Science

COVID-19

Building on what we've learned p. 1098

TEST. FAIL. LEARN. REPEAT.

Neuroinflammatory response in Alzheimer's disease (AD) results in altered morphology of microglia and astrocytes surrounding amyloid-beta plaques. This image shows brain from a mouse model of AD labeled with Iba1/AIF-1 antibody #36618 (yellow) and several other targets.

OUR PROCESS — YOUR SUCCESS.

As a privately held company founded and led by active research scientists, we understand your needs as a researcher. Like you, we only want to use products that are specific and deliver reproducible results. We validate every one of our thousands of antibodies and research products in-house. If any product fails to meet our stringent standards, we won't sell it. That's why you can trust CST to help you achieve experimental success.

www.cellsignal.com

Cell Signaling
TECHNOLOGY®

CONTENTS

ON THE COVER
On 11 March 2020, the World Health Organization officially characterized COVID-19 as a pandemic, triggered by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Science* marks this anniversary with a special issue that highlights the international scientific community's extraordinary achievements in response to the pandemic—and, as SARS-CoV-2 becomes endemic, outlines ways to bridge the gaps in our understanding of this changing virus and our responses to it. See page 1098. *Illustration: Stephan Schmitz/Folio Art*

NEWS

INSIGHTS

CONTENTS

1169

The water content of lava from volcanoes like Cleveland in Alaska helps constrain magma storage depth.

RESEARCH

1194

"Back to normal" is not enough

As the world reflects on 2 years of the COVID-19 pandemic, we need to change how to tackle the enormous challenges of the future. The good news is that the past 2 years of the COVID-19 pandemic have shown that change is possible.

Mitigating the impact of COVID-19 forced rapid change everywhere. Countries built new testing and contact-tracing infrastructure. Many instituted lengthy stay-at-home orders. Major institutions moved operations online. Entire health systems were rapidly redesigned. Millions learned to work and learn from home. Meanwhile, vaccines were produced on record time scales, and countries undertook the largest mass vaccination programs of the modern era. Governments passed stimulus and relief packages to provide unprecedented financial support for workers and businesses. Necessity was indeed the mother of invention. Some problems, however, remain unsolved, such as ensuring global vaccine equity or reducing health inequalities.

The past 2 years should prompt everyone to revisit the long-standing problems considered too difficult to solve because traditional practice is too sticky, or because people "will never go for it," or because it's just too expensive, or it will cost the next election. We need to ask: Are these problems really too difficult? Or do we just not consider solving them necessary? Who decides what is necessary? Government is about balancing trade-offs, and not all problems will be solved—but honesty is needed about which solutions are deemed necessary and why.

Then there are the problems so ingrained that they are not even recognized as solvable. Consider annual hospitalizations and deaths from influenza and other pneumonias. These have been treated as a known winter burden of ill health, although many public health experts argue that surges in winter deaths are not inevitable and are mainly due to poverty and health inequalities. So accepted, however, is the annual burden from influenza and pneumonias, that their average annual numbers of hospitalization and deaths have been used as a benchmark for COVID-19—a signal for when "we can go back to normal."

But measures taken against one infectious disease can be effective against others, as proven with COVID-19 mitigations. Influenza almost disappeared from the world in the first year of the pandemic. This winter, many were expecting a comeback, but in England, for instance, some pandemic measures remained, such as mask wearing, remote working, and a reduction in social contacts. Since October 2021, England has had just over 2000 influenza admissions—10% of equivalent periods in the 2 years prepandemic. The annual winter burden of respiratory diseases is clearly not inevitable. Rather, the question is how far society wants to go to reduce them. Instead of calling for a return to normal, we should be asking if normal can't be better.

For many countries, it was mostly the threat of COVID-19 overwhelming health care systems, and the thousands of potential deaths, that created support for large-scale change. But the time scales of infection and the nature of exponential growth are such that hospitalizations and deaths only reached the threshold that changed perception from "it's fine" to "we must do something" weeks after transmission was already out of control. This doomed many countries to acting late and needing to impose harsher measures for longer periods—and not just once but repeatedly.

Humans have evolved to perceive urgency on the order of hours and days, not years or decades. This tendency already delayed action on COVID-19, which works on a time scale of weeks. We can see it playing out now in the unparalleled and rapid global unified response to the invasion of Ukraine. A sense of immediacy combined with necessity has enabled a response there unimaginable even a few weeks ago. But the biggest global crisis—the climate emergency—struggles to provoke such a response because catastrophes are fully experienced only decades after they became inevitable. The Intergovernmental Panel on Climate Change issued a stark warning last month: "Any further delay in concerted global action will miss a brief and rapidly closing window to secure a liveable future." The climate emergency requires solutions far greater in their scope than those for COVID-19, and it requires them now. It is imperative that the world uses science to inject that sense of immediacy for tackling the greatest challenges. The last 2 years have shown that together we can do great things, if we decide they are necessary. The old normal is no longer enough.

