clinical trials (*3*–*8*), because many potential CAR tumor antigen targets in solid cancers are also found in normal epithelial tissues. Bioinformatic analyses, however, indicate that combinatorial antigen pattern recognition could enable more-specific cancer targeting (*9*–*11*). Thus, we aim to design tailored recognition programs that exploit discriminatory combinatorial features of target disease cells.

Engineered cell therapies are an opportune platform for constructing combinatorial recognition programs, because, in principle, multiple molecular recognition systems could be assembled into higher-order cellular recognition programs through coordinated intracellular regulation (Fig. 1A, right). We previously designed synthetic Notch (synNotch) receptors that link an antigen recognition domain [e.g., a single-chain antibody (scFv)] to a transcriptional regulatory domain, which, only upon target antigen binding, induces expression of a user-defined, genetically encoded payload (*12*). By using a synNotch receptor to drive expression of a CAR, directed against a second antigen, we engineered a two-antigen AND gate in which the circuit leads to T cell activation and target cell killing only when the target cell expresses both the synNotch and CAR antigens on its surface (*13*). This AND gate, however, represents only a small fraction of the possible combinatorial recognition circuits that could be engineered. In this study, we used multiple synNotch receptors as flexible regulatory connectors to daisy-chain, or link together, a broader range of receptors and outputs into circuits that can achieve a more complete set of cell-cell recognition behaviors (Fig. 1A).

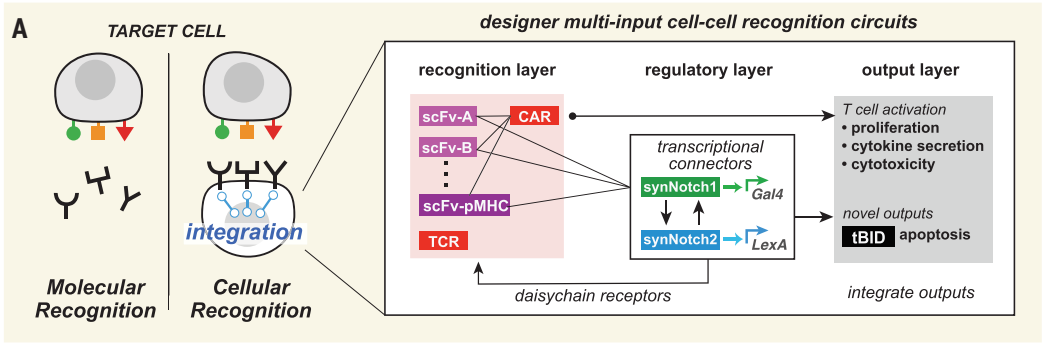
One simple but impactful recognition program would integrate extracellular and intracellular recognition (Fig. 1B). CAR T cells are generally restricted to recognizing extracellular antigens, which limits their ability to read and harness the ~75% of the proteome that is intracellular (including many cancer antigens) (*14*). We designed such an external-internal recognition circuit to recognize melanoma cells. The intracellular melanocyte antigen MART1 is presented on the cell surface by major histocompatibility complexes (MHCs). A T cell receptor (TCR) directed against MART1 showed antimelanoma effects in clinical trials, but anti-MART1 T cells also induced severe rash or vitiligo, uveitis, and hearing loss due to on-target off-tumor attack of normal melanocytes that also express MART1 (Fig. 1C) (*6*). RNA sequencing data from 53 human melanoma cell lines (*15*, *16*) shows that many patient-derived melanoma lines also overexpress the surface antigen MET (hepatocyte growth factor receptor) (~50% of samples with transcripts per million > 5; fig. S1, A and B), which is only weakly expressed in normal adult melanocytes (*17*–*19*). Thus, a recognition circuit that requires MET and MART1 could be used to distinguish between melanoma cells and normal melanocytes. We constructed an anti-MET synNotch→anti-MART1 [human leukocyte antigen A2 (HLA-A2)] TCR circuit (fig. S1, C and D) and showed that, unlike standard constitutive anti-MART1 T cells, $CD8^+$ (killer) T cells expressing this circuit did not show T cell activation, cytokine production, proliferation, or killing directed against normal adult melanocytes (Fig. 1D and fig. S1, E and F). In contrast, when challenged with MET^+ (M202) patient-derived melanoma cell lines, the MET→MART1 TCR T cells showed strong proliferation and cytotoxicity against these $MET^+/MART1^+$ tumor cells (Fig. 1E). MET^- (M262) melanoma cell lines, however, did not activate these T cells. MART1 TCR T cells and MET→MART1 TCR T cells had similar cytotoxic capability against dual-antigen-positive targets, demonstrating that the circuit's increased precision does not sacrifice efficacy (Fig. 1E). Thus, synNotch→TCR circuits can function robustly, and they offer a general solution for improving the specificity of therapeutic TCRs [such as MART1 or MAGEA3 (melanoma-associated antigen 3)] (fig. S1, G to I) with toxic cross-reactivities.

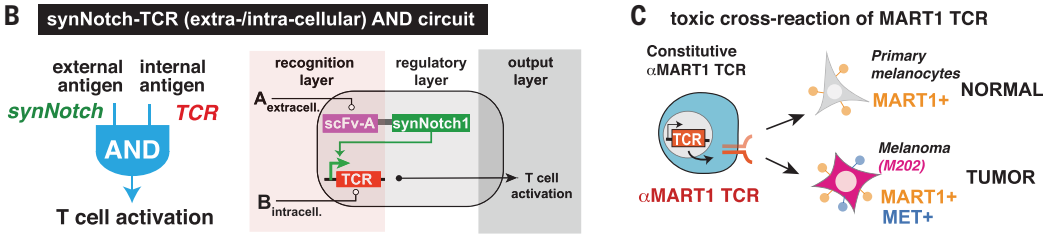
scFvs raised against specific peptide-MHC (pMHC) complexes have been used to engineer CARs that detect MHC-presented intracellular antigens (*20*). We used these anti-pMHC scFvs as recognition heads for synNotch receptors (referred to as "inNotch" receptors) (fig. S2A). As an example, we validated that an inNotch receptor with an HLA-A2/AFP (alpha-fetoprotein) recognition domain (*20*) drives induction of a blue fluorescent protein (BFP) reporter in an antigen-dependent manner (fig. S2B). We tested whether the anti-HLA-A2/AFP inNotch receptor could be incorporated into AND gates regulating either CAR or TCR expression. T cells engineered with the anti-AFP inNotch controlling either an anti-HER2 (human epidermal growth factor receptor 2) CAR (internal→external circuit) or anti-HLA-A2/NY-ESO1 (cancer/testis antigen 1)

[1]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA. [2]Cell Design Institute, University of California, San Francisco, San Francisco, CA 94158, USA. [3]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94158, USA. [4]Department of Medicine, University of California, San Francisco, San Francisco, CA 94158, USA. [5]Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA. [6]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94158, USA.
*Present address: Department of Microbiology and Immunology, Helen Diller Family Comprehensive Cancer Center, Parker Institute for Cancer Immunotherapy, Chan Zuckerberg Biohub, University of California, San Francisco, CA 94158, USA.
†Corresponding author. Email: wendell.lim@ucsf.edu

TCR (internal→internal circuit) acted as AND gates and specifically killed the intended dual-positive target cells (fig. S2, C to E).

A second broad and general class of combinatorial recognition strategy involves the incorporation of negative antigen inputs. Recent bioinformatic analysis indicates that negative information (i.e., NOT gate antigens expressed in normal tissues but not in the target cancer cells) could be powerful in preventing toxic cross-reactions (*11*). Earlier studies have explored the design of an inhibitory-CAR receptor (iCAR) (*21*), which fuses NOT antigen recognition domains with intracellular signaling domains of immune inhibitory receptors to





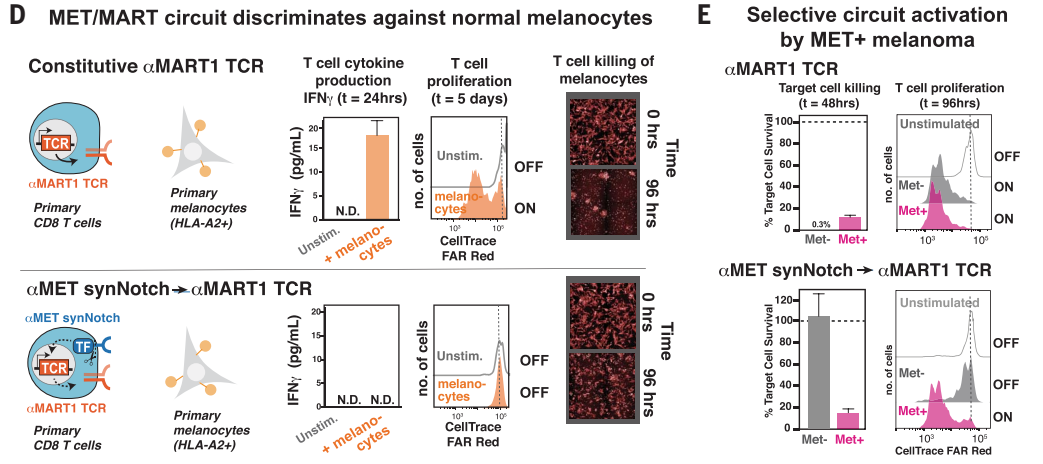


Fig. 1. SynNotch modules can be used to daisy-chain multiple receptor systems to design diverse multi-input cell-cell recognition circuits, including combinatorial recognition of internal and external antigens. (**A**) Recognition of disease cells currently focuses on individual molecular recognition events (e.g., antibodies and CARs). In principle, cell therapies can be engineered to integrate multiple molecular recognition events into a composite cellular recognition program. Transcriptional connectors, such as synNotch receptors, provide a modular way to functionally link multiple receptors. Here the recognition layers are composed of scFvs or similar antigen binding molecules used by CAR or synNotch receptors. SynNotch receptors act as a regulatory layer inducing antigen-triggered expression of synthetic transcription factors (in this case, GAL4 or LEXA) that control cell function within the output layer. (**B**) Intracellular and extracellular antigen information could be integrated into a T cell recognition program by using a synNotch receptor that drives expression of a TCR. (**C**) MART1 TCR is example of TCR that requires additional inputs to eliminate toxic cross-reactions—the TCR targets both melanomas and normal melanocytes (MART1 is also referred to as MLANA). We find that a subpopulation of patient-derived melanoma cells also express additional tumor-associated antigens, such as MET (fig. S1A). (**D**) Experimental validation that anti-MET synNotch→anti-MART1 TCR circuit improves discrimination of anti-MART1 TCR against normal melanocytes. Primary human CD8 T cells were engineered with a constitutive anti-MART1 TCR or an anti-MET Gal4-VP64 synNotch→anti-MART1 TCR circuit. These T cells were cocultured at 1:1 effector: target cell ratio (E:T) for varying times with primary human melanocytes that were HLA-A2$^+$/MART1$^+$/MET$^-$. The supernatant from these cocultures was collected at 24 hours and analyzed for interferon-γ (IFN-γ) levels via an enzyme-linked immunosorbent assay. After 5 days of coculture, remaining cells were collected and T cell proliferation was determined by measuring CellTrace dye dilution via flow cytometry. synNotch→TCR circuit T cells showed no activation in response to the normal melanocytes (cytokine production: $n = 3$, error bars are SD; N.D., not detected; $^*P < 0.01$; proliferation: representative of three independent experiments). In separate cell-killing experiments, melanocytes were labeled with CellTrace Far Red, and confocal microscopy images were taken after coculture with constitutive or synNotch-induced MART1 TCR T cells. Representative images show that constitutive anti-MART1 TCR T cells cleared the primary melanocytes, whereas neither untransduced nor synNotch→TCR circuit T cells showed any reactivity (see fig. S1E for larger images). (**E**) The anti-MET synNotch→anti-MART1 TCR circuit is still effective at recognizing and killing MET$^+$MART1$^+$ melanoma cells. T cells with constitutive or synNotch-induced MART1 TCR were cocultured with MET$^-$ (M262) as well as the MET$^+$ (M202) cell lines at an E:T of 1:1. Cytotoxicity was measured by flow cytometry as the number of tumor cells remaining compared with untransduced T cells after 48 hours, and proliferation was measured by dilution of CellTrace Far Red after 5 days. $^*P < 0.01$. Representative of three independent experiments.

override CAR killing. However, these iCAR receptors require very high levels of expression of both the iCAR and the NOT antigen and have not proven to be highly robust. Thus, it appears that directly overriding CAR activity at the membrane-proximal signaling level is challenging (*21*).

An alternative strategy for engineering a NOT gate antigen input might be to override T cell function downstream from CAR signaling. For example, potent CAR T cell function requires both T cell activation and proliferation. Thus, an effective NOT gate strategy could be to direct negative inputs to induce apoptosis, thereby blocking critical T cell expansion. We therefore constructed a synNotch circuit that activated expression of the proapoptotic factor tBID (truncated BH3-interacting domain death agonist) (fig. S3A). This circuit, which we refer to as OFF-Notch, rapidly induced T cell death when stimulated (fig. S3, B to E). To incorporate this OFF-Notch receptor as a NOT gate in a three-antigen circuit that also includes an AND gate requires two orthogonal synNotch platforms with distinct transcriptional regulatory domains. We found that GAL4-VP64 and LEXA-VP64 synNotch receptors could yield robust independent responses with no cross-talk when coexpressed in primary human T cells (fig. S4).

With these components in hand, we combined an anti-HER2 NOT gate with an AND gate [recognizing the model antigens GFP (green fluorescent protein, surface-expressed) and CD19] (Fig. 2A). The composite circuit showed the predicted target selectivity—when $CD8^+$ T cells expressing the circuit were mixed with K562 tumor cells engineered to express different combinations of the antigens HER2, surface GFP, and CD19 (fig. S5), all tumor cells expressing HER2 (NOT antigen) were spared (Fig. 2B and fig. S6B). In contrast, only those target cells expressing the on-target antigen combination of surface GFP, CD19, and NOT HER2 were killed. Selectivity was maintained at the level of T cell proliferation—$GFP^+/CD19^+$ tumor cells induced potent T cell expansion (3.7 ± 0.2 divisions), but this proliferation was significantly blocked by target cells that also expressed HER2 (in addition to $GFP^+/CD19^+$) (1.9 ± 0.1 divisions, $P = 0.0002$) (Fig. 2C). Thus, we were able to build a negative regulatory circuit that counteracts CAR activity at a downstream level (i.e., cell survival), effectively blocking both killing and proliferation responses that play a key role in on-target off-tumor toxicity. For clinical application of this circuit, local induction of cell death must be carefully balanced with sufficient expansion in targeted tumor and CAR T cell efficacy, and it may require induction of alternative payloads that have a smaller effect on T cell survival (fig. S3E).

Interestingly, this NOT gate worked best when counteracting a positive AND gate circuit (synNotch→CAR) rather than when counteracting a constitutively expressed CAR (fig. S6A). The delay required for antigen-induced tBID expression (fig. S3, C and D) is likely too long to effectively counteract the immediate signaling of a constitutively expressed CAR. Placing CAR expression under analogous synNotch control likely works better because the positive and negative branches of the circuit are dynamically well matched—both the CAR and tBID are induced at similar time scales, allowing more even competition between the pathways. Thus, matching kinetics is likely a critical design principle for circuits incorporating both positive and negative regulation.

Multiantigen AND gates—circuits that require three or more antigens to trigger T cell recognition and killing—are a third class of recognition strategy. We reasoned that extensible AND gate circuits could be built by combining multiple orthogonal synNotch receptor platforms. To construct a three-input AND gate, we could use our Gal4- and LexA-based synNotch platforms to recognize the first and second priming antigens and a CAR to recognize the third antigen. These receptor components can be linked in different configurations, akin to electronic circuits: in series (where

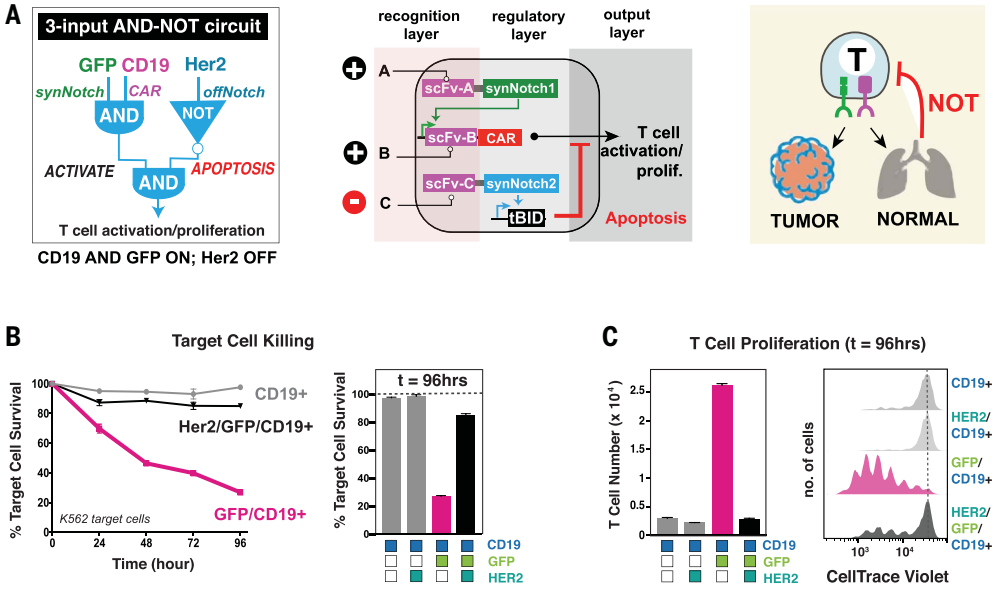


Fig. 2. Engineering recognition circuits that combine positive and negative logic by using synNotch modules that drive opposing outputs of activation and apoptosis. (**A**) Design of a three-input circuit integrating an AND gate with a NOT gate to control T cell activation. The anti-GFP LEXA synNotch induces expression of the anti-CD19 CAR, and the orthogonal anti-HER2 GAL4 synNotch induces expression of proapoptosis protein tBID ("offNotch"). These T cells must first bind surface GFP to induce CAR expression and should only activate and kill target cells if the T cells sense both surface GFP and CD19. If the T cells sense HER2, then tBID induction will trigger apoptosis, which should deplete the effector T cell population and serve as an antigen-dependent NOT gate to prevent target cell killing. (**B**) Quantification of replicate target cell killing data gathered via flow cytometry after 1:1 E:T cocultures of varied times between engineered CD8 T cells with the circuit shown in (A) and K562 target cells expressing different combinations of CD19, surface GFP, and HER2. Although in the first 24 hours the T cells kill some of the HER2/GFP/$CD19^+$ off-target cells, there is no further killing at later time points. In contrast, the T cells continue to kill the GFP/$CD19^+$ on-target cells over the 96-hour time course ($n = 3$, error bars are SD), $^*P < 0.01$ compared with $CD19^+$ control. Representative of three independent experiments. (**C**) Histograms of T cells stained with CellTrace Violet dye in the coculture experiment described in (B) and processed via flow cytometry after 96 hours. After 96 hours, cell counts (left) and CellTrace plots (right) show proliferation of T cells only in the presence of GFP/$CD19^+$ target cells and NOT HER2/GFP/$CD19^+$ target cells (representative of at least three independent experiments) ($n = 3$, error bars are SD), $^*P < 0.01$ compared with $CD19^+$ control. See figs. S3 to S6 for more details on construction and analysis of this circuit and its components.

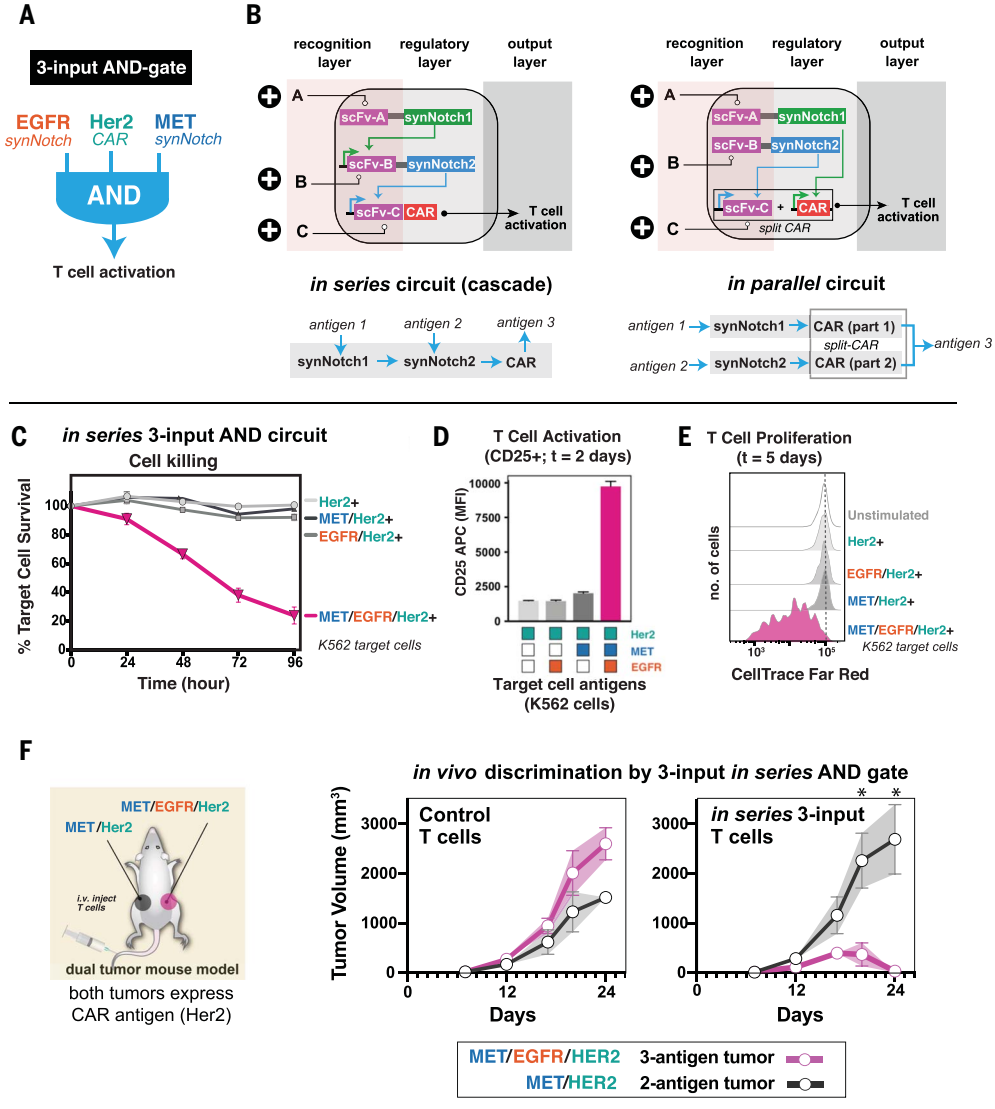


Fig. 3. Multi-synNotch circuits: Design and in vivo validation of ultraprecise three-input AND gates. (**A**) Target design of a three-input AND gate circuit (must have EGFR, HER2, and MET for activation). (**B**) Two alternative design schemes for using synNotch and CAR components to build three-antigen AND gate recognition circuits. Both schemes use two synNotch receptors, but akin to an electrical circuit, the receptors can be functionally arranged to be in-series or in-parallel. The in-series three-input circuit has a first synNotch receptor (recognizes antigen 1) that, when activated, induces expression of a second synNotch receptor (recognizes antigen 2), which in turn, when activated, induces expression of a CAR (recognizes antigen 3, the killing antigen). In contrast, the in-parallel circuit uses the two different synNotch receptors (recognize antigens 1 and 2) to each induce components of a split CAR, which only when coexpressed will recognize antigen 3 (killing antigen). See figs. S7 to S9 for details on construction and testing of both of these circuits and their components. (**C**) Selective killing by CD8$^+$ T cells transduced with the in-series three-input AND gate circuit. T cells with the full EGFR synNotch 1→MET synNotch 2→HER2 CAR circuit only killed K562 cells expressing all three antigens (pink line). No killing was observed of target cells with two or fewer of the antigens. Killing was measured by flow cytometry every 24 hours and normalized to effect of UnT CD8$^+$ T cells (n = 3, error bars are SD). Representative of three independent experiments. (**D**) Selective T cell activation, as measured by CD25 staining of T cells [as in (C)], was only seen when T cells were stimulated with K562 cells expressing all three antigens. (**E**) Selective proliferation of three-input T cells (using CellTrace Far Red–stained T cells) was only observed under stimulation with K562 cells expressing all three antigens, *P < 0.01 compared with HER2$^+$ control. (**F**) In vivo validation of three-input T cells. These T cells selectively kill triple-antigen-positive tumors while sparing double-antigen-positive tumors in a bilateral tumor mouse model. 5e6 K562 tumor cells were implanted into the flanks of NSG mice (left flank: MET$^+$/HER2$^+$; right flank: MET$^+$/HER2$^+$/EGFR$^+$). Seven days after tumor implantation, 3e6 control untransduced T cells or MET-AND-EGFR in-series CAR T cells were administered by tail vein injection. Complete clearance of triple-antigen-positive tumor was seen, while tumor lacking one of the priming antigens (EGFR) was spared (n = 5 mice, error bars are SEM, significance determined by Holm-Sidak t-test, *P < 0.05). Individual mouse plots and 6e6 T cell dose plotted in fig. S10.

synNotch A induces expression of synNotch B, which in turn induces expression of CAR C), or in parallel (where synNotch A induces expression of part 1 and synNotch B induces expression of part 2 of a split CAR) (Fig. 3, A and B). We engineered both such configurations, using a first synNotch receptor that recognizes epidermal growth factor receptor (EGFR) and a second that recognizes MET. For the in-series circuit, the dual synNotch cascade drove expression of an anti-HER2 CAR (fig. S7A). For the in-parallel circuit, the same two synNotch receptors drove two different parts of a split anti-HER2 CAR [in this case, part 1 was a secreted CAR adapter that had an anti-HER2 scFv linked to a peptide neo-epitope (PNE) tag (*22*), and part 2 was a CAR that recognizes the PNE tag] (fig. S9A).

The in-series three-input cascade circuit showed precise recognition in multiple in vitro assays, including cell killing and proliferation (Fig. 3, C to E, and fig. S8C). These T cell outputs were only observed when presented with target cells expressing all three of the required antigens. In contrast, the in-parallel circuit was imperfect, showing partial killing of one of the two antigen combinations (EGFR$^+$/HER2$^+$) (fig. S9B). This is likely because of small amounts of leaky expression by the anti-MET synNotch, such that when the partner (anti-EGFR) synNotch is stimulated, some amount of the complete anti-HER2 split CAR is generated, resulting in killing activity. The linkage of these same synNotch receptors in series, which is predicted to lead to multiplicative specificity, provides an effective mechanism to suppress the impact of individual receptor leakiness (*23*, *24*), although with slowed killing kinetics. We also

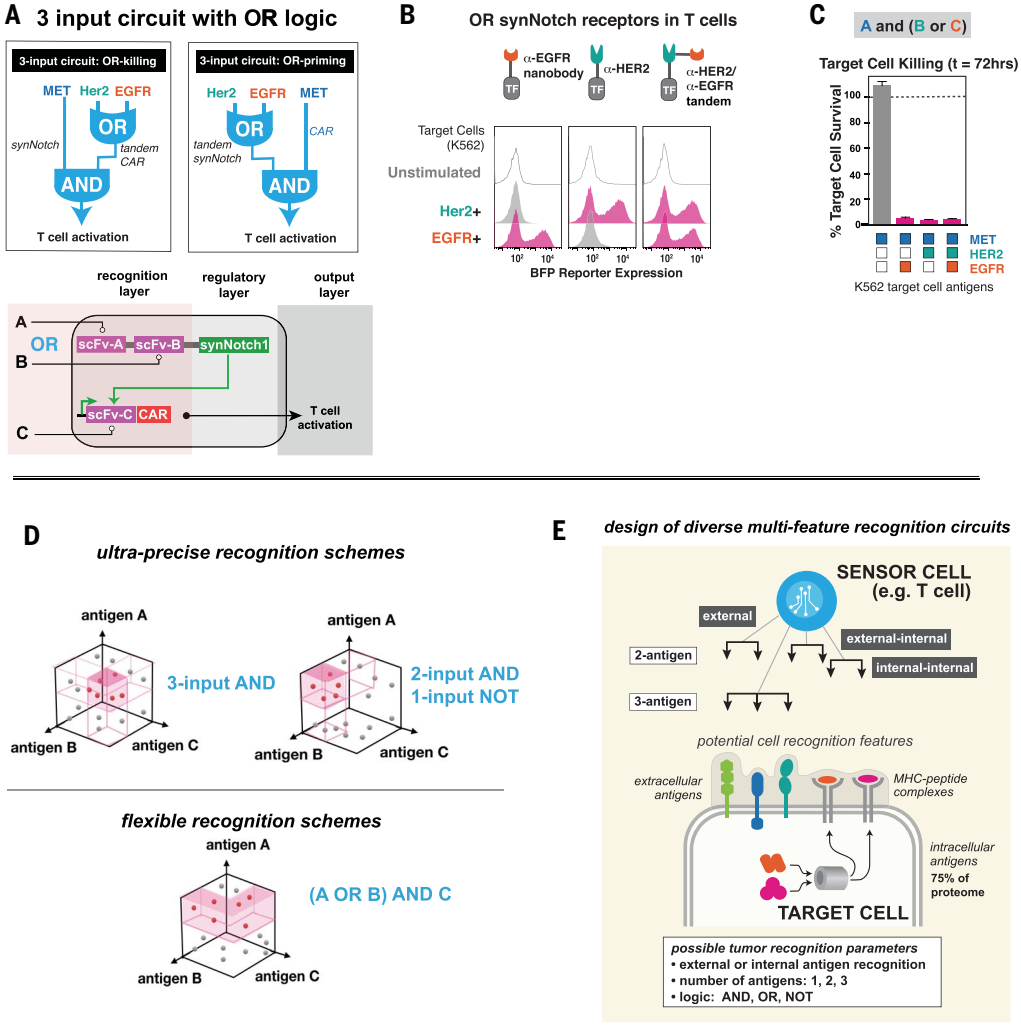


Fig. 4. Diverse cell-cell recognition schemes can be engineered using synNotch integrative circuits. (**A**) Three-input circuits that incorporate AND and OR logic to introduce increased recognition flexibility of alternative antigens (in this case, recognition of [HER2 OR EGFR] AND MET). Alternative designs involve use of either tandem CARs or tandem synNotch receptors (each with two antigen recognition domains). Detailed scheme for constructing an OR-AND three-input circuit using a tandem synNotch receptor (see fig. S11 for more detail). (**B**) Validating OR recognition function of a tandem synNotch receptor that recognizes both EGFR and HER2. Primary human CD8 T cells were engineered with an anti-EGFR synNotch, anti-HER2 synNotch, or anti-HER2/EGFR tandem synNotch controlling the expression of a BFP reporter. All synNotch receptors used a Gal4-VP64 transcription factor. These T cells were cocultured at 1:1 E:T for 24 hours with K562 target cells expressing HER2, EGFR, or neither antigen as labeled. Histograms of BFP levels measured via flow cytometry show that the dual synNotch specifically responds to either HER2 OR EGFR positive target cells, whereas the two single-antigen synNotch receptors only respond to one of the target cells (representative of three independent experiments). (**C**) Efficient OR killing output by CD8 T cells transfected with the above circuit. T cells efficiently kill target K562 cells that express [MET and Her2] or [MET and EGFR], but not MET alone (n = 3, error bars are SD). (**D**) We demonstrate that three-antigen circuits can be engineered with different user-desired recognition behaviors—one can engineer ultraprecise A-and-B-and-C or A-and-B-not-C circuits, or more-flexible recognition schemes that tolerate heterogeneity in particular dimensions of antigen space [A-or-B]-and-C. It may now be possible to program therapeutic cells capable of recognizing any possible combinatorial antigen signature desired. (**E**) T cells can be engineered with an assortment of diverse combinatorial antigen recognition circuits. Information from up to three receptors targeting different antigen inputs can be integrated to generate highly specific recognition programs that integrate intracellular and extracellular targets as well as AND, NOT, and OR logic. Such recognition circuits should also function in other engineered cell types. See fig. S12 for a summary "periodic table" of general multiantigen cell-cell recognition circuits that are possible.

identified other design principles for constructing efficient synNotch recognition cascades, focusing on the sequence of deploying the Gal4- and LexA-based receptors (Gal4 has stronger output than LexA) (fig. S7C). We built versions of the synNotch cascade with either order: Gal4→LexA (strong→weak) or LexA→Gal4 (weak→strong) (fig. S7D). We found that only the strong→weak cascade effectively propagated signal to yield detectable output in response to multiantigen stimulation (the importance of antigen density to these designs is discussed in the supplementary text section of the supplementary materials). Thus, these synthetic receptor cascades can be an effective way to integrate multiple inputs, subject to several key design constraints.

Given the precise in vitro function of this three-input AND gate in-series circuit, we tested whether T cells bearing this circuit could selectively kill tumors in vivo using mouse tumor models. We constructed target K562 tumor lines that were either MET$^+$/EGFR$^+$/HER2$^+$ (target) or MET$^+$/HER2$^+$ (control) (fig. S8A) and implanted them in a bilateral tumor model in NSG mice (Fig. 3F). Both of these tumors contain the CAR killing antigen (HER2) and thus could be killed if the circuit was substantially leaky. The three-input AND gate T cells were administered by tail vein injection and allowed to autonomously explore and act on both tumors. We found that the T cells showed the same precise recognition observed in vitro—the three-input T cells were able to rapidly clear the three-antigen tumors, while ignoring the two-antigen tumors on the opposing flank (Fig. 3F; see fig. S10, B and C, for repeat experiment at different dosage). These results show an extensible strategy for building higher-input AND gate recognition circuits.

A fourth recognition strategy that we explored involves building circuits with flexibility of recognition. Although increasing selectivity of recognition is generally desirable, intrinsic heterogeneity or loss of particular antigens could require increased flexibility of recognition

in particular dimensions. Recent studies have shown that tandem CARs—CARs with two different antigen recognition heads—provide a mechanism for engineering OR gates that can kill on the basis of either of two possible antigens (*25*–*29*). Such OR-killing functionality could be achieved by having a synNotch receptor induce expression of a tandem CAR. We tested whether we could also build tandem-headed OR gate synNotch receptors and incorporate them into higher-order circuits to accommodate priming antigen-specific recognition flexibility (Fig. 4A). A tandem synNotch receptor targeting EGFR and HER2 functions well to induce reporter expression in response to either individual target antigen (Fig. 4B). We incorporated this tandem synNotch into a three-antigen circuit (tandem synNotch→MET CAR), yielding T cells that robustly kill target cells that express MET combined with either EGFR or HER2 (Fig. 4C). Thus, it is possible to build OR gates into either synNotch or CAR receptors, and synNotch→CAR AND gates can flexibly recognize both priming and/or killing antigens. This gives control of how tightly or loosely a region of antigen-space is targeted to optimize effective recognition (Fig. 4D and fig. S13).

Together, these results show the power of cellular engineering as opposed to molecular engineering (*21*, *30*, *31*)—by engineering cellular regulation, we can integrate multiple molecular recognition events to achieve more precise and robust recognition and discrimination of target cells (e.g., cancer cells). Components such as synNotch do not carry out a prespecified physiological output function but rather execute the generic function of inducing transcriptional expression. Thus, they function as connectors that allow different molecular recognition events and outputs to be linked in different sequential configurations. By combining molecular recognition events with intracellular regulatory circuits, a diverse array of selective recognition programs can be engineered. Here we show that a sensor cell (for example, a therapeutic T cell) can be flexibly engineered to recognize combinations of intracellular and extracellular antigens and to integrate information from one, two, or three antigens using AND/OR/NOT logic (Fig. 4E). These modular and hierarchically designed cell circuits are compatible with new target antigens and new kinds of molecular recognition platforms that may emerge.

REFERENCES AND NOTES

1. E. O. Long, H. S. Kim, D. Liu, M. E. Peterson, S. Rajagopalan, *Annu. Rev. Immunol.* **31**, 227–258 (2013).
2. C. A. Klebanoff, S. A. Rosenberg, N. P. Restifo, *Nat. Med.* **22**, 26–36 (2016).
3. R. A. Morgan *et al.*, *Mol. Ther.* **18**, 843–851 (2010).
4. R. A. Morgan *et al.*, *J. Immunother.* **36**, 133–151 (2013).
5. F. C. Thistlethwaite *et al.*, *Cancer Immunol. Immunother.* **66**, 1425–1436 (2017).
6. L. A. Johnson *et al.*, *Blood* **114**, 535–546 (2009).
7. M. R. Parkhurst *et al.*, *Mol. Ther.* **19**, 620–626 (2011).
8. C. H. Lamers *et al.*, *Mol. Ther.* **21**, 904–912 (2013).
9. F. Perna *et al.*, *Cancer Cell* **32**, 506–519.e5 (2017).
10. M. MacKay *et al.*, *Nat. Biotechnol.* **38**, 233–244 (2020).
11. R. Dannenfelser *et al.*, *Cell Syst.* **11**, 215–228.e5, E5 (2020).
12. L. Morsut *et al.*, *Cell* **164**, 780–791 (2016).
13. K. T. Roybal *et al.*, *Cell* **164**, 770–779 (2016).
14. M. S. Almén, K. J. V. Nordström, R. Fredriksson, H. B. Schiöth, *BMC Biol.* **7**, 50 (2009).
15. J. N. Søndergaard *et al.*, *J. Transl. Med.* **8**, 39 (2010).
16. J. Tsoi *et al.*, *Cancer Cell* **33**, 890–904.e5 (2018).
17. Y.-J. Lee *et al.*, *Ann. Dermatol.* **23**, 33–38 (2011).
18. P. G. Natali *et al.*, *Br. J. Cancer* **68**, 746–750 (1993).
19. S. Nambiar *et al.*, *Arch. Dermatol.* **141**, 165–173 (2005).
20. H. Liu *et al.*, *Clin. Cancer Res.* **23**, 478–488 (2017).
21. V. D. Fedorov, M. Themeli, M. Sadelain, *Sci. Transl. Med.* **5**, 215ra172 (2013).
22. D. T. Rodgers *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E459–E468 (2016).
23. S. Hooshangi, S. Thiberge, R. Weiss, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3581–3586 (2005).
24. Y. Li *et al.*, *Nat. Chem. Biol.* **11**, 207–213 (2015).
25. D. Schneider *et al.*, *J. Immunother. Cancer* **5**, 42 (2017).
26. E. Zah, M.-Y. Lin, A. Silva-Benedict, M. C. Jensen, Y. Y. Chen, *Cancer Immunol. Res.* **4**, 498–508 (2016).
27. T. J. Fry *et al.*, *Nat. Med.* **24**, 20–28 (2018).
28. N. N. Shah *et al.*, *Blood* **132** (suppl. 1), 4193 (2018).
29. N. Hossain *et al.*, *Blood* **132** (suppl. 1), 490 (2018).
30. M. J. Lajoie *et al.*, *Science* **369**, 1637–1643 (2020).
31. Z. Chen *et al.*, *Science* **368**, 78–84 (2020).

ACKNOWLEDGMENTS

We thank J. Bluestone, M. Broeker, K. Chang, J. Garbarino, C. Ghosh, X. Huang, S. Stark, P. Lopez Pazmino, D. Phineas, N. Blizzard, H. Okada, F. Szoka, and the Lim and Troyanskaya labs for technical assistance, advice, and helpful discussion.
Funding: This work was supported by a Genentech Pre-Doctoral Fellowship (to J.Z.W.), a Jane Coffin Childs Memorial Fund Post-Doctoral Fellowship (to G.M.A.), NIH grant R35 CA197633 (to A.R.) NIH grants P50GM081879, R01 CA196277 (to W.A.L.), and the Howard Hughes Medical Institute (to W.A.L.). **Author contributions:** Conceptualization: W.A.L., K.T.R., G.M.A., and J.Z.W. Funding acquisition: W.A.L. Investigation: J.Z.W., G.M.A., D.S., I.S.S., K.H.K., V.P.G., G.E.S., and W.Y. Methodology: J.Z.W. and G.M.A. Melanoma bioinformatics: C.P.-S., J.T., and A.R. Project administration: W.A.L. Supervision: W.A.L., J.Z.W., and G.M.A. Visualization: W.A.L., J.Z.W., and G.M.A. Writing: W.A.L., J.Z.W., G.M.A., and D.S. **Competing interests:** W.A.L., K.T.R., J.Z.W., G.M.A., and D.S. are inventors on patents related to this work. W.A.L., K.T.R., and J.Z.W. receive licensing fees for patents that were licensed by Cell Design Labs and are now part of Gilead. W.A.L. is on the scientific advisory board of Allogene Therapeutics. K.T.R. is a cofounder of Arsenal Biosciences. J.Z.W. is currently an employee of Arsenal Biosciences. A.R. has received honoraria from consulting with Amgen, Bristol-Myers Squibb, Chugai, Genentech, Merck, Novartis, Roche, and Sanofi; is or has been a member of the scientific advisory board and holds stock in Advaxis, Apricity, Arcus Biosciences, Bioncotech Therapeutics, Compugen, CytomX, Five Prime, FLX-Bio, ImaginAb, Isoplexis, Kite-Gilead, Lutris Pharma, Merus, PACT Pharma, Rgenix, and Tango Therapeutics; and has received research funding from Agilent and from Bristol-Myers Squibb through Stand Up to Cancer (SU2C). **Data and materials availability:** All data are available in the main text or the supplementary materials. Expression plasmids are available from Addgene under a material transfer agreement. Cell lines, other than primary T cells, are available upon request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1099/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (*32*–*36*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

5 May 2020; accepted 12 October 2020
10.1126/science.abc6270

RIBOSOME STRUCTURE

Elongational stalling activates mitoribosome-associated quality control

Nirupa Desai[1]*, Hanting Yang[1]*, Viswanathan Chandrasekaran[1]*, Razina Kazi[1], Michal Minczuk[2], V. Ramakrishnan[1]†

The human mitochondrial ribosome (mitoribosome) and associated proteins regulate the synthesis of 13 essential subunits of the oxidative phosphorylation complexes. We report the discovery of a mitoribosome-associated quality control pathway that responds to interruptions during elongation, and we present structures at 3.1- to 3.3-angstrom resolution of mitoribosomal large subunits trapped during ribosome rescue. Release factor homolog C12orf65 (mtRF-R) and RNA binding protein C6orf203 (MTRES1) eject the nascent chain and peptidyl transfer RNA (tRNA), respectively, from stalled ribosomes. Recruitment of mitoribosome biogenesis factors to these quality control intermediates suggests additional roles for these factors during mitoribosome rescue. We also report related cryo–electron microscopy structures (3.7 to 4.4 angstrom resolution) of elongating mitoribosomes bound to tRNAs, nascent polypeptides, the guanosine triphosphatase elongation factors mtEF-Tu and mtEF-G1, and the Oxa1L translocase.