—Christina Pagel

> "Instead of calling for a return to normal, we should be asking if normal can't be better."
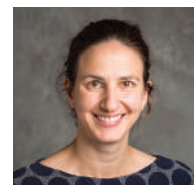
Christina Pagel
is a professor of Operational Research and director of the Clinical Operational Research Unit, Department of Mathematics, University College London, London, UK. c.pagel@ucl.ac.uk; @chrischirp

Scientists in the line of fire

The devastation and despair gripping Ukraine following the unprovoked invasion by neighboring Russia is heartbreaking and unthinkable. Such a loss of life and homeland has stirred wide concern around the world. This war sets back progress to establish a peaceful and sustainable world and to address important problems faced by all humanity, including climate change, environmental degradation, public health, and inequality. The international community of scientists cooperates extensively to address the challenges of our time, and a war that is destroying a stable and healthy nation and provoking a refugee crisis is no exception. What can the scientific community do most immediately to provide support and aid to its Ukrainian colleagues in their time of need? The community should focus on strengthening regional partnerships in Eastern Europe, networking to find refugees safe havens, speaking out forcefully against this invasion, and preparing to help rebuild Ukrainian science when the time is right.

As a first step, scientists and scientific organizations around the world, including the United States, need to redouble efforts to fortify links with scientific communities in countries that border the conflict area. The scientific communities in these young democracies, including Poland, Romania, and Bulgaria, often have strong connections to colleagues in Ukraine. Such ties will be critical in efforts to strengthen civil society in Ukraine and in other parts of the region. These nations will bear the brunt of the refugee crisis. Connections can be strengthened through international scientific bodies such as the InterAcademy Partnership, direct outreach between leaders, and the development of joint programs to support refugees, both financially and professionally.

Support is also needed for those colleagues in Ukraine's neighboring nations who are providing, often through heroic efforts, safe havens for refugees. Many Ukrainian scientists are women who have left with their families while their husbands stay to fight for their country. Many of the families of male scientists are also in need of a welcoming home outside of Ukraine, at least for now. Efforts to highlight these refugees and their plight must ensure that communications will not lead to unintentional harm by identifying individuals in electronic communications or social-media posts.

> **"…expressions of solidarity surely lift the spirits of our Ukrainian colleagues…but words are not enough…"**

The experience of the United States National Academy of Sciences (NAS) in providing new temporary homes for Afghan scientists showed that the families crave an opportunity to establish some semblance of normalcy and dignity: a safe home, schools, and a distraction from the stress of conflict for the children. For these scientists, an opportunity to remain connected with the research community through a provisional affiliation with a research lab or university is invaluable. The US NAS has used its own discretionary funds to support travel and short-term settlement costs in welcoming nations for Afghan scientists and their families, whose prior sponsorship by US government programs put them in additional jeopardy. The US NAS intends to extend this policy to scientists in conflict zones whose US ties place them at increased risk.

Many international and national science academies and societies have issued strong statements condemning the invasion and expressing support for the scientists (and all citizens) affected. Such expressions of solidarity surely lift the spirits of our Ukrainian colleagues and encourage them in their struggles, but words are not enough. The international scientific community will need to help rebuild science and research infrastructure in Ukraine when the time is right. It is not too early to begin these discussions.

As the world seeks to support scientists from Ukraine, it must also be careful not to condemn indiscriminately Russian scientists by assuming that all of them support this conflict. Many have, at great personal peril, spoken out against the invasion. At the same time, the global scientific community must be cautious in determining which international scientific activities can continue to involve Russian science so as not to support indirectly the Russian government, economy, and military through scientific exchange and recognition.

Leaders and members of science academies in Europe and elsewhere, including the US NAS, must work together now to accommodate Ukraine's scientists and their families until it is safe for them to return to their homes. Ukraine will urgently need the talents of these experts as they seek to rebuild their country following this devastating war. Sadly, this is not the first time that scientists have been caught in the line of fire, and it will not be the last.

–Marcia McNutt and John Hildebrand

Marcia McNutt
is president of the United States National Academy of Sciences, Washington, DC, USA. mmcnutt@nas.edu

John Hildebrand
is the international secretary of the United States National Academy of Sciences, Washington DC, USA, and the Regents Professor of Neuroscience at the College of Science, University of Arizona, Tucson, AZ, USA. jhildebr@arizona.edu

NEWS

IN BRIEF

Edited by
Jeffrey Brainard

The Recovery trial has analyzed data from U.K. COVID-19 wards, such as this one in Airdrie, to identify effective therapies.

THERAPEUTICS

Big COVID-19 trial notches another effective drug

The world's largest study of COVID-19 treatments has yielded another drug that can reduce mortality. On 3 March, researchers with the United Kingdom's Recovery trial announced that baricitinib, an oral drug that dampens an overactive immune system and is commonly used by people with rheumatoid arthritis, reduced hospitalized COVID-19 patients' risk of dying by 13%. Baricitinib inhibits enzymes in the Janus kinase family, which play an important role in regulating immune responses. Several smaller randomized trials had concluded it helped against COVID-19, but this arm of the Recovery study, which enrolled more than 8000 people, is by far the largest. It also showed the drug benefits patients who take other drugs known to cut COVID-19 deaths, such as the steroid dexamethasone. Baricitinib, sold under the brand name Olumiant, comes in generic versions that low- and middle-income countries can afford.