Mitochondria are double-membrane organelles with important roles in metabolism and adenosine 5′-triphosphate generation. They have retained many features of their α-proteobacterial ancestry, including their own genome and ribosomes (mitoribosomes). We and others have determined structures of mitoribosomes from various eukaryotes (*1*–*4*). Human mitoribosomes comprise a 39*S* large subunit (LSU) [52 proteins, a 16*S* ribosomal RNA (rRNA) and mitochondrial $tRNA^{Val}$ (mt-$tRNA^{Val}$)] and a 28*S* small subunit (SSU) (30 proteins and a 12*S* rRNA) and synthesize 13 inner mitochondrial membrane proteins encoded by mitochondrial DNA. These proteins are cotranslationally inserted into the mitochondrial inner membrane by the Oxa1L translocase, where they are assembled with nuclear-encoded components into oxidative phosphorylation complexes.

Little is known about mechanisms of cotranslational ribosome-associated quality control (RQC) in mitochondria and, if they exist, whether they operate any differently from those present in bacteria and the eukaryotic cytosol (*5*). Stalling can occur at any step of the translation cycle due to defective mRNAs, including absent in-frame stop codons, truncations or inappropriate polyadenylation within the protein-coding region (*6*), and tRNA mutations or insufficiency (*7*). In these and other cases, RQC is critical to maintain cellular homeostasis by rescuing stalled ribosomes, preventing protein misfolding and other aberrant translation phenotypes (*8*, *9*). Mitochondria, however, lack counterparts to bacterial rescue mechanisms such as trans-translation, stringent control, and ArfA (*10*). ICT1, the presumed mitochondrial ortholog of another bacterial rescue factor, ArfB, was instead identified as an integral component of the LSU (*1*, *3*, *11*). In silico studies have proposed that other putative mitochondrial release factors may exist in addition to the canonical termination factor mtRF1a, but without biochemical evidence (*10*).

We reasoned that any attempt to induce translational stalling might trap mitoribosomes at various stages of the translation cycle in the act of nascent chain insertion into the inner mitochondrial membrane and generate intermediates suitable for structural analyses by cryo–electron microscopy (cryo-EM). We therefore purified mitoribosomes from a genetically engineered human cell line that lacks 2′-5′ phosphodiesterase 12 (PDE12). PDE12 facilitates the maturation of mt-tRNA, mt-$tRNA^{Lys}$ in particular, by deadenylating spurious poly(A) tails from mt-tRNAs and restoring the 5′-CCA-3′. A PDE12 knockout can therefore lead to aberrant polyadenylation of the 3′ ends of mt-tRNAs and, consequently, ribosome stalling (*12*). During purification, we included GMPPCP, a nonhydrolyzable GTP analog, to prevent dissociation of guanosine triphosphatases (GTPases) mtEF-G1 and mtEF-Tu from mitoribosomes, and *n*-dodecyl β-D-maltoside, to solubilize mitoribosomes anchored to the mitochondrial inner membrane via the Oxa1L translocase (*13*) (fig. S1). Exhaustive classification of over 3 million mitoribosomes imaged using cryo-EM yielded several structures at resolutions of 3.1 to 4.4 Å (table S1 and figs. S2 to S4).

As expected, we observed a substantial proportion of stalled mitoribosomes. We also observed two classes of LSUs with attached tRNA(s) and nascent polypeptide (Fig. 1). The presence of a nascent chain attached to a tRNA confirms that these classes represent LSUs of ribosomes that have been split before termination and nascent chain hydrolysis have occurred. Therefore, they represent intermediates in a hitherto-unreported mitoribosome-associated quality control (mtRQC) pathway.

The first LSU class was resolved to 3.1-Å resolution and harbors both P- and E-site tRNAs (Fig. 1A). The tRNAs are secured by bridging interactions with elements of vicinal proteins and other unassigned densities that span the inter-tRNA space (fig. S5A). The E-site tRNA is further stabilized by D-loop interactions with uL1m, similar to those seen in bacterial and mammalian structures (Fig. 1A, inset) (*14*). We conclude that this class represents an early mitoribosome-associated quality control intermediate that has been split by an unknown mechanism upon stalling.

The second LSU class, determined to 3.3 Å resolution, represents a rescue intermediate prior to hydrolysis of the nascent chain. The resolution of this complex permitted unambiguous identification and modeling of two proteins. The complex contains a peptidyl-tRNA that has recruited a heterodimer of the release factor homolog C12orf65 (here renamed mtRF-R, for mitochondrial release factor in rescue) and MTRES1, a double-stranded RNA (dsRNA) binding protein (Fig. 1, B and C, and fig. S5B). There was excellent agreement between the protein sequences and most side-chain densities (Fig. 1C). mtRF-R occupies the A-site of the LSU next to the peptidyl tRNA, whereas MTRES1 wraps around the anticodon stem-loop of the peptidyl tRNA (Fig. 2A and fig. S6A). The two proteins interact via the C-terminal helix of mtRF-R and the S4-like domain of MTRES1. The observed binding of mtRF-R and MTRES1 is consistent with their coimmunoprecipitation from human embryonic kidney 293T cells (*15*).

mtRF-R structurally resembles the GGQ-containing catalytic domain of the bacterial release factors RF1 and RF2. Like RF1 and -2, mtRF-R inserts a conserved GGQ motif into the peptidyl transferase center (PTC) (Fig. 1D). One major difference is that the C-terminal helix of mtRF-R splits into two α helices (α_1 and α_2), whereas in bacterial release factors, the equivalent helix (α_7) is continuous (Fig. 2, A and B). The long α_2 helix of mtRF-R, which extends to contact MTRES1, is stabilized by the tip of helix 69 of 16*S* rRNA (fig. S6, A and B). In addition to the α_2 helical contacts, conserved Arg^{99} of the GGQ domain hydrogen-bonds with C-71 of the terminal 3′-CCA, and Gln^{103} contacts position 1 of the tRNA (fig. S6C).

The modeled residues of MTRES1 (129 to 216) bind the anticodon stem-loop of the P-site tRNA as well as helix 69 rRNA and the C terminus of the α_2 helix of mtRF-R (figs. S6 and S7). MTRES1 was predicted to contain an S4-like dsRNA binding domain (*15*). This region

[1]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [2]MRC Mitochondrial Biology Unit, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: ramak@mrc-lmb.cam.ac.uk

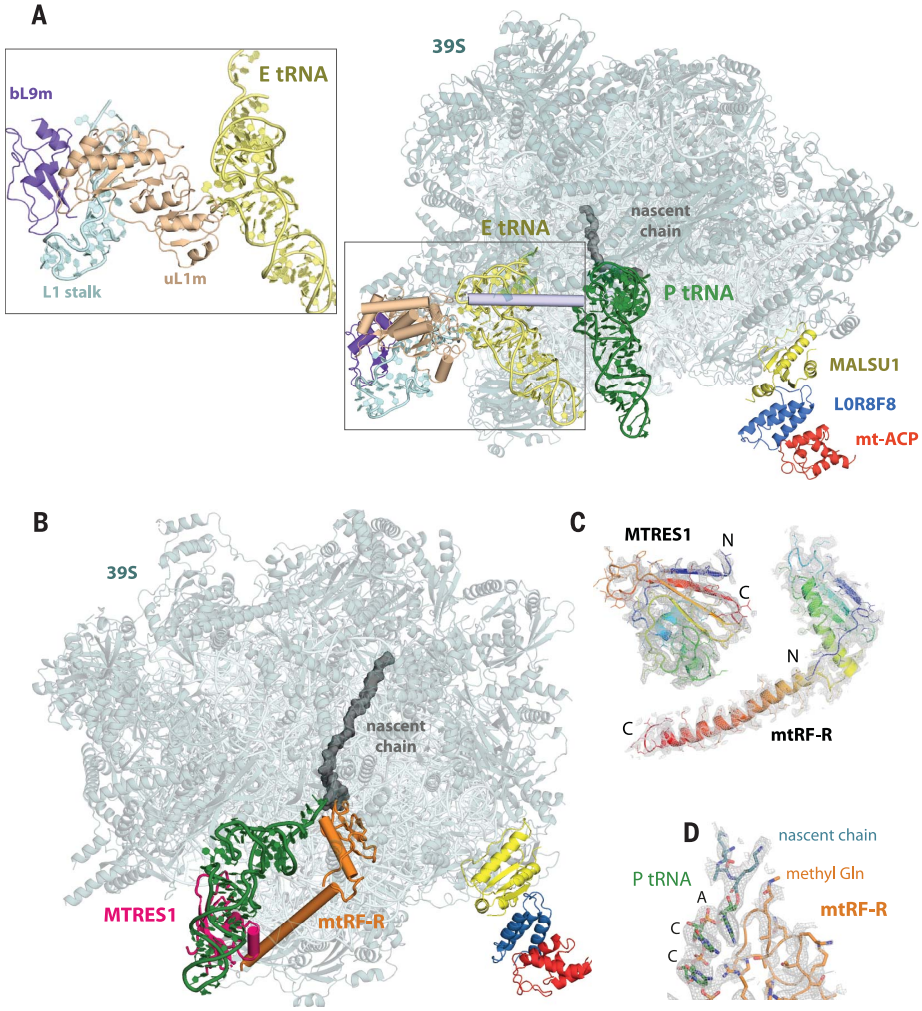


Fig. 1. Structures of split mitoribosome intermediates undergoing quality control. (**A**) Overall structure of mitochondrial LSU complexed with P, E tRNAs, nascent chain, and elements of proteins of the large subunit. (Inset) Packing interactions between the D-loop of the E-site tRNA and the L1 stalk. (**B**) Overall structure of mitochondrial LSU in complex with a P tRNA, nascent chain, mtRF-R, and MTRES1. (**C**) Map–model fits for MTRES1 (left) and mtRF-R (right). The N (blue) and C (red) terminals are indicated. The good local resolution permitted unambiguous assignment of the backbone and most side chains (shown as lines). (**D**) Close-up view of the ribosomal PTC, showing cryo-EM density for interactions between the CCA end of the P tRNA and the GGQ domain of mtRF-R poised to hydrolyze the nascent chain.

binds helix 69 and resembles the corresponding domain of the *Escherichia coli* ribosomal protein uS4, despite having only 15% sequence identity (Fig. 2, C and D). Other regions outside the S4-like domain have diverged considerably to confer tRNA and mtRF-R binding ability. The conserved α_1 and α_2 helices of the S4-like domain make specific contacts with helix 69 of the 16S rRNA (fig. S6D). MTRES1 also interacts with a double-stranded region of the peptidyl tRNA (fig. S6E), where stabilizing contacts are made between positive residues of β_1 and β_4 against the backbone of positions 24, 36, and 37 of the tRNA. MTRES1 and mtRF-R interact with each other predominantly using the C termini of their respective α_2 helices, via a salt bridge and a network of hydrogen bonds and hydrophobic interactions (fig. S6F).

Comparison of the two postsplitting classes reveals the order of events along the mtRQC pathway. We conclude that the first class is an early postsplitting state of the LSU that serves as the substrate for mtRF-R to bind the empty aminoacyl site in the PTC. Concomitant binding and repositioning of the peptidyl tRNA by MTRES1 (Fig. 2, A and B, compare peptidyl tRNA positions) would eject the E-site tRNA and allow the mtRF-R/MTRES1 heterodimer to hydrolyze the nascent chain and extract the tRNA. RQC pathways must target the small population of stalled ribosomes while sparing the majority of active ribosomes from futile rescue. This pathway is specific for stalled and split mitoribosomal LSUs because (i) actively translating ribosomes rarely persist with an empty A site, (ii) the kinked α_1-α_2 arrangement of mtRF-R would sterically clash with an SSU if present, and (iii) mtRF-R is stabilized by packing against MTRES1, an ordered helix 69 of the 16*S* rRNA, and the peptidyl tRNA, all of which are absent in empty LSUs.

In this LSU class undergoing rescue (Fig. 1B), nascent chain hydrolysis has not occurred, owing to unfavorable geometry between the attacking and leaving groups (Fig. 1D) (*16*). Although our map likely contains averaged density from many mitochondrial nascent peptides, it is also consistent with a lysyl residue in the PTC in a flipped position that we and others have observed in mammalian cytosolic ribosomes stalled on poly(A) and with a polylysine nascent chain (*16*). The ε nitrogen of the glutamine is methylated, as seen in other release factors, but is ~4.8 Å away from the carbonyl carbon of the lysyl residue, which is too far for a nucleophilic attack on the ester bond and could explain why the nascent chain has not yet been hydrolyzed. Our structure raises the question of whether mtRF-R can hydrolyze a suitable substrate in vitro. Among the putative mitochondrial release factors, only mtRF1a, the canonical mitochondrial termination factor, and ICT1, a constitutive 39*S* protein, have proven peptidyl-tRNA hydrolase activities, but previous in vitro experiments that failed to detect mtRF-R activity were performed without MTRES1 and used intact bacterial 70*S* ribosomes as substrate (*11*, *17*, *18*). We therefore reasoned, based on our structure, that mtRF-R too may hydrolyze bacterial ribosome-nascent chains (RNCs) in vitro, provided that (i) 70*S*-RNCs are first split into 50*S*-RNCs and (ii) MTRES1 is also present. This is indeed the case (fig. S8), showing that mtRF-R acts as a release factor on stalled LSU tRNA-nascent chains.

The importance of this rescue pathway is highlighted by the facts that mtRF-R is essential for mitochondrial protein synthesis

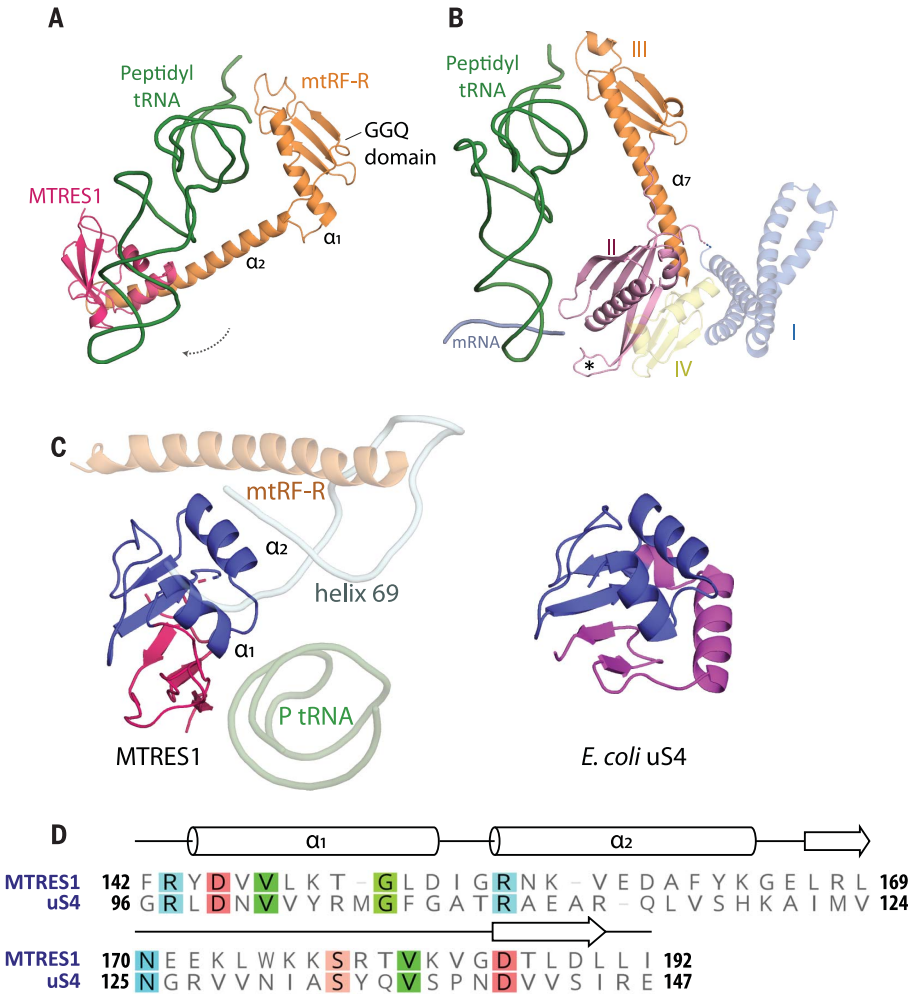


Fig. 2. Comparison of mtRF-R and MTRES1 with bacterial homologs. (**A**) mtRF-R (orange) and MTRES1 (dark pink) together resemble domains II and III in bacterial RF2. The peptidyl tRNA has been repositioned toward the E site (dotted arrow), and the long α_2 helix of the GGQ domain binds MTRES1, which wraps around the peptidyl tRNA. (**B**) Bacterial class I release factor (domain I, blue; domain II, pink; domain III, orange; domain IV, pale yellow) in the A site of a terminating 70*S* ribosome. The peptidyl tRNA (green) and mRNA (pale blue) are also shown. The asterisk indicates the PXT motif that recognizes the stop codon (PDB 5MDV). (**C**) Comparison of *E. coli* uS4 and human MTRES1. The structurally conserved dsRNA binding region is shown in blue. Note that the rest of the proteins are topologically different (magenta versus dark pink). (**D**) Structure-based alignment and secondary structure of the two proteins. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; W, Trp; and Y, Tyr.

(*19*, *20*) and that mutations lead to human disease phenotypes, especially the triad of optic atrophy, peripheral neuropathy, and spastic paraparesis (*10*). Autosomal recessive mutations result in mtRF-R truncations, the extent of which largely, but not exclusively, correlates with severity of disease (*17*, *19*, *20*) (fig. S9). At the same time, MTRES1-knockout human cells exhibit reduced mitochondrial translation and oxidative phosphorylation deficiency (*15*).

The mtRF-R/MTRES1 mtRQC pathway is reminiscent of, but differs mechanistically from, the eukaryotic cytosolic system of the yeast tRNA endonuclease Vms1 (ANKZF1 in humans) and the ABCF-type adenosine triphosphatase Arb1 (*21*, *22*). Both systems address the cellular need to release nascent chains from stalled ribosomal large subunits but operate via nascent chain hydrolysis and tRNA nucleolytic cleavage, respectively. Analogously to bacterial and mammalian RQC, the mt-mRNA and nascent chain must be extracted after hydrolysis and targeted for degradation.

Previous work identified a module of three factors, MALSU1 (mitochondrial assembly of ribosomal large subunit 1; C7orf30), L0R8F8, and mt-ACP (mitochondrial acyl carrier protein), that functions during mitoribosome biogenesis to bind immature LSUs and prevent premature subunit association (*23*). Intriguingly, we find the same MALSU1/L0R8F8/mt-ACP module on both the early postsplitting and later rescue intermediates (Fig. 1, A and B), suggesting that the module is recruited to block subunit joining during tRNA and nascent chain removal from the split LSU. Other regulatory factors, such as eukaryotic translation initiation factor 6 (eIF6), ABCE1, and eIF3, prevent premature subunit joining during ribosome biogenesis, initiation, and ribosome recycling after termination, as well as after ribosome rescue (*23*–*27*). We propose that the MALSU1/L0R8F8/mt-ACP module functions during both biogenesis and rescue by preventing subunits of mitoribosomes from associating prematurely.

In addition to the two rescue pathway intermediates described above, our dataset also captured all of the main states during normal translational elongation by mitoribosomes, except for one (Fig. 3 and 4). One intermediate (Fig. 3A and fig. S10A) stalls with an empty A site and is likely rescued by mtRQC. Another observed class captures a state during decoding (Fig. 3B and fig. S10B) and contains a pre-accommodated mt-tRNA delivered to the A site of the mitoribosome by mtEF-Tu•GMPPCP. The mitochondrial ternary complex (figs. S10C and S11, A to C) closely resembles the equivalent GMPPCP-bound complex on the bacterial ribosome. The decoding nucleotides A1561 and A1562 of helix 44 in the 12*S* rRNA (*E. coli* A1492 and A1493) both flip out to read the minor groove of the double-stranded helix formed between the +1, +2, and +3 A-site mRNA nucleotides and the anticodon of the A/T tRNA (figs. S10B and S11D). As in other ribosomes, correct decoding induces domain closure, a large-scale movement of the body of the SSU that activates GTP hydrolysis and tRNA accommodation (*28*).