Cancer trials need old people

CLINICAL RESEARCH | Older adults make up more than half of cancer patients but only about 25% of participants in clinical studies, and last week the U.S. Food and Drug Administration (FDA) urged drug companies to tackle the problem. It recommended that companies boost representation of people over 65 years old in their trials, including in the earliest phase studies. Those studies could help refine drug doses in older patients to guide larger, later trials. Many older people don't qualify for trials or find them difficult to access. But leaving them out can lead to surprises, including serious side effects, when they take cancer drugs tested primarily in younger, otherwise healthier patients. Oncologists hailed the recommendations, although they aren't binding.

U.N. OKs push on plastics treaty

ENVIRONMENT | International negotiators are aiming to draft a global treaty within 2 years to control plastic pollution, after

the United Nations Environment Assembly resolved last week to tackle the growing problem. Already, an estimated 11 million tons of plastic waste reach the ocean each year. Although several treaties deal indirectly with plastic pollution, such as marine litter, no agreement covers the full scope of the problem, for example by reducing the manufacture of new plastic. The assembly approved a resolution calling for nations to set their own legally binding waste reduction targets and create a new global scientific advisory body. The body also agreed to create a new science policy panel, akin to the Intergovernmental Panel on Climate Change, to advise nations on emerging problems in chemical pollution and research trends.

USGS, ARPA-E heads nominated

LEADERSHIP | President Joe Biden this week moved to fill two long-vacant research spots in his administration by nominating geologist David Applegate to lead the U.S. Geological Survey (USGS) and Massachusetts Institute of Technology (MIT) professor Evelyn Wang to head the Advanced Research Projects Agency-Energy (ARPA-E) within the Department of Energy. Applegate has focused on geological hazard response and planning since joining USGS in 2004 and has been its acting head since the start of the Biden administration. Wang, who studies thermophotovoltaics, joined the MIT faculty in 2007 less than a year after earning her Ph.D. and since 2018 has led its mechanical engineering department. Biden's emphasis on combatting climate change through resilience, mitigation, and sustainable energy technologies has raised the profile of both USGS and ARPA-E. The nominations require approval by the U.S. Senate.

Test, treatment plan knocked

COVID-19 | The U.S. government this week began to ship antiviral pills for a program of free testing and immediate, free drug treatment for those with COVID-19 at hundreds of pharmacy clinics and long-term care facilities. The "test to treat" plan—one of several new proposals announced last week by President Joe Biden's administration to manage the pandemic as U.S. caseloads drop—is aimed at helping people who lack ready access to a physician. The antiviral pills being dispensed work best if given quickly after diagnosis. Many public health officials and scientists applauded the proposal. But the American Medical Association complained that the COVID-19

Death disparity underscores power of vaccines

Hong Kong is paying a heavy price for the vaccine hesitancy among its older residents, as a comparison with New Zealand underscores. Both countries minimized COVID-19 cases until the Omicron variant spread in February, but are now enduring big surges. Although New Zealand's deaths from COVID-19 have remained flat—at 65 for the whole pandemic as of 7 March—Hong Kong's have skyrocketed, with the weekly average nearing 200 per day, among the world's highest recorded fatality rates. The key difference seems to be vaccination: New Zealand authorities say 100% of residents 75 and older are fully vaccinated (and 96% of those ages 12 and up), but in Hong Kong only about 30% of those 80 and older have received two doses of a COVID-19 vaccine. (The figure is 70% for all adults and children ages 3 and up.)

A stark contrast
Extensive vaccination in New Zealand and lagging rates in Hong Kong have led to a sizable difference in death rates.

New COVID-19 cases per million people
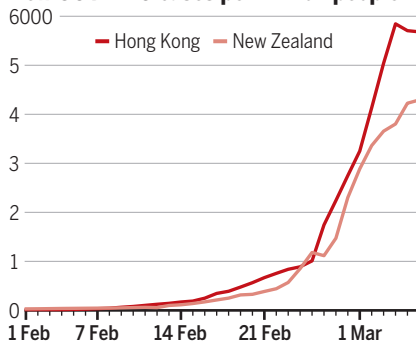
New COVID-19 deaths per million people
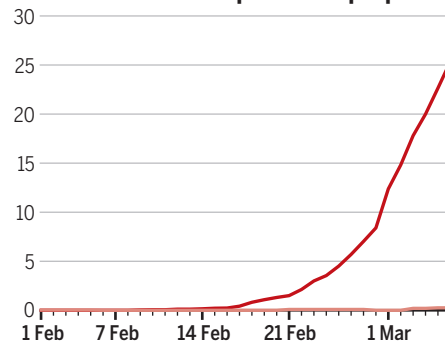
drugs could interact harmfully with drugs for other conditions and that physicians, who don't typically staff pharmacies, are best equipped to prescribe them safely.

Extinct beasts re-created digitally

VISUALIZATION | A team has created the first scientifically accurate, animated digital models of extinct mammoths, ground sloths, and other ice age animals excavated from the La Brea Tar Pits in Los Angeles. An artist and video game developer based the "paleoart" on photos of mounted
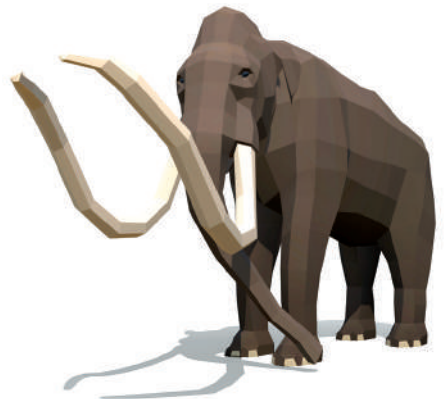
Besides the Columbian mammoth, new virtual models include the dire wolf and saber-toothed cat.

skeletons and videos of behavior in related modern-day species. Visitors to the tar pits can use a smartphone app or wear an augmented reality viewer to watch the animated creatures moving around the grounds. Simpler versions of the models, listed at https://scim.ag/3sUHnCw, can also be viewed on Snapchat.

Split decision on ResearchGate

LEGAL AFFAIRS | Two publishing giants last week said they will appeal a ruling by a German court in their copyright infringement lawsuit against ResearchGate, a social media platform. Many scientists have posted paywalled journal articles there that users can read for free. In 2017, the American Chemical Society (ACS) and Elsevier sued over 50 such papers, a sliver of what they allege are 4 million from all publishers. A regional court in Munich ruled that ResearchGate hosted them illegally but that ACS and Elsevier are not entitled to damages because not all authors on those papers had assigned them ownership. A similar lawsuit continues in a U.S. court. Two other large publishers—Springer Nature and Wiley—allow ResearchGate users to access their copyrighted material through universities' online portals.

In February, before the invasion of Ukraine, Russian President Vladimir Putin (left) met with German Chancellor Olaf Scholz. Germany has led sanctions that isolate Russia.

SCIENCE DIPLOMACY

Science ties to Russia cut after Ukraine invasion

Institutions struggle to remain neutral as Western nations take hardline stances

By **Richard Stone**

I n 2011, Russia signed a deal that would pay the Massachusetts Institute of Technology (MIT) $300 million to help found the Skolkovo Institute of Science and Technology (Skoltech), an English-language research university on the outskirts of Moscow. Dozens of researchers from around the world leaped at the chance for a foreign assignment, which they thought would help mend lingering rifts from the Soviet era, while boosting innovation in a place aspiring to be Russia's Silicon Valley. "It was a very exciting time," says Ed Seidel, a computer scientist and current president of the University of Wyoming, who was tasked with building up Skoltech's research capacity.

Citing "the Russian government's violent invasion of a peaceful neighbor," MIT on 25 February dissolved its partnership with Skoltech, 1 day after President Vladimir Putin began a bloody war with Ukraine. The decision had a personal echo for MIT President L. Rafael Reif, who in a statement says his parents fled the western Ukraine-Moldova region on the eve of World War II. MIT's withdrawal only affects a handful

of faculty and students, but it will still be an enormous loss of prestige, says Skoltech Provost Keith Stevenson, an electrochemist who moved to Russia from the University of Texas, Austin, to establish a center for energy storage technology. "It's kind of like getting a divorce from someone you love," says Stevenson, who plans to stay in Russia. "We'll have to think creatively how we can maintain morale."

A rising chorus is calling on the West to cut other ties with Russian science. The European Commission suspended Russia's participation in its flagship research program, Horizon Europe, and the national research councils of several European nations, including France, Germany, Italy, and the Netherlands, froze collaborations with Russia. "While there is a war and people are fighting, the only decision is to stop everything," says Augusto Marcelli, a physicist at Italy's National Institute of Nuclear Physics and adviser to Italy's Ministry of Foreign Affairs.

Science diplomacy veterans who spent careers forging ties with Russia agree. "It's important to be consistent in our policy toward Russia, which is to isolate and punish them. If there were high-level scientists in

a position to influence decision-making, I would be more comfortable in making an argument to keep ties open," says Cathleen Campbell, former president of CRDF Global, a nonprofit that specializes in nuclear and bioweapons nonproliferation work in the former Soviet Union. "Why should we treat scientific exchanges any differently than Champions League soccer matches, ballet performances, financial transactions, and investment projects—which have all been canceled in recent days?" asks Alfred Watkins, chairman of the Global Solutions Summit and a former World Bank official who led science capacity-building projects in Russia and Ukraine.

Other science organizations have resisted being drawn into what they see as a political minefield. Last week, for example, the International Astronomical Union rejected a petition from Ukrainian astronomers to ban Russian astronomers from IAU activities. "That would definitely be making a political statement, which the IAU cannot do," IAU President Debra Elmegreen wrote in a 1 March email to Yaroslav Yatskiv, president of the Ukrainian Astronomy Association. "The IAU was founded right after WWI [World War I] in order to bring

colleagues together, so we do not wish to drive them apart by deciding whom to support based on what their governments are doing." And the experimental ITER fusion reactor in France has no plans at present to expel Russia, which is a full member of one of the world's biggest science collaborations. "ITER is a child of the Cold War and is deliberately nonaligned," says ITER spokesperson Laban Coblentz.