Domain closure is also observed in mitoribosomes with an accommodated cognate tRNA in the A site (A/A, P/P, E/E) (Fig. 3C) but not in ribosomes with an empty A site (Fig. 3A). The poor local resolution in the former class precluded direct visualization of the ester bond between either tRNA and the

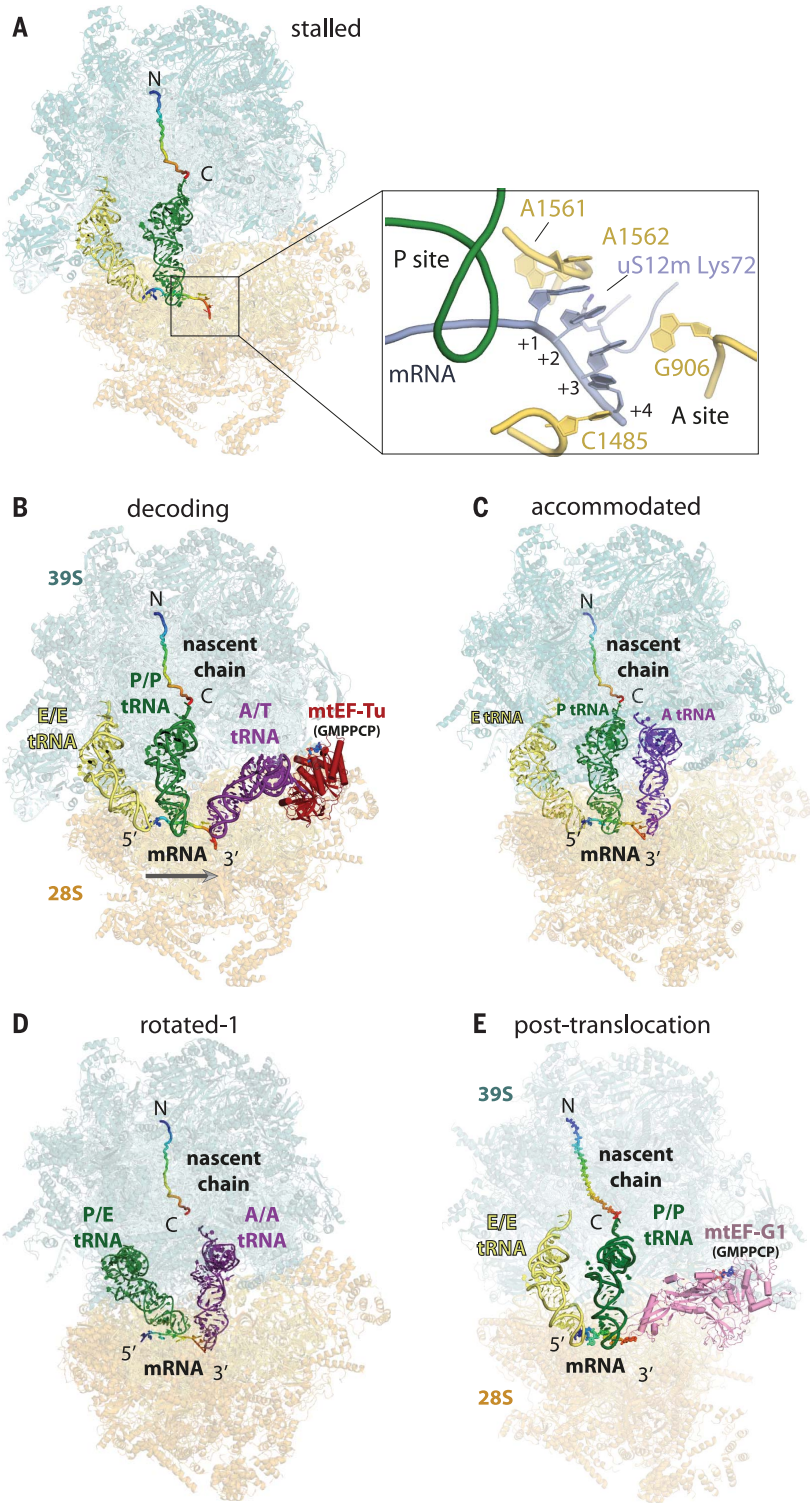


Fig. 3. Structures of active and stalled elongation intermediates. (**A**) Structure of an elongating mitoribosome stalled with an empty A site and bound to P- (green) and E-site (yellow) tRNAs. (Inset) Close-up view of interactions between the mRNA (pale blue) and decoding bases of the 12*S* rRNA (yellow). (**B**) Elongating mitoribosome-nascent chain complex during decoding of an incoming A/T tRNA (purple) bound to mitochondrial elongation factor Tu (mtEF-Tu). The arrow indicates the direction of translation. (**C**) Canonical state bound to A, P, and E tRNAs. (**D**) Rotated-1 state with A/A and P/E tRNAs. The nascent chain is attached to the A/A tRNA, but the ester bond is not seen due to poor resolution. These coordinates were not deposited. (**E**) Posttranslocated state bound to mitochondrial elongation factor G1 (mtEF-G1).

nascent chain, but the absence of a P/E hybrid tRNA implies that peptidyl transfer has not yet occurred. Particles in which peptidyl transfer has already occurred are also present in the dataset, resulting in mitoribosomes in the rotated-1 state with A/A and P/E tRNAs (Fig. 3D). mRNA, tRNA, rRNA, and protein interactions observed in the mtEF-Tu–bound map (Fig. 3B and figs. S10B and S11D) are preserved in the decoding center of the mitoribosome bound to A, P, and E tRNAs (Fig. 3C).

By contrast, decoding centers of stalled mitoribosomes were markedly rearranged (Fig. 3A, inset, and fig. S10A). In the absence of PDE12, these ribosomes have likely stalled with AAG or AAA codons in the A site (*12*). Consecutive AAA codons such as those in poly(A) tails have a particular propensity to base-stack in the decoding center even without stabilizing Watson-Crick base-pairing interactions with an A-site anticodon (*16*). Thus, our map would also be consistent with stalling on poly(A) stretches of mitochondrial mRNAs. As seen previously in cytosolic ribosomes, base-stacked poly(A) nucleotides induce A1562 but not A1561 to flip out and participate in the stack. The stack is capped at the 3′ end by C1485.

After peptidyl transfer, the GTPase mtEF-G1 translocates the A/P and P/E hybrid tRNAs (rotated-2) into the P/P and E/E states, respectively, accompanied by a corresponding movement of the mRNA. We determined a 3.7-Å resolution structure of an mtEF-G1•GMPPCP–bound mitoribosome in the posttranslocation state with P/P and E/E tRNAs and a nascent polypeptide chain (Fig. 3E).

The bacterial elongation factor EF-G•GTP is bifunctional and not only translocates mRNA-tRNA through the ribosome but also recycles the ribosome after termination, in conjunction with ribosome release factor. In human mitochondria, these roles are fulfilled by two EF-G paralogs, mtEF-G1 and mtEF-G2, respectively (*29*). mtEF-G1 is a homolog of bacterial EF-G with an identical five-domain architecture (fig. S10D and S11, E to H) (*30*). The GTP binding domain I(G) is in a prehydrolysis conformation, as previously described for bacteria (*30*). Switch loops I and II are ordered around the bound GMPPCP (fig. S11G, inset).

Electron cryotomographic studies have observed human mitoribosomes at the mitochondrial inner membrane, likely in complex with the translocase Oxa1L (*13*). A direct connection between mitoribosomes and Oxa1L would facilitate cotranslational insertion of mitochondrial proteins. To investigate nascent chain translocation, we performed focused classification with signal subtraction (FCwSS) on the ribosomal exit tunnel by using particles from the whole dataset (fig. S3). Oxa1L was present in our samples as confirmed by Western immunoblotting and mass spectrometry (fig. S1, B and C) and is also seen in both LSU and

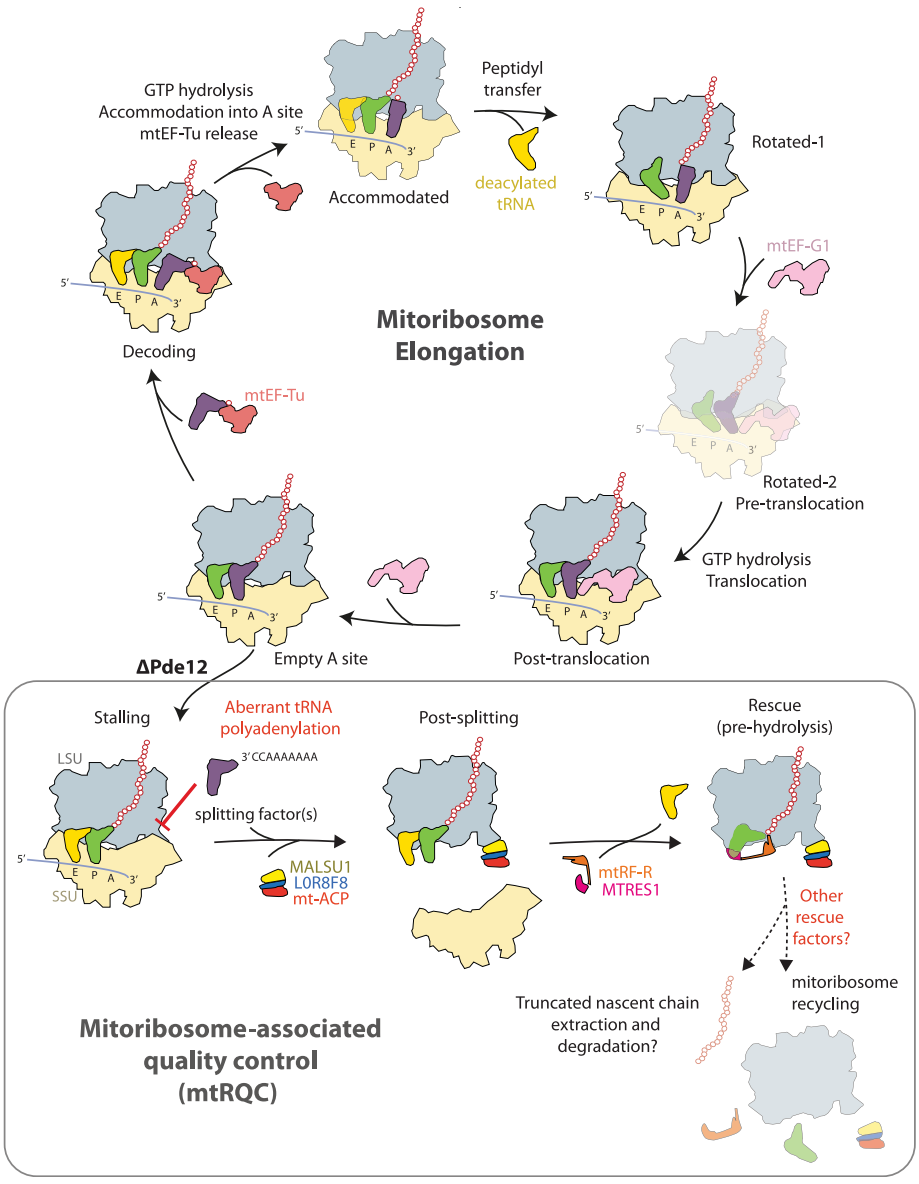


Fig. 4. Schematic of the mitoribosomal elongation cycle (top) and mtRQC (bottom). States visualized in this study are shown as opaque surfaces, whereas the states not visualized in this study are translucent. PDB codes of bacterial counterparts to these states are 4V5F (posttranslocation), 6WD0 (empty A site), 4V5L (decoding), and 6WDD (accommodated).

monosome maps (figs. S3 and S12A), providing evidence that normal translational elongation as well as rescue and recycling occur during nascent chain insertion into the mitochondrial inner membrane. The best map of Oxa1L was obtained from one additional round of FCwSS around the exit tunnel on the 39*S* particles and refined to 3.1 Å (fig. S3 and S12). Most of Oxa1L was at low resolution, presumably due to inherent flexibility, and we could only model the C-terminal tail as alanines. Oxa1L contacts two mitoribosomal proteins, uL24m and mL45. The C-terminal tail (30 residues) wraps over the surface of uL24m (fig. S12B). On the other side of the exit tunnel, it contacts an α helix of mL45. The nascent chain can be traced from the 3′-CCA of the P-site tRNA throughout the polypeptide tunnel and as it emerges from the tunnel exit, but it becomes disordered before contacting Oxa1L.

A paradigm shift occurred in the field of translation when it was discovered that cells use the translation machinery itself to detect errors and activate quality control mechanisms. Although the existence of quality control in mitochondria was predicted on the basis of analogy with bacteria and eukaryotic cytosol, very little was known about the molecular mechanisms. Our discovery of a mitoribosome-associated quality control pathway (Fig. 4, bottom) and structural characterization of elongating mitoribosomes (Fig. 4, top) now set the stage to understand the regulation of mitochondrial translation. We also demonstrate that exhaustive in silico classification of large cryo-EM datasets can serve as a viable alternative strategy to uncover processes involved in translation in the absence of suitable in vitro tools.

REFERENCES AND NOTES

1. A. Amunts, A. Brown, J. Toots, S. H. W. Scheres, V. Ramakrishnan, *Science* **348**, 95–98 (2015).
2. N. Desai, A. Brown, A. Amunts, V. Ramakrishnan, *Science* **355**, 528–531 (2017).
3. B. J. Greber *et al.*, *Science* **348**, 303–308 (2015).
4. D. J. F. Ramrath *et al.*, *Science* **362**, eaau7735 (2018).
5. O. Brandman, R. S. Hegde, *Nat. Struct. Mol. Biol.* **23**, 7–15 (2016).
6. P. A. Frischmeyer *et al.*, *Science* **295**, 2258–2261 (2002).
7. R. Ishimura, G. Nagy, I. Dotu, J. H. Chuang, S. L. Ackerman, *eLife* **5**, e14295 (2016).
8. W. E. Balch, R. I. Morimoto, A. Dillin, J. W. Kelly, *Science* **319**, 916–919 (2008).
9. F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **75**, 333–366 (2006).
10. S. A. Ayyub, F. Gao, R. N. Lightowlers, Z. M. Chrzanowska-Lightowlers, *J. Cell Sci.* **133**, jcs231811 (2020).
11. R. Richter *et al.*, *EMBO J.* **29**, 1116–1125 (2010).
12. S. F. Pearce *et al.*, *eLife* **6**, e27596 (2017).
13. R. Englmeier, S. Pfeffer, F. Förster, *Structure* **25**, 1574–1581.e2 (2017).
14. S. Mohan, H. F. Noller, *Nat. Commun.* **8**, 14285 (2017).
15. S. Gopalakrishna *et al.*, *Nucleic Acids Res.* **47**, 9386–9399 (2019).
16. V. Chandrasekaran *et al.*, *Nat. Struct. Mol. Biol.* **26**, 1132–1140 (2019).
17. H. Antonicka *et al.*, *Am. J. Hum. Genet.* **87**, 115–122 (2010).
18. H. R. Soleimanpour-Lichaei *et al.*, *Mol. Cell* **27**, 745–757 (2007).
19. H. Shimazaki *et al.*, *J. Med. Genet.* **49**, 777–784 (2012).
20. M. Wesolowska *et al.*, *J. Neuromuscul. Dis.* **2**, 409–419 (2015).
21. T. Su *et al.*, *Nature* **570**, 538–542 (2019).
22. M. C. J. Yip *et al.*, *Nat. Struct. Mol. Biol.* **26**, 343–349 (2019).
23. A. Brown *et al.*, *Nat. Struct. Mol. Biol.* **24**, 866–869 (2017).
24. M. Gartmann *et al.*, *J. Biol. Chem.* **285**, 14848–14851 (2010).
25. A. Heuer *et al.*, *Nat. Struct. Mol. Biol.* **24**, 453–460 (2017).
26. V. G. Kolupaeva, A. Unbehaun, I. B. Lomakin, C. U. T. Hellen, T. V. Pestova, *RNA* **11**, 470–486 (2005).
27. E. Mancera-Martínez, J. Brito Querido, L. S. Valasek, A. Simonetti, Y. Hashem, *RNA Biol.* **14**, 1279–1285 (2017).
28. R. M. Voorhees, T. M. Schmeing, A. C. Kelley, V. Ramakrishnan, *Science* **330**, 835–838 (2010).
29. N. Mai, Z. M. A. Chrzanowska-Lightowlers, R. N. Lightowlers, *Cell Tissue Res.* **367**, 5–20 (2017).
30. Y.-G. Gao *et al.*, *Science* **326**, 694–699 (2009).

ACKNOWLEDGMENTS

We thank J. Grimmett and T. Darling for advice, data storage, and high-performance computing; S. Chen, J. Brown, G. Cannone, and G. Sharov for technical support; P. Emsley, G. Murshudov, and P. Afonine for help with model building and refinement; S.-Y. Peak-Chew and M. Skehel for mass spectrometry analysis; T. Nakane for help with RELION; the Ramakrishnan lab members J. L. Llácer, A. Brown, and R. S. Hegde for useful discussions and reagents. We acknowledge the MRC Laboratory of Molecular Biology Electron Microscopy Facility for access and support of electron microscopy, sample preparation, and data collection. **Funding:** This work was supported by the UK Medical Research Council (MC_U105184332 to V.R. and MC_UU_00015/4 to M.M.), a Wellcome Trust Senior Investigator award (WT096570), the Agouron Institute, and the Louis-Jeantet Foundation (V.R.). N.D. is funded by a Wellcome Trust Clinical PhD Fellowship (110301/Z/15/Z). H.Y. is funded by an EMBO Postdoctoral Fellowship (EMBO ALTF 806-2018). **Author contributions:** N.D. and H.Y. performed cell biology, biochemistry, sample preparation, and cryo-EM data collection; R.K. purified recombinant proteins; V.C. performed the hydrolysis experiments. M.M. provided key

research material; V.C. and N.D. processed data and determined and interpreted structures; V.C. built and refined models; N.D. and V.C. wrote the manuscript with input from all authors; V.R. oversaw the project and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Eight maps have been deposited into the Electron Microscopy Data Bank (EMDB) with the accession codes EMDB 11636, 11637, 11641, 11642, 11643, 11644, 11645, and 11646. Six atomic coordinates have been deposited into the Proteins Data Bank under the accession codes 7A5F, 7A5G, 7A5H, 7A5I, 7A5J, and 7A5K. The mass spectrometry samples report is available in the supplementary materials. Further information and material requests may be made to the corresponding author.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1105/suppl/DC1
Materials and Methods
Figs. S1 to S12
Table S1
References (*31*–*55*)
MDAR Reproducibility Checklist
Mass spectrometry data

View/request a protocol for this paper from *Bio-protocol*.

18 May 2020; accepted 24 September 2020
10.1126/science.abc7782

CORONAVIRUS

REGN-COV2 antibodies prevent and treat SARS-CoV-2 infection in rhesus macaques and hamsters

Alina Baum[1], Dharani Ajithdoss[1], Richard Copin[1], Anbo Zhou[1], Kathryn Lanza[1], Nicole Negron[1], Min Ni[1], Yi Wei[1], Kusha Mohammadi[1], Bret Musser[1], Gurinder S. Atwal[1], Adelekan Oyejide[1], Yenny Goez-Gazi[2], John Dutton[2], Elizabeth Clemmons[2], Hilary M. Staples[2], Carmen Bartley[2], Benjamin Klaffke[2], Kendra Alfson[2], Michal Gazi[2], Olga Gonzalez[2], Edward Dick Jr.[2], Ricardo Carrion Jr.[2], Laurent Pessaint[3], Maciel Porto[3], Anthony Cook[3], Renita Brown[3], Vaneesha Ali[3], Jack Greenhouse[3], Tammy Taylor[3], Hanne Andersen[3], Mark G. Lewis[3], Neil Stahl[1], Andrew J. Murphy[1], George D. Yancopoulos[1], Christos A. Kyratsous[1]*

An urgent global quest for effective therapies to prevent and treat coronavirus disease 2019 (COVID-19) is ongoing. We previously described REGN-COV2, a cocktail of two potent neutralizing antibodies (REGN10987 and REGN10933) that targets nonoverlapping epitopes on the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein. In this report, we evaluate the in vivo efficacy of this antibody cocktail in both rhesus macaques, which may model mild disease, and golden hamsters, which may model more severe disease. We demonstrate that REGN-COV-2 can greatly reduce virus load in the lower and upper airways and decrease virus-induced pathological sequelae when administered prophylactically or therapeutically in rhesus macaques. Similarly, administration in hamsters limits weight loss and decreases lung titers and evidence of pneumonia in the lungs. Our results provide evidence of the therapeutic potential of this antibody cocktail.

Fully human monoclonal antibodies (mAbs) are a promising class of therapeutics against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection (*1*). To date, multiple studies have described the discovery and characterization of potent neutralizing mAbs targeting the spike glycoprotein of SARS-CoV-2 (*2*–*11*). However, evaluation of the efficacy of these antibodies in vivo is only beginning to emerge and has largely focused on the prophylactic setting (*6*, *10*, *12*). Furthermore, because the animal models of SARS-CoV-2 infection and coronavirus disease 2019 (COVID-19) are still being developed, no single model has emerged as being more relevant for human disease. Indeed, based on the extremely diverse manifestations of COVID-19 in humans, multiple animal models may be needed to mimic various settings of human infection. The rhesus macaque model is widely used to assess efficacy of therapeutics and vaccines and displays a transient and mild course of the disease (*13*–*20*). On the contrary, the golden hamster model manifests a much more severe form of the disease, accompanied by rapid weight loss and severe lung pathology (*21*–*23*).