Universities UK, which represents the nation's vice chancellors, advised its members to review collaborations with Russia on a case-by-case basis. "We do not support a blanket boycott," it said in a statement. Although Germany is taking a hardline stance, cutting off funding for supported researchers in Russia, Peter-André Alt, president of the German Rectors' Conference, encouraged scientists to keep informal channels open. He notes that many scientists in Russia have spoken out against the war—at least until last week, when Russia passed a law that threatens jail time for anyone departing from Russia's characterization of the assault on Ukraine as a "special military operation." "We want to support these colleagues," Alt says.

CERN, the world's largest particle physics laboratory, in Switzerland, has long prided itself as an East-West crossroads. "One of CERN's mottos is 'science for peace,'" says John Ellis, a theoretical physicist from King's College London who works at CERN and was on the lab's staff for more than 40 years. He notes that CERN did not expel Russian scientists when the Soviet Union invaded Czechoslovakia in 1968 or Afghanistan in 1979. "My personal attitude is that we should really strive to maintain that collaboration, if it's all politically possible."

The governing CERN Council appears to be walking a fine line. In a special session on 8 March, representatives from the lab's 23 member states voted to suspend Russia's "observer" status and barred its representatives from auditing the council's deliberations. But it did not expel the more than 1000 Russian scientists who make up roughly 8% of CERN's international users. In a statement, the council says it will continue to monitor the situation and "is ready to take further measures as appropriate."

Many Ukrainian scientists excoriate Western colleagues for attempting to remain neutral. Civilian deaths are mounting after Russia shifted its strategy and amped up indiscriminate shelling of civilian targets. "We need help in the isolation of Russia and Russian people from the world," says Sergiy Ryabchenko of Ukraine's Institute of Physics. "And we need more weapons."

"We are calling on the world scientific community to immediately stop the bloodshed and barbaric destruction of a civilized

European country," says Anatoly Zagorodny, president of the National Academy of Sciences of Ukraine. "Do not leave us alone facing the brutal aggressor." Western institutions must take a stand, says global security expert Gerson Sher, who led early efforts to engage Russian scientists after the Soviet breakup in 1991. "It's a moral issue, and silence is not an option."

Conditions in Kyiv and other besieged cities are rapidly deteriorating. Last week, missile fragments damaged a pipeline that provides heat and hot water to central Kyiv where Sergei Mosyakin, director of the Institute of Botany, resides. "I do not care much because I'm rather cold-resistant," he says. Rather, he worries how it will affect the elderly and young children.

tion," says Irina Yehorchenko, a mathematical physicist at the Institute of Mathematics who has spent her days shuttling between a bomb shelter and her Kyiv apartment. She hopes to join a humanitarian convoy bound for Lviv, in western Ukraine.

Labs in Europe, the United States, and elsewhere are throwing open their doors. Taras Oleksyk, a Ukrainian evolutionary biologist at Oakland University in Michigan, and colleagues have compiled a list of more than 1000 labs willing to host refugee scientists. The Polish Young Academy, part of the Polish Academy of Sciences (PAN), has already found accommodations and jobs for dozens of Ukrainian researchers, says Jacek Kolanowski of PAN's Institute of Bioorganic Chemistry. "I think in the upcom-

On 5 March, Ukrainians fleeing Kyiv crowd under a destroyed bridge over the Irpin River.

Mosyakin says many of his colleagues have taken up arms to defend their homes. Others have sought refuge abroad. Irina Belskaya, an expert on asteroids and comets at V.N. Karazin Kharkiv National University, reached Poland, she says, thanks to help from colleagues at the Poznań Observatory. Some are seeking sanctuary for their scientific prizes: Entomologist Valery Korneyev of the I.I. Schmalhausen Institute of Zoology in Kyiv was trying to get to Berlin with two plastic duffels stuffed with type specimens. "Keeping my fingers crossed," he says.

Other Ukrainian scientists have yet to escape the war zone. "I have no strength to stand for hours and hours at the train sta-

ing weeks we will have a logarithmic rise in requests," he says.

Russian scientists will be among the refugees. Rumors are swirling that Putin will declare martial law and close borders. Some Russian scientists who oppose the war are fleeing before a new Iron Curtain traps them. "They leave in fear of what is coming," says Igor Krichever, a Russian American mathematician at Columbia University who has spent several months each year in Moscow as the head of Skoltech's Center for Advanced Studies and is one of more than 500 Russian mathematicians who signed an open letter denouncing the invasion.