We previously described a cocktail of two fully human antibodies, REGN10933 and REGN10987, that bind to spike protein, potently neutralize SARS-CoV-2, and were selected as components of an antiviral antibody cocktail (REGN-COV2) to safeguard against mutational virus escape (*8*, *9*). In this study, we used two different animal models, rhesus macaque and golden hamster, that capture the diverse pathology of SARS-CoV-2 infection and evaluated the in vivo efficacy of this antibody cocktail when used prophylactically or therapeutically. This assessment allows us to compare the performance of the antibodies in diverse disease settings to more comprehensively understand the mechanisms by which mAb therapies may limit viral load and pathology in infected individuals.

To evaluate the ability of REGN-COV2 to protect rhesus macaques from SARS-CoV-2 infection, we initially assessed the impact of antibody administration before virus challenge [nonhuman primate (NHP) study 1]. Six animals were dosed with REGN-COV2 at 50 mg per kilogram of body weight (mg/kg) (25 mg/kg of each antibody) and six with placebo through intravenous administration and challenged with 1×10^5 plaque-forming units (PFU) of virus through intranasal and intratracheal routes 3 days after mAb dosing. Because of the relatively transient nature of the SARS-CoV-2 infection in rhesus macaques, the in-life portion of the study was limited to 5 days. To determine the impact of mAb prophylaxis on viral load in the upper and lower airways, we collected nasopharyngeal swabs on a daily basis and bronchoalveolar lavage (BAL) fluid on days 1, 3, and 5 after challenge (Fig. 1A). Both genomic RNA (gRNA) and subgenomic RNA (sgRNA) (which is made during replication) were measured to assess the impact of mAb prophylaxis on the dynamics of viral replication; whereas gRNA may reflect remaining viral inoculum as well as newly replicating virus, sgRNA should only result from newly replicating virus. For placebo-treated animals, the kinetics of viral load measures was as previously reported, with a peak in viral load on day 2 after challenge, although the majority of animals were still positive for viral RNA in nasal swabs on day 5; even though the kinetics of gRNA and sgRNA were similar, sgRNA levels were about a hundred-fold lower, consistent with what others have reported (*6*, *15*, *16*, *18*). For animals receiving REGN-COV2 prophylaxis, we observed accelerated clearance of gRNA with almost complete ablation of sgRNA in the majority of the animals, showing that REGN-COV2 can almost completely block establishment of virus infection; this pattern was observed across all measurements in both nasopharyngeal swabs and BAL compared with that from placebo animals, demonstrating that mAbs administered prophylactically can greatly reduce viral load in both the upper and lower airways (Fig. 1B).

A second prophylatic study (NHP study 2) was designed to test whether REGN-COV2 could protect against a 10-fold higher viral inoculum (1.05×10^6 PFU) and compared four animals treated with the 50 mg/kg dose of REGN-COV2 (25 mg/kg of each antibody) with four animals treated with a much lower dose of 0.3 mg/kg and four animals that were administered placebo (Fig. 2A). Nasopharyngeal and oral swabs were collected and used to measure viral gRNA and sgRNA. BAL samples were not collected in this study to minimize the potential impact of the procedure on histopathological analysis of the lung tissue. We observed that 50 mg/kg of REGN-COV2 administered 3 days before virus challenge was once again able to minimize virus replication even when animals were challenged with this 10-fold higher viral inoculum (Fig. 2B), whereas the prophylactic effect was greatly diminished with the 0.3 mg/kg dose. Interestingly, in this study we observed an increased impact of mAb treatment on viral load in oral swabs versus nasopharyngeal swabs, potentially indicating that mAb treatment may affect different physiological sources of virus replication differentially. Additional studies in animal models and humans will be needed to assess this.

Next, we assessed the impact of REGN-COV2 in the treatment setting by dosing four animals challenged with the higher 1×10^6 PFU of SARS-CoV-2 virus at 1 day after infection with 25 or 150 mg/kg of the antibody cocktail (Fig. 2A). By day 1 after challenge, the

[1]Regeneron Pharmaceuticals, Inc., Tarrytown, NY 10591, USA. [2]Southwest National Primate Research Center, Texas Biomedical Research Institute, San Antonio, TX 78245, USA. [3]BIOQUAL, Rockville, MD 20850, USA.
*Corresponding author. Email: christos.kyratsous@regeneron.com

animals had already reached peak viral load, as measured by both gRNA and sgRNA, mimicking a likely early treatment clinical scenario of COVID-19 disease, because it has been shown that most SARS-CoV-2–infected individuals reach peak viral loads relatively early in the disease course and often before or just at the start of symptom onset (*24*, *25*). Compared with four placebo treated animals, REGN-COV2–treated animals displayed accelerated viral clearance in both nasopharyngeal and oral swab samples, including both gRNA and sgRNA samples (Fig. 2C), clearly demonstrating that the monoclonal antibody cocktail can affect virus load even when administered after infection. Similar to the prophylactic study, the decrease in viral load appeared more dramatic in oral swabs versus nasopharyngeal swabs. Both treatment groups displayed similar kinetics of virus clearance, suggesting that 25 and 150 mg/kg doses demonstrate similar efficacy in this study. The treated animals in the 150 mg/kg group displayed about 10-fold higher titers on day 1, at the time of mAb administration, therefore potentially masking an enhanced effect of a higher drug dose. A similar impact of mAb treatment was observed on gRNA and sgRNA for both nasopharyngeal and oral samples, indicating that the mAb treatment is directly limiting viral replication in these animals (Fig. 2C).

The two antibody components of REGN-COV2 were selected to target nonoverlapping sites on the spike protein to prevent selection of escape mutants, which were readily detectable with a single-mAb treatment (*9*). To assess whether any signs of putative escape mutants are observed in an in vivo setting with authentic SARS-CoV-2 virus, we performed RNA sequencing (RNA-seq) analysis on all RNA samples obtained from all animals from the study. Analysis of the spike protein sequence identified mutations in NHP samples that were not present in the inoculum virus (fig. S1), further indicating that the virus is actively replicating in these animals. However, we did not observe any mutations that were specific to treated animals; all identified mutations were present either in the inoculum or in both treated and placebo animals, indicating that they were likely selected as part of virus replication in NHPs and were not selected by mAb treatment.

We next performed pathology analyses of the lungs of infected animals. All four placebo monkeys showed evidence of lung injury, which was characterized in three monkeys by interstitial pneumonia (Fig. 2D), with minimal to mild infiltration of mononuclear cells (lymphocytes and macrophages) in the septa, perivascular space, and/or pleura. In these three animals, the distribution of

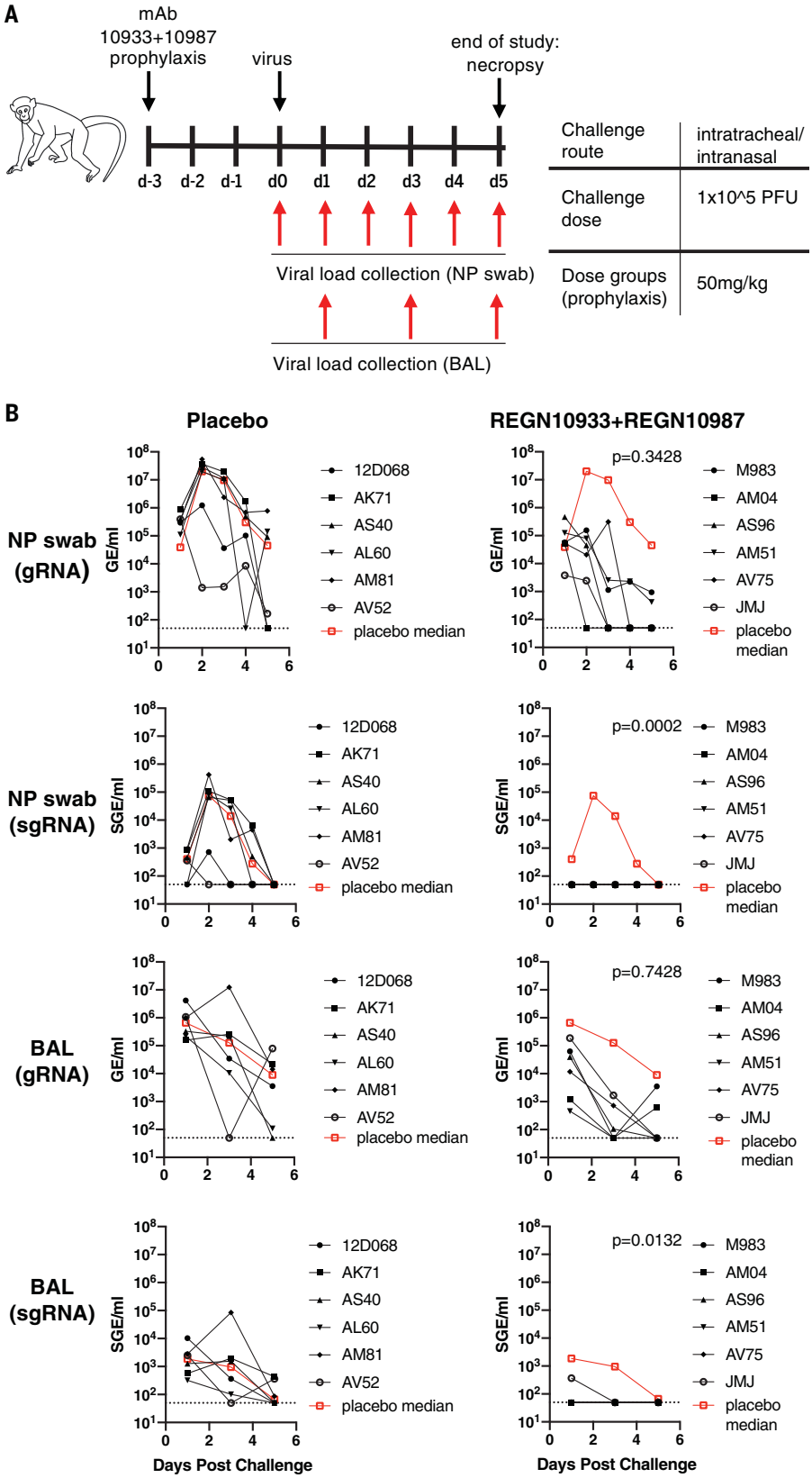
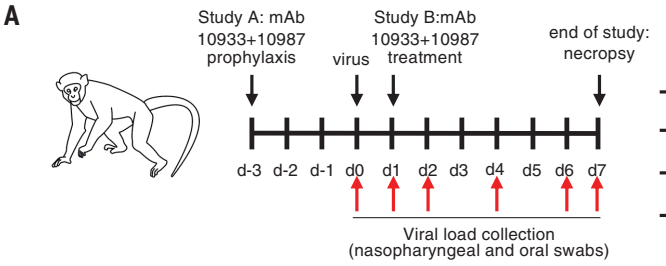
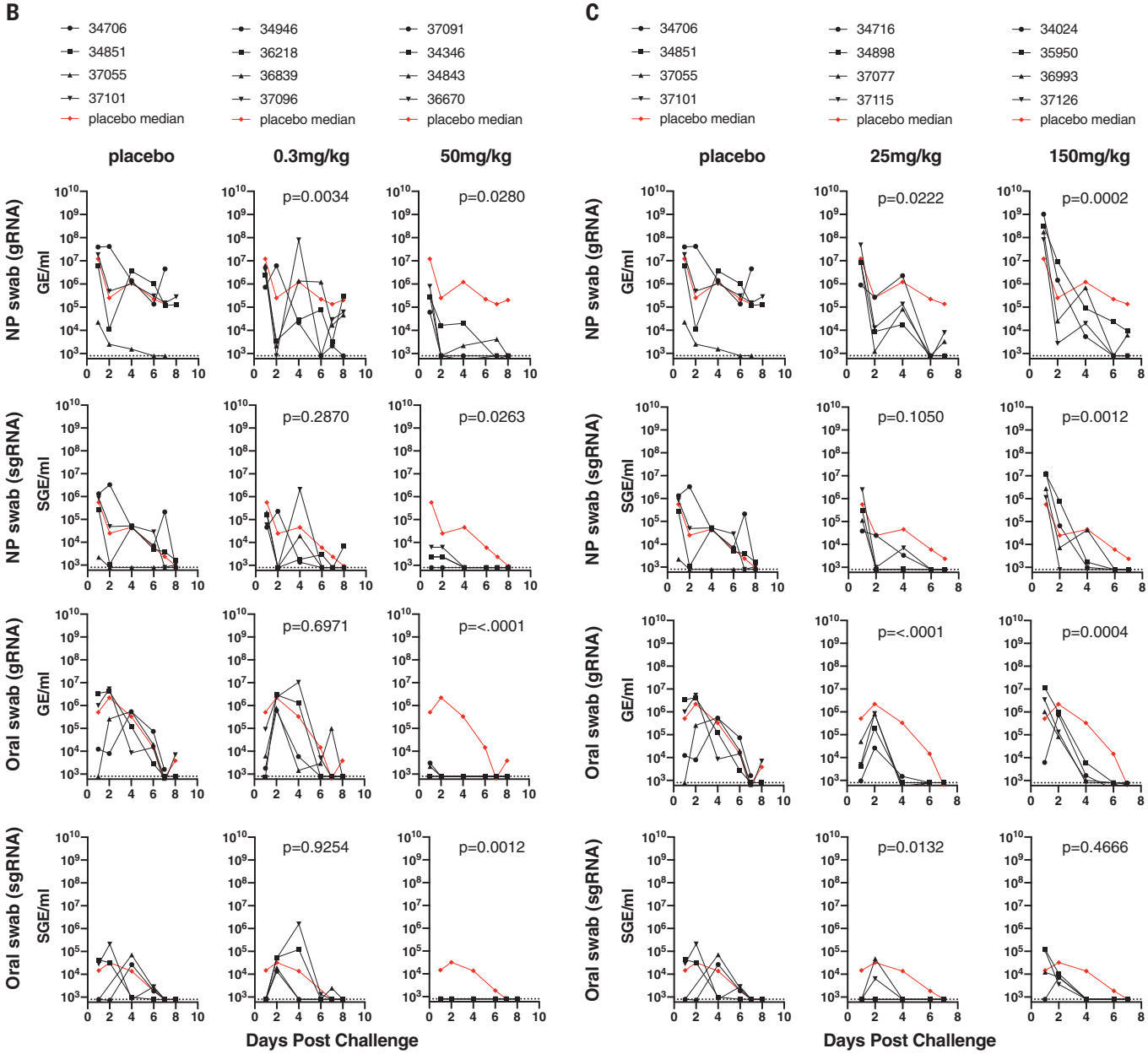


Fig. 1. Prophylactic efficacy of REGN-COV2 in the rhesus macaque model of SARS-CoV-2 infection (NHP study 1). (**A**) Overview of study design. d, day. (**B**) Impact of REGN-COV2 prophylaxis on viral gRNA and sgRNA in nasopharyngeal (NP) swabs and BAL fluid. The numbers in the graph legends represent animal codes. The dotted lines indicate limit of detection (LOD = 50 GE/ml for gRNA and LOD = 50 SGE/ml for sgRNA). For detailed statistical analysis, refer to tables S2 and S3. GE, genomic equivalents; SGE, subgenomic equivalents.



Challenge route	intratracheal/ intranasal
Challenge dose	1.05x10^6 PFU
Study A dose groups (prophylaxis)	0.3,50 (mg/kg)
Study B dose groups (treatment)	25,150 (mg/kg)



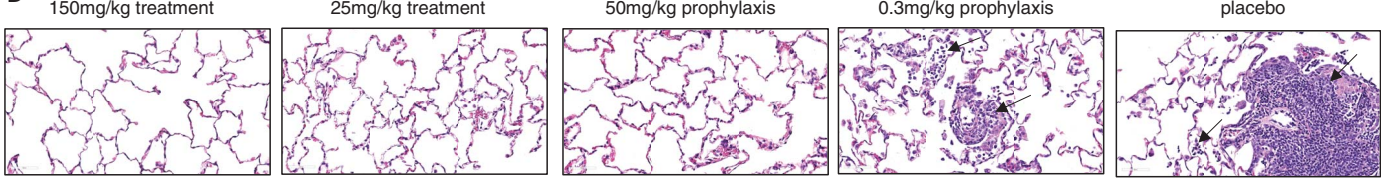


Fig. 2. Prophylactic and therapeutic efficacy of REGN-COV2 in the rhesus macaque model of SARS-CoV-2 infection (NHP study 2). (**A**) Overview of study design. (**B**) Impact of REGN-COV2 prophylaxis on viral gRNA and sgRNA in nasopharyngeal swabs and oral swabs [Study A, as shown in (A)]. (**C**) Impact of REGN-COV2 treatment on viral gRNA and sgRNA in nasopharyngeal swabs and oral swabs [Study B, as shown in (A)]. In (B) and (C), the numbers in the graph legends represent animal codes, and the dotted lines indicate limit of detection (LOD = 800 GE/ml for gRNA and LOD = 800 SGE/ml for sgRNA). (**D**) Representative images of histopathology in lungs of treated and placebo animals. The black arrows point to inflammatory cells. For detailed statistical analysis, refer to tables S2 and S3.

lesions was multifocal and involved two to three of the four lung lobes. Accompanying these changes were alveolar infiltration of lymphocytes, increased alveolar macrophages, and syncytial cells. Type II pneumocyte hyperplasia was also observed in occasional alveoli. In the fourth placebo monkey, lung injury was limited to type II pneumocyte hyperplasia, suggestive of a reparative process secondary to type I pneumocyte injury. Overall, the histological lesions observed in the placebo animals were consistent with an acute SARS-CoV-2 infection. In the prophylactic groups, three of four animals in the low-dose (0.3 mg/kg) and one of four animals in the high-dose (50 mg/kg) groups showed evidence of interstitial pneumonia (table S1) that was generally minimal and with fewer histological features when compared with that of the placebo group. In the one affected high-dose group animal, only one of the four lung lobes had a minimal lesion. In the therapeutic treatment groups, two of four low-dose (25 mg/kg) and two of four high-dose (150 mg/kg) treated animals showed evidence of interstitial pneumonia. In all affected low- and high-dose animals, only one of four lung lobes had lesions. Finally, there were no drug-related toxicities observed at any of the doses tested. In summary, the incidence of interstitial pneumonia (the number of animals as well as the number of lung lobes affected) and the severity were reduced in both prophylactic and therapeutic treatment modalities compared with placebo. The analyses demonstrate that prophylactic and therapeutic administration of REGN-COV2 greatly reduced virus-induced pathology in rhesus macaques and showed a clean safety profile.

Unlike rhesus macaques, which present with a mild clinical course of disease and transient virus replication when infected with SARS-CoV-2 that may mimic mild human disease, the golden hamster model is more severe, with animals demonstrating readily observable clinical disease, including rapid weight loss accompanied by very high viral load in the lungs, as well as severe lung pathology. Thus, this model may more closely mimic more severe disease in humans, although more extensive characterization of this model and severe human disease is needed to better understand similarities and differences in pathology. To evaluate the ability of REGN-COV2 to alter the disease course in this model, we designed a study that evaluated the prophylactic and treatment efficacy of the antibodies (Fig. 3A). In the prophylactic study, 25 hamsters were divided into five arms (five animals in each). Administration of 50, 5, or 0.5 mg/kg of REGN-COV2 at 2 days before challenge with a 2.3×10^4 PFU dose of SARS-CoV-2 virus resulted in dramatic protection from weight loss at all doses. This protection was accompanied by decreased viral load in the lungs at the end of the study in the majority of treated animals (day 7 after infection) (Fig. 3C). Evaluation of lung tissue from infected hamsters that were prophylactically treated with placebo or isotype control drug revealed distorted alveoli lined by swollen, hyperplastic type II pneumocytes interspersed with occasional type I single-cell necrosis and alveolar spaces that were filled with large numbers of lymphocytes, macrophages, and neutrophils, occasional syncytial cells, and hemorrhage. These changes were accompanied by variably severe interstitial pneumonia characterized by mixed-cell inflammation (lymphocytes, macrophages, and neutrophils) in the alveolar septa and perivascular spaces accompanied by edema and septal fibrosis. The severity and incidence of alveolar infiltration and interstitial pneumonia were greatly reduced in animals that received REGN-COV2 (Fig. 3D). Compared with placebo- and isotype-treated animals, the percent area of pneumonia in the lungs determined using HALO image analysis software was significantly reduced in all REGN-COV2–treated animals irrespective of doses. Intriguingly, we did observe high gRNA and sgRNA levels in the lungs of a few treated animals, although these individual animals did not show decreased protection from weight loss or more extensive pathology than the animals with much lower viral loads. It is possible that mAb treatment may provide an additional therapeutic benefit in this model that is not directly associated with viral load decrease. Alternatively, it is possible that the increased amounts of detected viral RNA may not necessarily be associated with infectious virus. Because viral replication and lung pathology in the hamster model occur very rapidly, the treatment setting represents a high bar for demonstrating therapeutic efficacy. We used 25 hamsters (five in each of five arms) in a therapeutic study and were able to observe therapeutic benefit in animals treated with 50 and 5 mg/kg doses of REGN-COV2 combination 1 day after viral challenge (Fig. 3B). Taken together, the two hamster studies clearly demonstrate that REGN-COV2 can alter the course of infection in the hamster model of SARS-COV-2 when administered either prophylactically or therapeutically.