But room for dissent is disappearing in Russia. Last week, it expunged the last ves-

tiges of its free press. And the leaders of nearly 200 Russian universities have come out staunchly in support of the war. In a signed statement posted to the website of the Russian Union of Rectors on 4 March, the leaders commended Putin's decision to "achieve the demilitarization and denazification of Ukraine." The nation's premier nuclear research center, the Kurchatov Institute, which helped Ukraine deal with the aftermath of the Chernobyl disaster, posted a similarly prowar statement to its website on 7 March.

The U.S. government was among the leaders in enacting financial sanctions on Russia and slapping new limits on technology transfers. Still, a governmentwide policy on cooperation with Russia on R&D has yet to be publicly issued.

Some decisions couldn't wait for a directive. The U.S. National Oceanic and Atmospheric Administration (NOAA) intervened in an ongoing expedition with Canada and Russia to better understand salmon ecology in the North Pacific Ocean, where salmon congregate before spawning in rivers in all three countries. A U.S. scientist was supposed to travel on the Russian vessel *Tinro*, but on 24 February, NOAA forbade the scientist from boarding the ship. And last week, Eric Regehr, a University of Washington, Seattle, biologist, abandoned plans to join Russian researchers on remote Wrangel Island in an annual campaign, supported by the U.S. Fish and Wildlife Service, to study polar bear migrations from Alaska to Siberia. "There's no way," Regehr says. "The idea of it being legal and safe and practical to go over there is zero."

Skoltech has survived previous rough patches. In 2014, Russia annexed Crimea, and later that year, Russian-backed separatists shot down a Malaysia Airlines flight over eastern Ukraine. The Dutch director of Skoltech's stem cell research center knew someone who died in the crash and resigned. Skoltech "persevered," Seidel says. But he fears this time it's different. He's already shelved an idea for a partnership he was exploring with a Skoltech colleague.

Skoltech hopes an infusion of young Russian talent will help overcome its impending isolation: Last week, it invited transfer applications from overseas Russian graduate students facing blowback from the war overseas. But so long as Russia remains a pariah, Skoltech will be crippled, Seidel says. "It leaves me heartbroken. I don't know how they'll recover from this." ∎

With reporting from Edwin Cartlidge, Adrian Cho, Daniel Clery, Warren Cornwall, Andrew Curry, and Jeffrey Mervis.

Rocks at the foot of Hiawatha Glacier contained shocked zircon crystals that revealed the impact's age.

EARTH SCIENCE

Impact crater under Greenland's ice is surprisingly ancient

New date of 58 million years undercuts idea that strike triggered recent Younger Dryas climate cooling

By **Paul Voosen**

In 2018, an international team of scientists announced a startling discovery: Buried beneath the thick ice of the Hiawatha Glacier in northwest Greenland is an impact crater 31 kilometers wide—not as big as the crater from the dinosaur-killing impact 66 million years ago, but perhaps still big enough to mess with the climate. Scientists were especially excited by hints in the crater and the surrounding ice that the Hiawatha strike was recent—perhaps within the past 100,000 years, when humans might have been around to witness it.

But now, using dates gleaned from tiny mineral crystals in rocks shocked by the impact, the same team says the strike is much, much older. The researchers say it occurred 58 million years ago, a warm time when vast forests covered Greenland—and humanity was not yet even a glimmer in evolution's eye. Kurt Kjær, a geologist at the Natural History Museum of Denmark and a co-author of the new study, says the new date is at odds with the team's initial impression, gleaned from ice-penetrating radar. "But this is the way science works and should work," he says.

The date is a blow to a group of scientists that for more than a decade has advanced a controversial hypothesis that the Younger Dryas, a drastic, 1000-year cooling about 12,800 years ago, was triggered when a comet struck Earth. They had seized on the first Hi-awatha paper as a smoking gun: The crater seemed about the right age, and it was in the right place—near a region of the North Atlantic Ocean that heavily influences Northern Hemisphere climate. Now, says Brandon Johnson, a co-author and impact modeler at Purdue University, West Lafayette, "It's probably safe to put the Younger Dryas impact hypothesis back to rest for a while."

James Kennett, a marine geologist at the University of California, Santa Barbara, and a leading Younger Dryas impact advocate, says the older date for the crater is a surprise, but Kjær's team "makes a very compelling case … I don't think it's related to the Younger Dryas now." That leaves his group where it was before the discovery of Hiawatha: arguing the Younger Dryas trigger was an airburst rather than a body slamming into the ground. Kennett says the team will continue to advance its case with evidence from more than 40 sites worldwide that contain glassy spherules or platinum-rich sediments, which the group believes are indicative of an impact. "It's all alive and well and very active."

Kjær's team originally thought dating the impact would be impossible without drilling through 1 kilometer or so of ice to sample rocks in the center of the crater. The radar data, however, yielded clues to what seemed to be a young age: reflections indicating ice layers older than 11,700 years are deformed, hinting at an impact around that time.

But in 2019, the team got a chance to date

the impact directly. Returning to the rivers that spill out from the foot of the glacier and deposit sediment from beneath the ice, they found fist-size rocks that had experienced melting, ostensibly from the heat of the impact. Slices of those rocks went to the lab of Gavin Kenny, a geochronologist at the Swedish Museum of Natural History, who sifted out crystals of the mineral zircon smaller than grains of sand.

Some of zircon crystals were "shocked"—inscribed with linear fracture patterns that are the hallmark of an impact. Trace amounts of radioactive uranium are present in the zircon, and its decay into lead provides an accurate way to date the samples. The impact kicked out the lead impurities in the shocked zircons, effectively resetting the uranium clock. In a study published this week in *Science Advances*, the researchers report that in 28 of these shocked zircon crystals, the decay clock points to an age of 58 million years ago, with an uncertainty of about 1 million years. Nearly 50 grains of sand collected from the same watershed, analyzed using the decay of radioactive potassium to argon, yielded about the same age.

Given the agreement between the two dating systems, "It seems fairly rigorous to me,"

than previously assumed, MacGregor says.

The 1.5-kilometer-wide asteroid that produced Hiawatha would have been regionally devastating, but there is no sign that the dust cloud and fires that might have followed the impact disrupted the global climate 58 million years ago. The strike would have come 3 million years before the Paleocene-Eocene Thermal Maximum (PETM), a 100,000-year temperature spike that some have used as an analogy for human-induced climate change.

But Sidney Hemming, a geochemist at Columbia University, says the age data are complex enough that the uncertainty might be as much as 5 million years, opening the possibility that the impact and the PETM are connected. "I'd be hard pressed to be that confident that it's not that," she says. She points out that glass spherules tied to the PETM, presumably forged and thrown up in an impact, have been found off the coast of New Jersey. For now, the Hiawatha team is combing through geological records for signs of disturbances 58 million years ago, Kjær says. "We're commencing that journey now."

Other impact mysteries in Greenland remain to be solved. Soon after the original Hiawatha paper came out, MacGregor

Hard knocks

Between the dinosaur-killing Chicxulub impact and a later one that carved the Chesapeake Bay, a smaller rock hit forest-covered Greenland. The new date is a blow to researchers who thought Hiawatha was much younger.
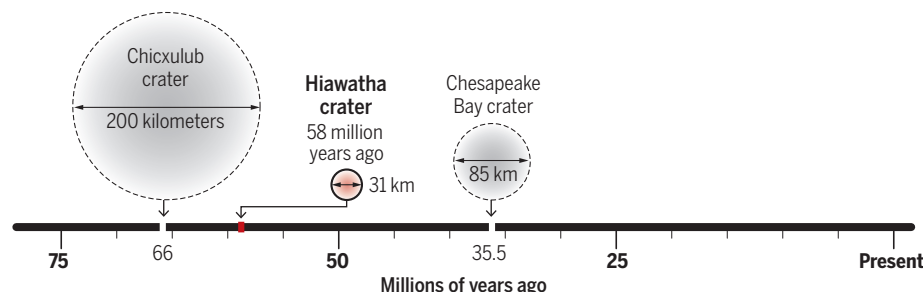
says Sandra Kamo, a geochronologist at the University of Toronto.

Now, the team is wondering whether the distorted ice it initially took as signs of a recent impact resulted instead from the sudden collapse of ice that bridged the ice sheets covering Greenland and Canada's Arctic archipelago during the last ice age. "Eventually they disconnected—presumably with some dynamic consequences," says Joseph MacGregor, a co-author and glaciologist at NASA's Goddard Space Flight Center.

Similarly, the sharp, well-preserved crater rim seen on radar might be a sign not of youth, but of slow erosion beneath Greenland's ice, Johnson adds. In that case the large, incised valleys detected beneath the ice elsewhere in Greenland could be much older

identified a possible second impact crater nearby, larger and more eroded than Hiawatha. (It remains unconfirmed.) And the region is also famed as the home of fragments from a massive iron meteorite that weigh in total some 58 tons. "It is a hot spot for impact up there," Kjær says.

The study is also a good reminder that, despite all the interest in catastrophic asteroid impacts, none has yet been clearly shown to have caused a global environmental change—other than the dino killer 66 million years ago at Chicxulub, on Mexico's Yucatán Peninsula. "I love impacts more than your average scientist," Johnson says. "But when you have some piece of data that is difficult to describe or understand, impacts are usually not the answer." ∎

When is a pandemic officially 'over'?

World Health Organization will confront thorny decision with big implications

By **Meredith Wadman**

Every 3 months since January 2020, when it first named the SARS-CoV-2 outbreak an international public health emergency, a committee of expert advisers to the World Health Organization (WHO) has convened to assess whether the pandemic still merits that label. And every 3 months, most recently in January, the advisers have unanimously agreed it does, and WHO Director-General Tedros Adhanom Ghebreyesus has accepted their verdict. When the committee meets again next month, it is likely to reach the same conclusion—and Tedros is again likely to accept it.

But at some point—estimates range from months to years from now—WHO will make a different call. Already, nations such as Denmark, the Netherlands, and the United Kingdom have functionally declared an end to the pandemic in their countries, lifting almost all health restrictions. On 7 March, influential U.S. scientists put out a "next normal" plan that specifies a metric for calling an end to the pandemic: 165 U.S. deaths per day from all major respiratory illnesses, including COVID-19.

For now, COVID-19 alone is causing nine times as many deaths in the United States—and New Zealand and Hong Kong are struggling with record-breaking surges. Deciding when to sound the all-clear is "not an enviable task," says Yonatan Grad, an infectious disease epidemiologist at the Harvard T.H. Chan School of Public Health (HSPH). "Do you call it over when there might still be a wave in one part of the world but it's a small part?"

"SARS-CoV-2 has caused such hardship and economic challenges that there will be a temptation to call it as over sooner rather than later," says Salim Abdool Karim, an epidemiologist who is the South African government's chief COVID-19 scientist. The prospect worries him. WHO's formal declaration of a Public Health Emergency of In-

A woman sheds her mask in Madrid, where pandemic restrictions were eased last month.

ternational Concern (PHEIC) legally binds 196 signatories to follow WHO's recommendations during the emergency. Drugmakers have also signed contracts agreeing to make anti–SARS-CoV-2 pills more affordable until the PHEIC is reversed. Other big, cooperative efforts help make diagnostics and vaccines affordable and distribute them worldwide, and "all of those things … will fall away. And those are the mechanisms that are needed by the poor countries," says Karim, who also runs an HIV research center in South Africa. "Getting it wrong will carry a high price."

To many outside China, where COVID-19 struck first, a statement by Tedros 2 years ago this week describing SARS-CoV-2 as a global pandemic marked its official start. But his 11 March 2020 comments triggered no public health requirements. Rather, the declaration with practical implications was the 30 January 2020 PHEIC announcement.

The regulations governing the PHEIC require signatory nations to report suspected outbreaks to WHO and to support its responses, although WHO has no way to enforce those rules. The expert committee that recommends whether to continue the PHEIC also lists, with each renewal, actions that nations should take, such as improving variant surveillance and expanding each nation's vaccination coverage. At the start of this year, for instance, the emergency committee added a new recommendation: Monitor and share data on cases and evolution in animals.

The decision to end a PHEIC has financial implications, too. Moderna in 2020 pledged not to enforce patents on its messenger RNA vaccine until the pandemic ends, although this week it ended that pledge for wealthier nations while making it permanent for poorer ones. Pfizer has not made a similar vaccine pledge, but it and Merck have agreed to allow generic drugmakers to make their drugs targeting SARS-CoV-2 until WHO declares the PHEIC is over. Dozens of companies have now signed up to make Merck's molnupiravir and Pfizer's Paxlovid for a long list of mostly low- and lower-middle-income countries.

Ending the PHEIC will also impact major pandemic-related programs such as the COVID-19 Vaccines Global Access (COVAX) Facility and its parent, the Access to COVID-19 Tools Accelerator (ACT-A)—cooperative global networks that aim to acquire and distribute affordable drugs, diagnostics, and vaccines. "The emergency operations of COVAX and ACT-A will probably go away—it's hard to keep that up," says Seth Berkley, CEO of GAVI, the Vaccine Alliance, which is integrally involved with both efforts. "The hope is that the core innovations—the ways of working all of that—will be kept warm" for the future.

WHO's 18-member committee uses three criteria to decide when to declare a PHEIC and when to lift it. A public health event must be "serious, sudden, unusual or unexpected"; likely to spread internationally; and likely to require immediate international action. When unwinding an emergency, the committee considers such metrics as vaccinations and case numbers. But the criteria are more social and political than scientific, says Caroline Buckee, an infectious disease epidemiologist at HSPH. At WHO, "There's not going to be a scientific threshold. There's going to be an opinion-based consensus," she says.

Complicating the decision is the prospect of additional harmful variants arising, including, potentially, from some 20 animal species now known to host the virus. "I don't know how it ends," says Michael Osterholm, an infectious disease epidemiologist at the University of Minnesota, Twin Cities.

Karim says the real end of the pandemic won't come until the arrival of a "final variant [that] even if it mutates, can't do better … than the previous version" in spreading and in escaping immunity. "If I was a betting man, I would say probably in about 2, 3 years we will get to that point."

WHO was conservative in lifting previous PHEIC declarations—there have been six including SARS-CoV-2 since the regulations took effect in 2007—says Horace Cox, director of vector-borne diseases at the Ministry of Public Health in Guyana, and he expects the same with SARS-CoV-2. Individual countries have been less gingerly signaling a return to normal, however. Several European countries have lifted restrictions already. And the U.S. Centers for Disease Control and Prevention late last month eased masking recommendations for some 70% of U.S. residents. In Congress, Republicans are this week resisting additional pandemic funding, and some have introduced a bill to end the federal emergency declared in March 2020.

"The expectation is that the United Kingdom and the U.S. will be well ahead … where they say: 'We don't think this is an issue anymore. We're making our own decision,'" Cox says. "[But] the WHO [must] consider what is good for the entire world."

Still, he is cautiously optimistic that a WHO determination that the PHEIC is over may not be too distant. "If I were to make an educated guess, I would say that perhaps by late