In this study, we assessed the in vivo prophylactic and treatment efficacy of the REGN-COV2 mAb cocktail in two animal models, one of mild disease in rhesus macaques and one of severe disease in golden hamsters. Our results demonstrate that the antibodies are efficacious in both animal models, as measured by reduced viral load in the upper and lower airways, by reduced virus-induced pathology in the rhesus macaque model, and by limited weight loss in the hamster model.

The ability of REGN-COV2 to almost completely block detection of subgenomic species of SARS-COV-2 RNA in rhesus macaques matches or exceeds the effects recently shown in vaccine efficacy studies using the same animal models (*18*–*20*, *26*, *27*). Additionally, the observed accelerated reduction of upper-airway virus load in rhesus macaques treated with REGN-COV2 contrasts with the lack of impact on viral load in remdesivir-treated animals, where reduced viral load could only be observed in lower airways with no differences in nasal viral RNA levels (*28*). These findings highlight the therapeutic potential of REGN-COV2 to both protect from and treat SARS-COV-2 disease. Additionally, the impact of REGN-COV2 prophylaxis on viral RNA levels in nasopharyngeal and oral swabs may indicate the potential not only to prevent disease in the exposed individual but also to limit transmission.

Importantly, in our studies, we did not observe any signs of increased viral load or worsening of pathology in the presence of antibodies at either high or low doses in either animal model. Potential for antibody-mediated enhancement of disease is a serious concern for antibody-based therapeutics and vaccines. And although a recent report showed the ability of some anti-spike mAbs to mediate pseudovirus entry into Fcγ receptor–expressing cell lines, these data do not address whether similar behavior would be observed with authentic SARS-CoV-2 virus

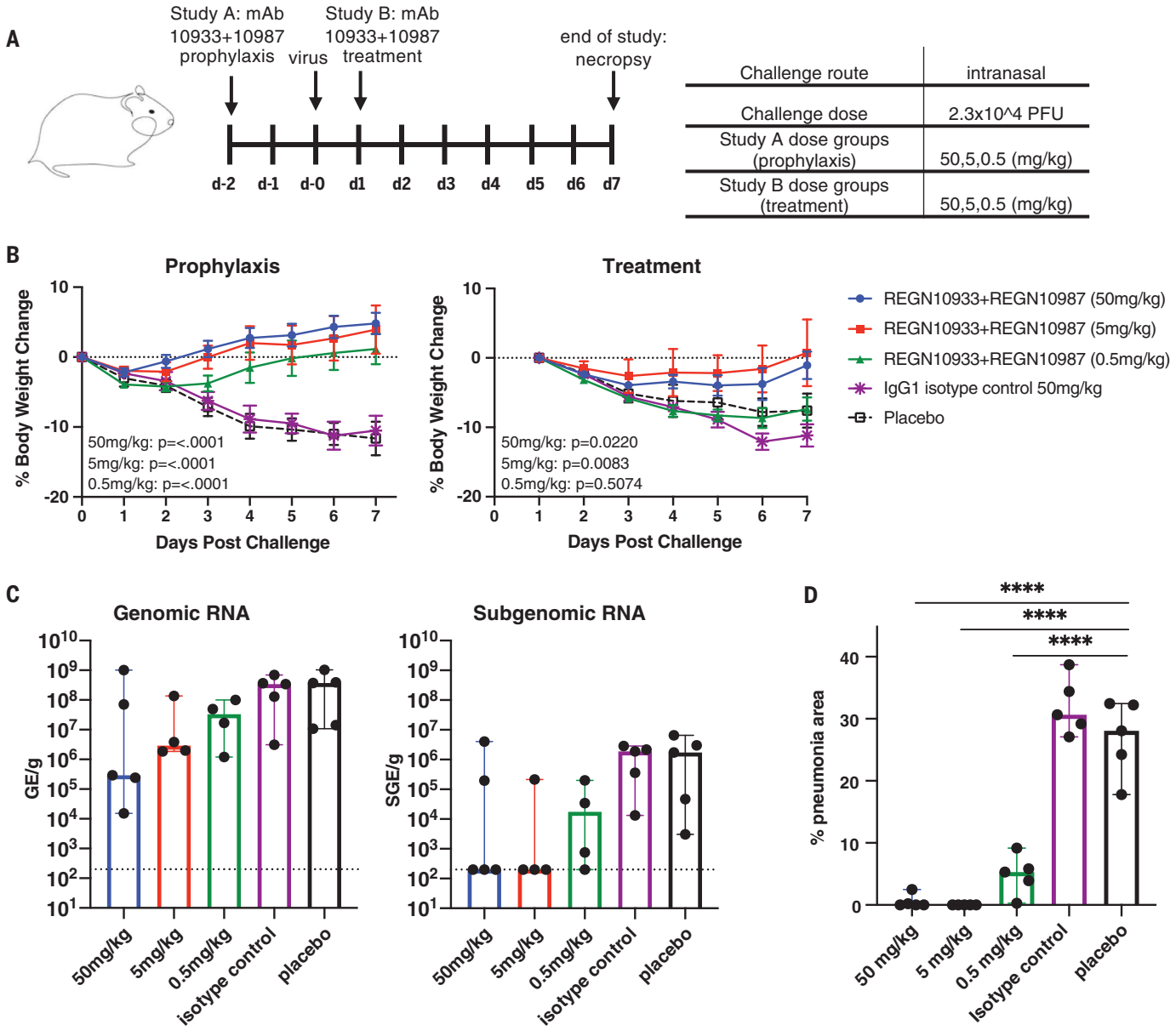


Fig. 3. Efficacy of REGN-COV2 in treatment and prophylaxis in the golden Syrian hamster model of SARS-CoV-2 infection. (**A**) Overview of study design. (**B**) Impact of REGN-COV2 on weight loss in prophylaxis and treatment groups. Error bars represent mean with error. IgG1, immunoglobulin G1. (**C**) Impact of REGN-COV-2 prophylaxis on levels of gRNA and sgRNA in hamster lungs (7 days after infection). No statistical significance was observed between any treatment groups and placebo. The dotted lines indicate limit of detection (LOD = 200 GE/ml for gRNA and LOD = 200 SGE/ml for sgRNA), and error bars represent median with 95% confidence intervals. (**D**) Impact of REGN-COV2 prophylaxis on percent area of lung exhibiting pathology typical of pneumonia (****$p <$ 0.0001 indicates significant differences). Error bars represent median with 95% confidence intervals. For detailed statistical analysis, refer to tables S4 and S5.

and primary immune cells (*29*). Our results are consistent with no evidence of enhanced disease in clinical studies that assess convalescent plasma therapy (*30*).

Similarly to most in vivo data generated to date, our in vivo studies were conducted with the D614 spike protein variant of the SARS-CoV-2 virus. A global shift in circulating SARS-CoV-2 to the D614G variant will likely necessitate a transition to use of that variant for in vitro and in vivo studies with SARS-CoV-2 virus in the future (*31*). It is yet not established if pathogenicity and replication dynamics of this variant differ in vivo, and it is equally unclear whether there is an association with severity of human infections (*32*–*34*). Importantly, we have previously demonstrated that the neutralization potency of REGN10933 and REGN10987, as well as the REGN-COV2 combination, was not altered in the presence of this variant, making it likely that the efficacy of the REGN-COV2 combination will extend to the 614G virus (*8*, *35*). Our data provide evidence that REGN-COV2–based therapy may offer clinical benefit in both prevention and treatment settings of COVID-19 disease, where it is currently being evaluated (clinicaltrials.gov NCT04426695, NCT04425629, and NCT04452318).

REFERENCES AND NOTES

1. J. Cohen, *Science* **368**, 564–565 (2020).
2. B. Ju *et al.*, *Nature* **584**, 115–119 (2020).

3. D. Pinto *et al.*, *Nature* **583**, 290–295 (2020).
4. C. Wang *et al.*, *Nat. Commun.* **11**, 2251 (2020).
5. Y. Cao *et al.*, *Cell* **182**, 73–84.e16 (2020).
6. S. J. Zost *et al.*, *Nat. Med.* **26**, 1422–1427 (2020).
7. D. F. Robbiani *et al.*, *Nature* **584**, 437–442 (2020).
8. J. Hansen *et al.*, *Science* **369**, 1010–1014 (2020).
9. A. Baum *et al.*, *Science* **369**, 1014–1018 (2020).
10. L. Liu *et al.*, *Nature* **584**, 450–456 (2020).
11. E. Seydoux *et al.*, *Immunity* **53**, 98–105.e5 (2020).
12. R. Shi *et al.*, *Nature* **584**, 120–124 (2020).
13. V. J. Munster *et al.*, *Nature* **585**, 268–272 (2020).
14. C. Shan *et al.*, *Cell Res.* **30**, 670–677 (2020).
15. J. Yu *et al.*, *Science* **369**, 806–811 (2020).
16. A. Chandrashekar *et al.*, *Science* **369**, 812–817 (2020).
17. W. Deng *et al.*, *Science* **369**, 818–823 (2020).
18. N. B. Mercado *et al.*, *Nature* **586**, 583–588 (2020).
19. K. S. Corbett *et al.*, *N. Engl. J. Med.* NEJMoa2024671 (2020).
20. N. van Doremalen *et al.*, *Nature* **586**, 578–582 (2020).
21. S. F. Sia *et al.*, *Nature* **583**, 834–838 (2020).
22. M. Imai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 16587–16595 (2020).
23. T. F. Rogers *et al.*, *Science* **369**, 956–963 (2020).
24. L. Zou *et al.*, *N. Engl. J. Med.* **382**, 1177–1179 (2020).
25. X. He *et al.*, *Nat. Med.* **26**, 672–675 (2020).
26. Q. Gao *et al.*, *Science* **369**, 77–81 (2020).
27. A. Patel *et al.*, bioRxiv 2020.07.28.225649 [Preprint]. 29 July 2020; https://doi.org/10.1101/2020.07.28.225649.
28. B. N. Williamson *et al.*, *Nature* **585**, 273–276 (2020).
29. S. Wang *et al.*, bioRxiv 2020.07.26.222257 [Preprint]. 27 July 2020; https://doi.org/10.1101/2020.07.26.222257.
30. L. Li *et al.*, *JAMA* **324**, 460–470 (2020).
31. B. Korber *et al.*, *Cell* **182**, 812–827.e19 (2020).
32. Y. Toyoshima, K. Nemoto, S. Matsumoto, Y. Nakamura, K. Kiyotani, *J. Hum. Genet.* **65**, 1075–1082 (2020).
33. E. M. Volz *et al.*, medRxiv 2020.07.31.20166082 [Preprint]. 1 September 2020; https://doi.org/10.1101/2020.07.31.20166082.
34. B. W.-Y. Mok *et al.*, bioRxiv 2020.08.28.271635 [Preprint]. 28 August 2020; https://doi.org/10.1101/2020.08.28.271635.
35. L. Yurkovetskiy *et al.*, bioRxiv 2020.07.04.187757 [Preprint]. 16 July 2020; https://doi.org/10.1101/2020.07.04.187757.

ACKNOWLEDGMENTS

The following reagent was deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate USA-WA1/2020, NR-52281. **Funding:** A portion of this project has been funded in whole or in part with federal funds from the Department of Health and Human Services, Office of the Assistant Secretary for Preparedness and Response, Biomedical Advanced Research and Development Authority, under OT number HHSO100201700020C. **Author contributions:** A.B., N.S., A.J.M., G.D.Y., and C.A.K. conceptualized and designed experiments. Y.G.-G., J.D., E.C., H.M.S., C.B., B.K., O.G., E.D., L.P., M.P., A.C., R.B., V.A., J.G., and T.T. performed experiments, and A.B., R.C., D.A., A.O., K.A., R.Ca., M.G., H.A., M.G.L., G.S.A., G.D.Y., and C.A.K. analyzed data. R.Co., K.L., N.N., M.N., and Y.W. prepared sequencing libraries and performed bioinformatics analysis. A.B. and C.A.K. wrote the paper. C.A.K. acquired funding. **Competing interests:** Regeneron authors own options and/or stock of the company. This work has been described in one or more pending provisional patent applications. N.S., A.J.M., G.D.Y., and C.A.K. are officers of Regeneron. **Data and materials availability:** All data are available in the main text or supplementary materials. Regeneron materials described in this manuscript may be made available to qualified, academic, noncommercial researchers through a material transfer agreement upon request at https://regeneron.envisionpharma.com/vt_regeneron/. For questions about how Regeneron shares materials, use the email address preclinical.collaborations@regeneron.com.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1110/suppl/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S5
References (*36*, *37*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

8 August 2020; accepted 7 October 2020
Published online 9 October 2020
10.1126/science.abe2402

SEX DETERMINATION

A chimeric gene paternally instructs female sex determination in the haplodiploid wasp *Nasonia*

Yuan Zou[1], Elzemiek Geuverink[1], Leo W. Beukeboom[1], Eveline C. Verhulst[1,2,3]*, Louis van de Zande[1]*†

Various primary signals direct insect sex determination. In hymenopteran insects, the presence of a paternal genome is needed to initiate female development. When absent, uniparental haploid males develop. We molecularly and functionally identified the instructor sex-determination gene, *wasp overruler of masculinization* (*wom*), of the haplodiploid wasp *Nasonia vitripennis*. This gene contains a P53-like domain coding region and arose by gene duplication and genomic rearrangements. Maternal silencing of *wom* results in male development of haploid embryos. Upon fertilization, early zygotic transcription from the paternal *wom* allele is initiated, followed by a timely zygotic expression of *transformer* (*tra*), leading to female development. *Wom* is an instructor gene with a parent-of-origin effect in sex determination.

Insect sex determination is characterized by a large variation of primary instruction signals. In many species, this instruction is transduced by the transformer protein (TRA). Female-specific splicing of *tra* premRNA yields a functional TRA protein (TRA-F), which splices *doublesex* (*dsx*) transcripts into a female-specific isoform, leading to female development (*1*–*4*). Absence of TRA-F results in male-specific splicing of *dsx* transcripts and male development. TRA-F also regulates *tra* transcript splicing into the female-specific isoform, thus establishing an autoregulatory loop of functional TRA-F production (*5*, *6*).

In many Diptera, dominant male-determining factors (*7*–*11*) instruct sex determination. In *Drosophila*, the dose of X-linked factors is the primary signal (*12*). Hymenoptera have haplodiploid sex determination: Haploid males develop from unfertilized eggs; diploid females develop from fertilized eggs. The only molecularly characterized instructor in Hymenoptera is *complementary sex determiner* (*csd*) of the honey bee (*13*–*15*), where female-specific splicing of transcripts from the *tra* ortholog, *feminizer* (*fem*), is needed for female development (*15*, *16*). The multi-allelic *csd* gene, a paralog of *fem*, contains a hypervariable region and instructs female development when heterozygous and male development when homozygous or hemizygous (*15*, *16*). However, most parasitoid wasps, including *Nasonia*, do not have *csd* (*17*), and the instructor signal in those haplodiploids is unknown.

For *Nasonia vitripennis*, a sex determination model was proposed in which a primary instructor gene is maternally silenced in unfertilized eggs, whereas fertilized eggs receive a nonsilenced paternal allele (*18*, *19*). This maternally silenced instructor gene is not *tra* itself, as in embryos from fertilized eggs both the

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Post Office Box 11103, 9700 CC Groningen, Netherlands. [2]Laboratory of Genetics, Plant Sciences Group, Wageningen University, Netherlands. [3]Laboratory of Entomology, Plant Sciences Group, Wageningen University, Netherlands.
*These authors contributed equally to this work.
†Corresponding author. Email: louis.van.de.zande@rug.nl

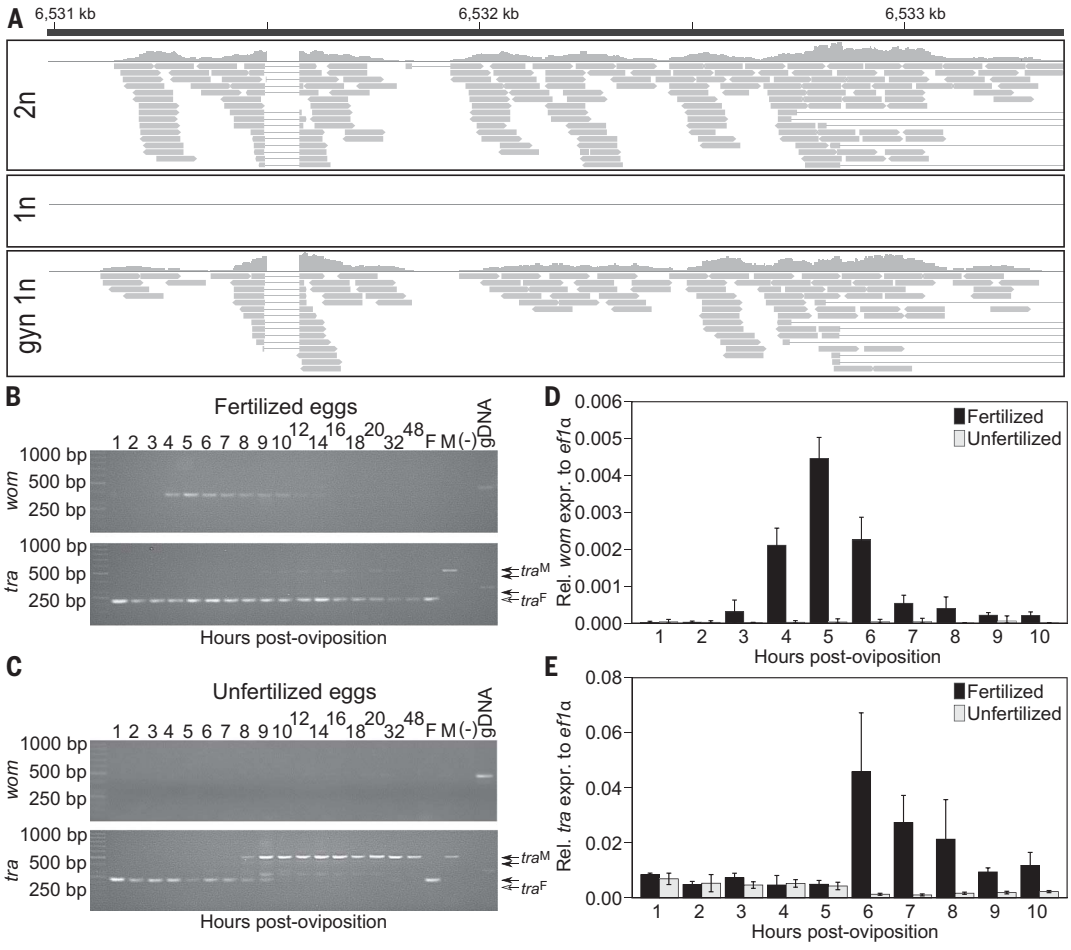


Fig. 1. Transcriptome analysis of early *Nasonia* embryos. (**A**) RNA-sequencing coverage (histograms) and read alignments (horizontal bars) from diploid female (2n, top), haploid male (1n, middle), and haploid gynandromorphic embryos (1n, bottom), located on chromosome 1, 6531.0-6533.5kb (NC_015867.2, reference *N. vitripennis* genome Nvit_2.1). (**B** and **C**) Temporal RT-PCR profiles of *wom* and *tra* for fertilized (B) and unfertilized eggs (C). Numbers indicate hours post-oviposition; F and M, adult female and male samples; (-), control without cDNA template; gDNA, adult male genomic DNA. Arrows indicate male-specific (filled) and female-specific (open) *tra* transcripts. (**D** and **E**). Temporal expression profiles of *wom* (D) and *tra* (E) for fertilized and unfertilized eggs. Reference gene: *EF1α*. Error bars: SEM.

maternal and paternal *tra* allele are expressed (*20*). Therefore, we suggested that an instructor gene is only expressed after fertilization in early diploid embryos and activates zygotic *tra* transcription (*21*, *22*). This initiates *tra* autoregulation, leading to female development.

To identify this regulatory gene, we sequenced transcriptomes of early embryos (see supplementary materials and methods for details). Haploid embryos (males) were obtained from unmated females; diploid embryos (females) were obtained from mated females, although their clutches contain ~20% unfertilized haploid eggs as well. Additionally, we used haploid embryos from unmated females of a gynandromorph strain (HiCD12) of which ~40% develop into adults with female characteristics (*23*).

Zygotic *tra* expression starts at 5 hours postoviposition (hpo) in the syncytial stage (*19*). Therefore, transcriptome analysis was performed on 2- to 5-hpo embryos. Differentially expressed gene (DEG) analysis yielded only two loci, *tra* and *LOC103317656* (reference *Nasonia vitripennis* genome Nvit_2.1, Annotation release 101), to be more highly expressed in diploid than in haploid embryos. Both loci were also among the more highly expressed genes in gynandromorphic embryos compared to normal haploid embryos (table S1). The differential expression of *tra* is in agreement with previous results and is an essential feature of *Nasonia* sex determination (*19*). Many reads from diploid and gynandromorphic embryos, but no reads from haploid embryos, mapped to the *LOC103317656* region (Fig. 1A). Therefore, *LOC103317656* on chromosome 1 is a compelling candidate for the regulatory gene, which we call *wasp overruler of masculinization* (*wom*).

Wom consists of three exons of 428, 268, and 1512 base pairs (bp), interrupted by two introns of 79 and 87 bp. It has a single splice form, and the coding sequence translates into a protein of 580 amino acids. The residues 91 to 285 form a P53-like DNA binding domain (fig. S1) that shares 30 to 43% identity with P53-family homologs and contains conserved zinc-binding, dimerization, and DNA binding motifs. These features correspond to the functional elements of P53 proteins for gene regulation and transcription in mammals (*24*–*26*). WOM also contains a coiled-coil domain at the C-terminal region (residues 550 to 577) (fig. S1), which may function in gene regulation (*27*, *28*).

Temporal expression analyses, with *tra* as internal control, revealed that *wom* is expressed in early diploid embryos but not in haploid embryos, demonstrating that it is not maternally provided (Fig. 1, B to D, and fig. S2A). Zygotic transcription of *wom* starts 2 to 3 hpo (blastoderm stage) and peaks at 4 to 5 hpo (Fig. 1D and fig. S2A), consistent with the DEG analysis. It coincides with the onset of zygotic *tra* expression that peaks at 6 to 7 hpo (*19*) (Fig. 1E and fig. S2B). *Wom* expression abruptly declines at 6 to 7 hpo (Fig. 1D and fig. S2A) and is not expressed during later female or male life stages (fig. S3). The sequential *wom* and *tra* expression within a defined time window is consistent with its proposed function of initiating zygotic *tra* transcription.

We determined the allelic origin of the *wom* transcripts using a synonymous single-nucleotide polymorphism (SNP) in exon 3 of *wom* (fig. S4A) that results in an Nhe I restriction fragment polymorphism. Sequencing and Nhe I digestion of reverse transcription–polymerase chain reaction (RT-PCR) products from the progeny of reciprocal crosses (Fig. 2A) demonstrated that *wom* mRNA is transcribed only from the paternal allele in diploid embryos (Fig. 2B and fig. S4B), in line with the hypothesis that *wom* is maternally silenced.

To confirm that *wom* is involved in initiating zygotic *tra* expression, we silenced *wom* in early diploid embryos by parental RNA interference (pRNAi) (*29*, *30*). This resulted in a reduction of both *wom* and *tra* expression levels in *wom* pRNAi diploid embryos, at 4 to 5 hpo for *wom* [$P < 0.001$; Welch's analysis of variance (ANOVA) with Tukey's test; Fig. 3A and fig. S5A] and 6 to 8 hpo for *tra* ($P < 0.05$; Welch's ANOVA with Tukey's test; Fig. 3B and fig. S5B). Furthermore, the diploid embryos had male-specific *tra* transcripts (fig. S5C) and developed as fully fertile males (Fig. 3C, fig. S6, and table S3). *Wom*-silenced females produced an increased number of haploid sons. This might be a derived *wom* silencing effect, as females can adjust the fertilization rate of their eggs. Conversely, *tra* pRNAi had no effect on the early embryonic *wom* expression in diploid embryos ($P > 0.05$; Welch's ANOVA with Tukey's test; Fig. 3A and fig. S5A), but it reduced *tra* expression levels ($P < 0.05$; Welch's ANOVA with Tukey's test; Fig. 3B and fig. S5B). This confirms that *wom* acts upstream of *tra* in the *N. vitripennis* sex-determining cascade and is essential for initializing zygotic *Nvtra* expression.

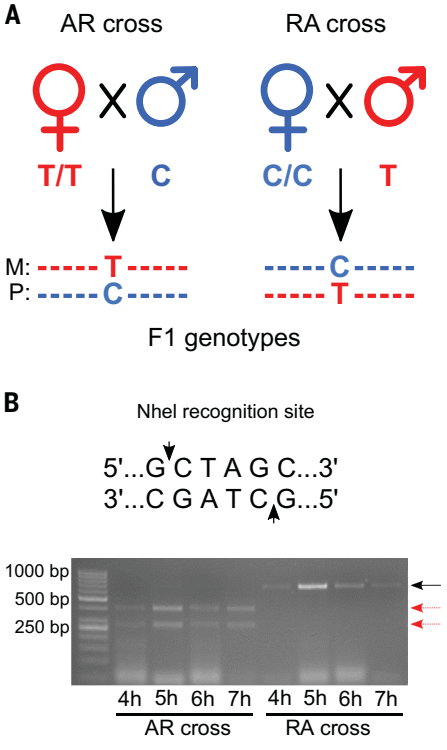
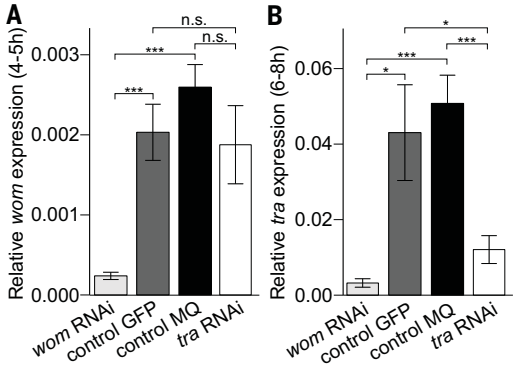


Fig. 2. *Wom* is expressed from the paternal allele only. (**A**) Crossing scheme to test the allelic origin of the *wom* transcripts. M, maternal; P, paternal. (**B**) Restriction fragment length polymorphism analysis of F_1 embryos 4 to 7 hpo. Arrows indicate undigested (black) and digested (red) PCR products.



C

Injection	P: ♀(mated)	F1: haploid♂	F1: diploid♂	F1: diploid♀
wom	80	3237	1865	0
tra	60	557	1662	0
GFP	50	664	0	2672
MilliQ	50	839	0	2543

Fig. 3. *Wom* and *tra* expression in early zygotes from dsRNA-treated females. *Wom* (**A**) and *tra* (**B**) expression in 4- to 5-hpo embryos of mated *wom* double-stranded RNA (dsRNA)– or *tra* dsRNA–injected females. Error bars depict SEM. $*P \leq 0.05$, $***P \leq 0.001$; n.s., not significant. (**C**) Number of *wom* dsRNA– or *tra* dsRNA–injected females and their offspring number. GFP (green fluorescent protein)– and MilliQ-injected females served as controls.

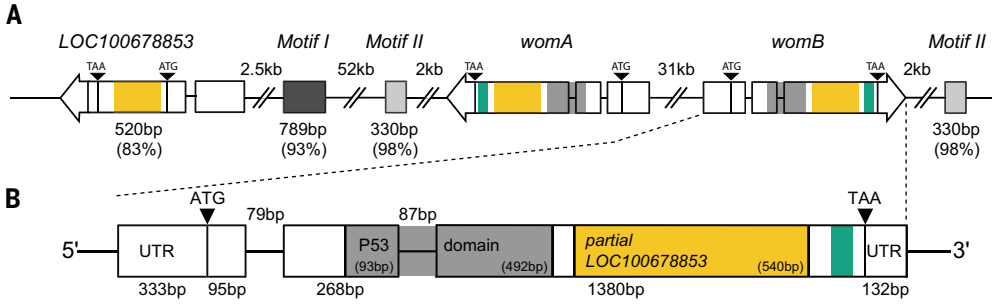


Fig. 4. Genomic organization of the *wom* locus. (**A**) Two copies, *womA* and *womB*, are present in antiparallel orientation. The medium gray block represents the p53-like domain. The downstream yellow block represents the partial *LOC100678853* sequence (yellow). P53-like domain sequence motifs I (dark gray box) and II (light gray box) are located between *womA* and *LOC100678853*. Motif II has an identical copy 2 kb downstream of *womB*. (**B**) Detailed structure of *womB*. Boxes represent exons, lines represent introns. The green block represents the coiled-coil region. UTR, untranslated region.

The genomic organization of *wom* is complex. The gene is duplicated as two antiparallel oriented copies (*womA* and *womB*) separated by an intergenic region of 31 kb (Fig. 4A). Both copies are transcribed (fig. S7, A to C). Downstream of the P53-like domain, a region of 540 bp (994 to 1533) shares 83% nucleotide identity with a 520-bp region of the annotated *Nasonia*-specific gene *LOC100678853* (reference *Nasonia vitripennis* genome Nvit_psr_1.1, Annotation release 104) (Fig. 4B). This is a gene of unknown function, 58.5 kb downstream of *womA*. The intergenic region between *womA* and *LOC100678853* contains two motifs interspersed by 52 kb (Fig. 4A). Motif I (789 bp) has 93% similarity and motif II (330 bp) 98% similarity with the P53-like domain of *wom* (Fig. 4A). The complete sequence of motif II is included in motif I. An identical copy of motif II is located 2 kb downstream of *womB* (Fig. 4A). Apparently, this region of chromosome 1 has been a site of dynamic genomic rearrangements.

Wom and *LOC100678853* homologs were only detected in *Nasonia* species and their close relative *Trichomalopsis* (fig. S8). By contrast, *wom*-like genes lacking the partial *LOC100678853* homologous region were only found in three other pteromalids (*Cecidostiba fungosa*, *Cecidostiba semifascia*, and *Pteromalus puparum*) (fig. S8). These species also lacked the entire *LOC100678853* gene. Sequence alignment and comparison of nonsynonymous to synonymous substitutions suggest that *wom* is an evolutionary new, chimeric, gene (supplementary text and table S5).

The molecular identification of primary sex determination genes has proved to be challenging (*7*–*11*). The existence of a female instructor gene in *Nasonia* was predicted (*19*, *20*), but its identity remained unknown. Here, we demonstrate that *wom* is a de novo evolved instructor gene with a parent-of-origin effect. It is maternally silenced, but is transcribed from the paternal allele in fertilized eggs, to initiate female development.

The existence of *wom* also may be relevant in explaining two cases of uniparental female offspring in *Nasonia* (*23*, *31*). *Wom* is not completely maternally silenced in haploid HiCD12 early embryos (fig. S9A) that partially develop as females (*23*). It is unlikely that this is due to a *wom* mutation, as no potential functional differences were detected in the HiCD12 gene sequence or its 2.5-kb upstream region (fig. S10). We hypothesize that the putative *gynandromorph* (*gyn*) gene, which maps to chromosome 4, is involved in maternal silencing of *wom*, and a loss-of-function mutation in HiCD12 leads to (partial) failure of imprinting.

Uniparental female offspring have been observed in a polyploid line (*32*). Triploid females occasionally produce diploid female offspring without fertilization (*31*). This may be due to dosage disruption in polyploid females, possibly combined with a *wom* mutation, resulting in defective *wom* imprinting.

Requirement of the paternal genome for female development is characteristic for haplodiploid reproduction. *Wom* is identified here as an instructor gene for female development with a parent-of-origin effect. Like *csd* in the honey bee, it is located in a dynamic genomic region (*16*, *33*). This suggests that, within Hymenoptera, various genes have evolved upstream of *tra* to enable paternally mediated onset of female development.

REFERENCE AND NOTES

1. K. C. Burtis, B. S. Baker, *Cell* **56**, 997–1010 (1989).
2. R. T. Boggs, P. Gregor, S. Idriss, J. M. Belote, M. McKeown, *Cell* **50**, 739–747 (1987).
3. K. Hoshijima, K. Inoue, I. Higuchi, H. Sakamoto, Y. Shimura, *Science* **252**, 833–836 (1991).
4. M. Tian, T. Maniatis, *Cell* **74**, 105–114 (1993).
5. A. Pane, M. Salvemini, P. Delli Bovi, C. Polito, G. Saccone, *Development* **129**, 3715–3725 (2002).
6. M. Hediger *et al*., *Genetics* **184**, 155–170 (2010).
7. A. B. Hall *et al*., *Science* **348**, 1268–1270 (2015).
8. E. Krzywinska, N. J. Dennison, G. J. Lycett, J. Krzywinski, *Science* **353**, 67–69 (2016).
9. F. Criscione, Y. Qi, Z. Tu, *eLife* **5**, e19281 (2016).
10. A. Sharma *et al*., *Science* **356**, 642–645 (2017).
11. A. Meccariello *et al*., *Science* **365**, 1457–1460 (2019).
12. S. M. Parkhurst, D. Bopp, D. Ish-Horowicz, *Cell* **63**, 1179–1191 (1990).
13. O. Mackensen, *Genetics* **36**, 500–509 (1951).
14. R. H. Crozier, *Am. Nat.* **105**, 399–412 (1971).
15. M. Beye, M. Hasselmann, M. K. Fondrk, R. E. Page Jr., S. W. Omholt, *Cell* **114**, 419–429 (2003).
16. M. Hasselmann *et al*., *Nature* **454**, 519–522 (2008).
17. A. R. Whiting, *Q. Rev. Biol.* **42**, 333–406 (1967).
18. L. W. Beukeboom, A. Kamping, L. van de Zande, *Semin. Cell Dev. Biol.* **18**, 371–378 (2007).
19. E. C. Verhulst, L. W. Beukeboom, L. van de Zande, *Science* **328**, 620–623 (2010).
20. E. C. Verhulst, J. A. Lynch, D. Bopp, L. W. Beukeboom, L. van de Zande, *PLOS ONE* **8**, e63618 (2013).
21. L. W. Beukeboom, L. van de Zande, *J. Genet.* **89**, 333–339 (2010).
22. L. van de Zande, E. C. Verhulst, *Sex Dev.* **8**, 74–82 (2014).
23. A. Kamping, V. Katju, L. W. Beukeboom, J. H. Werren, *Genetics* **175**, 1321–1333 (2007).
24. A. J. Levine, J. Momand, C. A. Finlay, *Nature* **351**, 453–456 (1991).
25. J. T. Zilfou, S. W. Lowe, *Cold Spring Harb. Perspect. Biol.* **1**, a001883 (2009).
26. O. Laptenko, C. Prives, *Cell Death Differ.* **13**, 951–961 (2006).
27. A. Lupas, *Trends Biochem. Sci.* **21**, 375–382 (1996).
28. P. Burkhard, J. Stetefeld, S. V. Strelkov, *Trends Cell Biol.* **11**, 82–88 (2001).
29. G. Bucher, J. Scholten, M. Klingler, *Curr. Biol.* **12**, R85–R86 (2002).
30. J. A. Lynch, C. Desplan, *Nat. Protoc.* **1**, 486–494 (2006).
31. L. W. Beukeboom, A. Kamping, *Genetics* **172**, 981–989 (2006).
32. P. W. Whiting, *Genetics* **45**, 949–970 (1960).
33. J. H. Werren *et al*., *Science* **327**, 343–348 (2010).

ACKNOWLEDGMENTS

We thank J. A. Lynch, A. H. Rensink, A. de Haan, S. de Vegt, M. Schenkel, F. Chen, K. Leung, E. Dalla Benetta, and D. Bopp for data and experimental, analytical, and intellectual help. **Funding:** Part of this work is financed by Netherlands Organization for Scientific Research (NWO Veni 863.13.014 to E.C.V and NWO TOP grant no. 854.10.001 to L.W.B) and China Scholarship Council (CSC no. 201506240202 to Y.Z.). **Author contributions:** Y.Z.: Investigation, formal analysis, writing; E.G.: Investigation, formal analysis, writing; L.W.B: Conceptualization, supervision, writing, project administration; E.C.V.: Conceptualization, investigation, formal analysis, writing; L.Z.: Conceptualization, supervision, writing, project administration. **Data availability:** The *wom* mRNA sequence is deposited in NCBI GenBank (MT063190). RNA sequencing and genome resequencing data are deposited in the Sequence Read Archive of NCBI (BioProject PRJNA606659 and PRJNA622909).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6520/1115/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S5
References (*34*–*52*)

9 April 2020; accepted 20 October 2020
10.1126/science.abb8949

THE SURGE OF HUANGPU RIVER, THE BLOOM OF SHANGHAI

Shanghai is called 'Shen' and also can be shortened to 'Hu'. It is surrounded by water on three sides, the Yangtze River to the north, the East China Sea to the east, and Hangzhou Bay to the south. The legend of Shanghai is recorded by the surging Yangtze River and the vastness of the Pacific Ocean.

Shanghai leaves behind much of the country when it comes to performance and thanks to the integration of the Yangtze River Delta, which has been accelerated, Shanghai's economy has been constantly making new records. In 1990, China's first financial and trade zone, Pudong Lujiazui Financial and Trade Zone, was established; in 2013, the first China Pilot Free Trade Zone was established in Pudong; in 2018, the first China International Import Expo was held and the world watched, amazed.

Shanghai's strength in technology and industrial system can put our country on the international map. The city's presence is everywhere, whether it is in major scientific and technological achievements such as Jiaolong, Tiangong, Beidou, FAST, Micius, and large passenger aircraft, or scientific developments in basic frontier research in China, to original innovation and independent R&D teams of key technologies such as integrated circuits and AI.

Shanghai is also renowned for its immense historical and cultural value. Books published by the Commercial Press and Zhonghua Book Company are sold across the country. Shanghai is also the birthplace of Chinese oil painting, caricature, the cosmopolitan style of Peking opera, Shaoxing opera, Shanghai opera, comics, drama, etc. Shanghai's newspaper and film industry have played an important role in China's cultural development.

Shanghai has a highly evolved education system with numerous famous schools and excellent talents. People want to attend reputable universities such as Fudan University, Tongji University, Shanghai Jiao Tong University, East China Normal University, Donghua University, East China University of Science and Technology and ShanghaiTech University, where they can grow by learning and exploring or develop their skills further through teaching or research.

Some universities in Shanghai have developed into world-class institutions. They have invested a lot of effort to strengthen the talent teams and improve their ability of independent innovation. They have learnt from the development experience of overseas higher education systems so they can make innovations within their own mechanisms, highlight their revolution, focus on their quality, and accelerate their development. Shanghai's universities precisely connect the needs of economic and social development with the Master Plan of National Development Strategy in terms of course set-up, industry-university-research cooperation, transfer and transformation of academic achievements, consulting and research of decision-making, and think-tank development.

A team of top-class scholars is essential for the development of a top-class university. Therefore, Shanghai universities actively look at growing the talent collection. With the development of domestic economy and the increase of scientific research funds, Shanghai's universities are now seeing some excellent teachers while the faculty has increasingly grown to be on par with international world class universities. However, everyone wants to step up to a better place instead of remaining in one place all the time. In the face of global competition, Shanghai's universities have provided a positive mechanism to retain excellent teachers. With exceptional talent development, accurate talent services, and perfect incentive and evaluation mechanism, Shanghai's universities are looked upon favorably by numerous high-level talents and it has led to many of them being drawn to Shanghai from all over the world in 2020.

Shanghai is a charming city. The Master Plan of Shanghai (2016-2040) paints a blueprint of Shanghai's future: "In the future, buildings will be enlightening, streets will be strolling-friendly, parks will be enjoyable, and the skyline will be breathtaking. Shanghai strives to create the image of humility, elegance, warmth, and enjoyment!" In future, more universities and disciplines in Shanghai will get into the top ranks in the world. Will you be in Shanghai then? We welcome more scholars to join our team so we can successfully develop Shanghai's higher education field!

We welcome excellent scholars interested in applying for talent programs in China to contact us through AcaBridge (consultant@acabridge.edu.cn), which provides one-on-one consultations.

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

SJTU Global Recruiting Program

Shanghai Jiao Tong University (SJTU) is one of the higher education institutions which enjoy a long history and a world-renowned reputation in China. Through 120 years' unremitting efforts, SJTU has become a comprehensive, research-oriented, and internationalized top university in China.

SJTU now has 31 schools/departments, 32 research institutions, 13 affiliated hospitals, with around 40,000 students and over 3,000 full-time faculty, including the leading number of academic masters such as academicians of the Chinese Academy of Sciences and the Chinese Academy of Engineering, candidates for overseas talent programs and winners of National Outstanding Youth Fund among institutes of higher education in China.

Today SJTU has 67 undergraduate programs, 42 first-level disciplines authorized to offer doctorate degree, 57 first-level disciplines authorized to offer master degree. According to the latest Thomson Reuter's Essential Science Indicator (ESI), SJTU has 19 disciplines listed World Top 1%, with 7 disciplines ranking World Top 1‰ and Engineering ranking World Top 1‰. In 2019 QS World University Rankings by Subject, SJTU has 25 subjects ranking World Top 100, of which 10 subjects rank World Top 50. By 2018, SJTU has led the country for the 9th consecutive year in terms of both the project number and the amount of funds supported by National Natural Science Foundation of China. SJTU also ranks top among all Chinese universities in the total number of published SCI papers.

Shanghai Jiao Tong University, carrying the mission of preserving cultural heritage, and seeking for the truth, bearing the responsibility of invigorating the Chinese nation and developing for the benefits of mankind, today this centennial university is sailing for the aim of becoming a comprehensive, research-oriented and internationalized world-class university. SJTU will provide free academic environment, strong research support and competitive compensation package for the talents.

SJTU, your stage to becoming academic master!

---

1. Recruiting Categories:

(1) Chair Professor/Distinguished Professor/Full Professor with Tenure

(2) Tenure-track Associate Professor

2. Work and Life Treatment:

(1) Salary Benefits: With reference to the corresponding positions in the world's top universities, competitive remuneration and welfare benefits will be provided;

(2) Start-up Package: Adequate research fund negotiable according to research demand;

(3) Housing: Assistance will be given in solving housing problem with furnished interim apartment provided;

(4) Child Education & Health Care: Assistance will be given in medical treatment and children's education depending on affiliated hospitals and schools.

3. How to Apply:

Application Materials:

(1) Cover letter for the position that you are applying for;

(2) CV (with publication list);

(3) Expertise and academic results;

(4) No less than 5 representative papers.

Application materials should be integrated into a PDF document named "name—position—discipline".

SJTU Talent Resources Department

Tel: 86-21-34206732, 34206724

E-mail: faculty@sjtu.edu.cn.

Job Opportunities at East China Normal University

Founded in Shanghai in October 1951, East China Normal University (ECNU) is one of China's top research universities. Directly under the Ministry of Education and sponsored by the national key university programs — "Project 211" and "Project 985" — the university is renowned for its teacher education and high-level research programs in both basic and applied sciences.

In 2017, ECNU was selected into Class A of the Double First Class Program (First Class University and First Class Academic Discipline in the World), an education initiative launched by the Ministry of Education aiming at developing elite Chinese universities into world-class institutions by the end of 2050.

Adhering to the strategy of pursuing interdisciplinary development, internationalization, and informatization, ECNU has made great achievements in talent training, scientific research, community service and international exchanges in the past few decades.

With the five action plans, i.e. Education Plus, Ecology Plus, Health Plus, Intelligence Plus and Internationalization Plus as the petals, the university is working hard to forge the Flower of Happiness, ECNU's recently formulated university development initiative. With all these efforts, ECNU intends to fulfill its vision of "Creativity, Character and Community".

Job Opportunities

ECNU invites applications for tenure-track faculty positions in multiple disciplines in all ranks from assistant professors to full professors. The positions of research scientists and postdoctoral fellowships are also open for application.

Qualifications

Applicants must possess a doctorate degree in a closely related field and show evidence of relevant teaching and research experience, including peer-reviewed publications. They must have a professional record that proves their external funding history or potential.

For more detailed information, please contact:
Contact: Mr. Wang
Office Tel: 86-21-62235058
E-mail: cwang@bio.ecnu.edu.cn
Talent Recruitment: www.jobs.ecnu.edu.cn
Personnel Department: http://hr.ecnu.edu.cn/

ECNU Recruitment Announcement

ECNU Talent Recruitment Platform

EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY

East China University of Science and technology (ECUST) is a key university directly under the administration of the Ministry of Education (MOE) of the People's Republic of China. Located in Shanghai, a center of China's economy, finance, commerce and shipping, ECUST is listed on Project 211 and 985 Project Innovation Platform. In 2017, ECUST was chosen to participate in the Double First-Class Plan. Based on Essential Science Indicators (ESI) database, Chemistry, Material Science, Engineering, Biology and Biochemistry, Pharmacology and Toxicology, Agricultural Science, Computer Science, Environmental Ecology in ECUST rank top 1% globally, Chemistry being among top 1‰. Chemical Engineering and Technics was assessed as No.1 in MOE's 4th Discipline Assessment. In the newly released 2020 QS University Rankings, ECUST was 28th among China mainland universities. According to Times Higher Education (THE) World University Rankings, ECUST situated at 28th among China mainland universities. In 2020 US News University Rankings, ECUST was at 41st among China mainland universities. And based on 2020 ESI Chinese University Rankings, ECUST was at 28th.

POSITIONS AND COMPENSATION PACKAGES

1.Recruitment Program for Outstanding Professionals

Applicants engaged in scientific researches who possess a PHD degree granted by prestigious overseas universities, with formal teaching and researching positions in overseas prestigious universities, institutions or enterprises for consecutive 36 or more months, will be able to work full time in China. As the elites among the peers in certain field, the outstanding professionals shall have the potential of becoming leading figures in the future. Outstanding PHD students can be recruited in exceptional cases. Applicants have no full domestic employment or less than a year upon application.

- Annual salary: 500,000 RMB or more (before tax).
- National research fund:1,000,000-3,000,000 RMB, a 3,000,000 RMB counterpart fund will be provided by ECUST.
- Housing benefits of 2.5 million RMB (including allowance from the National and Shanghai Municipal Governments). Fully furnished Teachers Apartment available at preferential rent.
- Special quotas for recruiting doctoral students provided.
- Help with job arrangements for the spouse as well as talents' children's entrance into kindergarten, primary school and secondary school.
- To those failed the final Recruitment Program for Young Professionals defense, ECUST offers associate Distinguished Research Fellow or above.

2.ECUST Distinguished Research Fellow/ Associate Distinguished Research Fellow

(1) Distinguished Research Fellow

Applicants should have high morality in science, dedicated to rigorous research. Applicants should possess a PhD degree granted by prestigious overseas universities, with formal teaching and researching positions in overseas prestigious universities, institutions or enterprises for 2 or more years, who will be able to work full time in China. As the elites among the peers in certain field, the young professionals shall have the potential of becoming leading figures in the future. Applicants have acquired important achievements equal or close to professionals selected in The Recruitment Program for Young Professionals or average top youth talents, have definite research direction and features, and display strong academic potentials.

- Annual salary: 350,000 RMB or more (before tax).
- Research fund:1,000,000 RMB for majors of science and engineering, 500,000 RMB for majors of arts.
- Special quotas for recruiting doctoral students provided.

(2) Associate Distinguished Research Fellow

Applicants should have high morality in science, dedicated to rigorous research. Applicants should possess a PHD degree granted by prestigious overseas universities,
have rich research background, have acquired important achievements equal or close to the qualifications for associate professors or research fellows in ECUST, and display quite strong academic potentials. Applicants should be able to work fully in ECUST after being employed.

- Annual salary: 250,000 RMB or more (before tax).
- Research fund: 500,000 RMB for majors of science and engineering, 200,000 RMB for majors of arts.
- Special quotas for recruiting master students provided.

3.Professor/Associate Professors/Excellent Scholars

Professionals in the fields of sciences, engineering and humanities at home and abroad, have acquired important achievements and displayed quite strong academic potentials. Applicants should be able to work fully in ECUST after being employed.
Competitive payment and development opportunities provided.

4.Postdocs

- Annual salary: 210,000 RMB or more (before tax) for majors of sciences and engineering, 180,000 RMB or more (before tax) for majors of humanities.
- Residence registration will be processed in line with post-doctor management regulations.
- Support will be given to working postdocs applying for national and Shanghai talents programs, China National Funds for natural sciences as well as postdoctoral foundations.
- Excellent post-doctors can apply for tenure tracks after their postdoctoral work.

Email: rsc@ecust.edu.cn
Web: http://personnel.ecust.edu.cn/

DONGHUA UNIVERSITY

Welcome Distinguished Scholars from Home and Abroad

Donghua University (DHU) is one of the National Key Universities directly under the Ministry of Education (MoE) of China and enlisted in China's "Project 211" and "Double First-Class University Plan". Located in Shanghai, DHU has three campuses: Songjiang Campus, Yan'an Road Campus and Xinhua Road Campus, covering an area of nearly 329 acres with an overall floorage of over 780,000 m^2. Over the past 70 years of development, DHU has evolved from a single-disciplinary textile institute into a multi-disciplinary national-key university, with engineering at its core supplemented by science, management, humanities. DHU strives to carry out talent cultivation, scientific research, social service, cultural preservation and innovation, and international cooperation and exchanges. It provides expertise and scientific and technological support for the development of Shanghai and China.

Faculty

DHU is committed to academic excellence and the recruitment and development of its academic faculty. Of the total number of more than 2,200 faculty and staff, over 1,300 are faculty members, among which over 900 are holders of senior professional/titles, forming a faculty team including academicians, experts selected under winners of National Science Fund for Distinguished Scholars, winners of National Fund for Outstanding Youth Talents, Chief Scientist of the National 973 Programme, winners of National Key Research and Development program, and other high-level talents as well as outstanding young and middle-aged academic elites.

Education

DHU has contributed to the process of comprehensive reform by developing staff as high-quality talents with innovative spirit and sense of social responsibility. It continues to deepen the reform of education and teaching, innovating in its cultivation of talent and demonstrating a commitment to quality assurance and enhancement at undergraduate and postgraduate levels. At present, nearly 30,000 students are enrolled at DHU, including over 14,000 undergraduates, nearly 7,000 postgraduates, over 2,000 students pursuing adult education and more than 4,800 international students.

Research

DHU continues to build a comprehensive discipline with textile as the "body", materials science and design as "two wings" and innovation as the "engine". Currently, DHU offers 6 postdoctoral research programs, 10 doctoral programs in first-level disciplines, 2 doctorate degree conferment cat-

egories, 28 Master's programs in first-level disciplines, 17 professional Master's programs, and 55 undergraduate programs across 9 disciplines including engineering, science, management and so on. DHU has won 28 National Natural Science Awards, National Invention Awards and National Science and Technology Progress Awards. It has also led 8 national key research and development programs in the past two years.

Social Responsibility

In recent years, more than 5,000 industry-university-research cooperation programs have been established and a large number of scientific research achievements have been widely applied in areas such as aerospace, military, architecture, environmental protection and new materials, making positive contributions to the successful launches of Shenzhou V to Shenzhou XI manned spacecraft and of Tiangong-1 and Tiangong-2 space stations. In addition, DHU works with local government to propel the construction of "Greater Donghua Fashion and Creative Industries Complex" and "Greater Songjiang University Town Innovation Park". Furthermore, DHU is promoting the development of the alumni association, foundation and governing board to enhance external relationships.

Colleges (Departments, Centers)

- College of Textile
- College of Fashion and Design
- Glorious Sun School of Business and Management
- College of Mechanical Engineering
- School of Information Sciences and Technology
- School of Computer Sciences and Technology
- College of Materials Science and Engineering
- College of Environmental Science and Engineering
- College of Humanities
- College of Marxism Studies
- College of Science
- College of Foreign Languages
- Department of Physical Education
- Continuing Education College
- International Cultural Exchange School
- Shanghai International College of Fashion and Innovation
- Innovation Center for Textile Science and Technology
- Center for Civil Aviation Composites
- Center for Advanced Low-dimension Materials
- Shanghai International Fashion Innovation Center

Shanghai Universities First-Class Disciplines Project

- Textile Science and Engineering
- Materials Science and Engineering
- Design
- Mechanical Engineering
- Control Science and Engineering
- Environmental Science and Engineering
- Chemistry

Recruitment Positions

Donghua University Distinguished Research Fellow

1. Under the age of 35 for researchers in natural science and engineering science, or under the age of 40 for researchers in humanities and social sciences.
2. Applicant should get PhD degree and have post-doctor experience or obtained assistant professorship or above in prestigious overseas universities; or professors in domestic high-level universities or institutions.

Donghua University Shangshi Forum for International Young Scholars will be held online on **25th Dec. 2020**. Welcome young scholars to exchange latest research trends. More information please contact us by email **rcb@dhu.edu.cn** or by telephone **(+86) 021-67792042**

OPPORTUNITIES TO SHINE AT SHANGHAITECH UNIVERSITY

ShanghaiTech is a young resource-rich university with a modern residential campus in the heart of Shanghai Pudong's Zhangjiang Hi-Tech Park. With an academic focus on STEM research, ShanghaiTech is committed to serving China's national challenges in economic and social development with the focus on science and technology, and nurturing the next generation of innovative scientists, inventors and entrepreneurs. With the backing and support of Shanghai Municipal Government and Chinese Academy of Sciences, ShanghaiTech's six schools and four research institutes are committed to tackling the challenges that China and the world are facing in the fields of energy, materials, biology, environment, human health, information technology and mathematics.

For more information, please visit: www.shanghaitech.edu.cn.

SHANGHAITECH IS NOW SEEKING TALENTS IN THE FOLLOWING FIELDS

School of Physical Science and Technology: energy, system materials, photon and condensed state, material biology, environmental science and engineering

School of Life Science and Technology: molecular and cell biology, structural biology, neuroscience, immunology, stem cells and regenerative medicine, system biology and biological data, molecular imaging, biomedical engineering

School of Information Science and Technology: computer science, electrical engineering, information engineering, artificial intelligence, network and communication, virtual reality, statistics, big data and data mining

School of Entrepreneurship and Management: economics, finance, accounting, management, marketing, strategy and entrepreneurship

School of Creativity and Art: intelligent design, entertainment design

Shanghai Institute for Advanced Immunochemical Studies: antibody therapy, Immunotherapy, cell therapy, regeneration medicine

iHuman Institute: bio-imaging, biology, chemistry, computational biology, AI/ML

Institute of Mathematical Sciences: pure mathematics, theory of computing, applied mathematics

Institute of Humanities: Chinese Philosophy, Western Philosophy, Logic, Science Philosophy, Aesthetics, Ancient literature, modern literature, literary theory, comparative literature and world literature, Chinese writing, Chinese history, World history, Historical theory, Art history, English language teaching & English literature, French, German, British and American literature, French literature, German literature Religious Theory, Cultural Anthropology, Political Theory, the Science of Law, Social Theory

FOLLOWING POSITIONS ARE OPENING

Tenured and Tenure-track positions: assistant professor, associate professor and full professor. Successful applicants will have a doctoral degree, and are expected to establish a record for independent, internationally recognized research, supervise students and teach high-quality courses.

Research positions: post-doctoral research fellow, research assistant professor, research associate professor and research professor. Successful applicants will have a doctoral degree, a good research record and great passion for research.

Assistant positions: teaching assistant, research assistant, and administrative officer. Successful applicants will have a Master's degree and relevant working experience.

SHANGHAITECH WILL OFFER ATTRACTIVE COMPENSATION PACKAGES, INCLUDING

Initial research support package: reasonable start-up funds, research associates and post-doctoral fellows, laboratory space to meet research needs.

Compensation and benefits: highly competitive salary commensurate with experience and academic accomplishments, a comprehensive benefit package.

Subsidized housing: on-campus 80/100/120 m^2 faculty apartments available at low rent for tenured and tenure-track faculty, on-campus postdoctoral dormitories, off-campus postdoctoral apartments and municipal apartments subsidized by Shanghai government.

Relocation & travel allowance: reimbursement of expenses for household relocation and family's one-way travel.

Family assistance: support with children's education; affiliated kindergarten, primary and middle schools.

To apply: using this format, please submit
1.Cover letter (Firstname_Lastname_Cover_Letter.pdf)
2.Research plan (Firstname_Lastname_Research_Plan.pdf)
3.CV (Firstname_Lastname_CV.pdf)
To **talents@shanghaitech.edu.cn**

Who's the top employer for 2020?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles at sciencecareers.org/topemployers

Science 2020
TOP EMPLOYER

ASSISTANT/ASSOCIATE/FULL PROFESSOR
Faculty Positions in Biochemistry & Molecular Biology

The Department of Biochemistry & Molecular Biology at the University of Oklahoma Health Sciences Center (OUHSC) invites exceptional scientists to apply for tenured/tenure-track eligible **faculty positions**. Applications from individuals with expertise in biochemistry and molecular biology are requested and areas of special interest include bioenergetics, metabolism, adult stem cells/tissue regeneration, cell signaling and mechanisms of neurodegeneration. The start date is negotiable beginning **Spring 2021**.

Minimum Qualifications and Requirements: Successful candidates are expected to develop innovative and externally funded research programs, participate in departmental education programs, and contribute to service opportunities. Candidates should have a PhD and/or MD with a strong record of research accomplishments. Candidates at the **Assistant Professor** level are expected to have significant postdoctoral research experience with the potential of developing an extramurally funded, nationally recognized research program, while candidates at the **Associate/Full Professor** level are expected to have an established nationally/internationally recognized research program.

The Department of Biochemistry & Molecular Biology is one of five basic science departments within the College of Medicine at OUHSC, providing a highly interactive group of scientists with interests in geroscience, mitochondrial function, metabolism and protein structure. The department is home to the Oklahoma Nathan Shock Center on Aging, a Geroscience T32 Training Program and a Cellular and Molecular Geroscience CoBRE grant and receives over $9M in extramural funding annually.

Successful applicants will receive a competitive salary commensurate with rank and experience, a generous, multi-year start-up package, and be assigned newly renovated laboratory space. For further information about the department go to **https://basicsciences.ouhsc.edu/bmb/**.

Candidates should submit a curriculum vitae, a description of current research interests, and future research plans as one pdf file. They should also include contact information for at least three references. Please send the application packet to **Melissa-K-Day@ouhsc.edu**. Confidential inquiries may be addressed to Dr. Ann-Louise Olson, Chair Biochemistry Search Committee (**Ann-Olson@ouhsc.edu**).

The University has a commitment to diversity, equality and inclusion. The University of Oklahoma Health Sciences Center is an equal employment opportunity and affirmative action institution **http://www.ou.edu/eoo**! *Individuals with disabilities and protected veterans are encouraged to apply.*

ASSISTANT/ASSOCIATE PROFESSOR OF CELLULAR BIOLOGY AND ANATOMY

The Department of Cellular Biology and Anatomy at the LSU Health Sciences Center School of Medicine in Shreveport is restarting a faculty search seeking a candidate holding a Ph.D. or equivalent degree for a tenure-track position at the level of **Assistant** or **Associate Professor**. To be considered for the position, the candidate should have expertise in the areas of Cellular Biology and the Anatomical Sciences. Candidates having an established research direction/expertise in cardiovascular diseases and related areas are particularly encouraged to apply.

This position is part of a continued expansion of the research program in the department. The successful candidate will participate in team-based teaching for Gross Anatomy in our Medical, Graduate, and Allied Health Curricula. Additional experience in teaching Cellular Biology, Neuroanatomy, or Histology is a plus. The requirements for **Assistant Professor** include a minimum of two years of postdoctoral research experience. For the **Associate Professor** rank a minimum of five years of experience at the level of Assistant Professor. The successful candidate is also expected to have a record of scholarly and research activity during their career. For the Associate Professor rank, extramural funding (NIH R01 or equivalent) is expected. Salary and laboratory start-up packages for individuals applying for this position will be commensurate with experience.

This is an exciting opportunity for the appropriate candidate to play a significant role in the future growth of the Department and its research programs. In addition, there are numerous opportunities for the development of collaborative research projects with faculty members in our multidisciplinary Center for Cardiovascular Disease and Sciences here at LSUHSC.

Please submit application with full curriculum vitae and names of three references via email to Ms. Jennifer Harper (**jharpe@lsuhsc.edu**), Departmental Coordinator, Department of Cellular Biology and Anatomy, LSU Health Sciences Center, School of Medicine in Shreveport, Shreveport, LA 71130. Please include with your packet a letter that summarizes your future research plans. Review of applications will begin immediately and will continue until the position is filled.

LSUHSC – Shreveport is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, disability status, protected veteran status, or any other characteristic protected by law.

By **Emily Fogarty**

Credit where credit is due

"While I recognise that she thinks she had a major contribution to the project it will be difficult to show that she did more than a technician's job." This was the infuriating response I received in a long email thread with former colleagues about work I had done as an undergraduate researcher. During my 7-month internship, my colleagues in the lab told me I would be an author when they wrote up the work. A few months ago, I was shocked to learn that the resulting paper had already been accepted for publication—and that my contribution had been relegated to the acknowledgments section. This was my wake-up call about the need to speak up for myself regarding authorship, and to speak out against the unfair convention of diminishing the contributions of undergrads and technicians to scientific research.

As an undergraduate, I was just excited to be involved in science and making discoveries. I was oblivious to the true currency of science: publications. I dived into my assigned project to develop and optimize a cell culturing system, then collected as much preliminary data as I could. In a conversation with the principal investigator and postdoc I worked with, they mentioned that I would be the first or second author when they published the work, but we didn't get into specifics.

After leaving the lab, I was in contact with the postdoc and the graduate student who took over the project, eager to be a co-author on my first paper. Then, I got some bad news: The cell lines I had been working with were contaminated. The graduate student had repeated my experiments—based on the protocol I had written, methods I had taught him, and observations and interpretations I had shared—and he was able to replicate the phenomena I had observed. He assured me that my contributions were still recognized and that I would be third author on the eventual paper.

I went on to start graduate school myself. I occasionally checked in with the graduate student and postdoc about the project and offered to help write the manuscript. When the student told me that he was postponing submission for personal reasons, I decided to back off, expecting that they would reach out when they needed my input.

Eventually, after not hearing anything from them for more than a year, I emailed to check in. That's when I learned that the paper had already been written, submitted, and accepted for publication—and I had been excluded from the process and the author list. Instead, I was acknowledged "for preliminary studies and technical expertise." When I insisted that I should be an author, as we had discussed, they told me I had "just" been a technician.

"When I insisted that I should be an author ... they told me I had 'just' been a technician."

Most journals' policies state that authorship should be based on making a significant contribution to the conception or design of the work, or the acquisition, analysis, or interpretation of the data. Authors may also be required to help write the manuscript, but guidelines from the International Committee of Medical Journal Editors—which the journal that published this paper adheres to—state that this requirement is not meant to disqualify individuals who otherwise meet the criteria. Based on that definition, my contributions *did* merit authorship. I was hurt to see my work being trivialized—in large part, I believe, because I had done it as an undergraduate technician.

I now work in a different discipline, so I hope I can share my story without jeopardizing my future career. In retrospect, I should have initiated explicit conversations about authorship from the beginning, with all members of the project. Before I left the lab, I should have written up my results and methods in a format that could readily be included in the eventual paper.

But the community needs to address a broader problem: the convention of excluding undergraduate students and technicians from authorship lists. A researcher's title doesn't make their contribution any less significant. Junior researchers are often critical to a project's success, and we need to make sure that everyone who contributes gets the credit they deserve. ■

Emily Fogarty is a Ph.D. student at the University of Chicago.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER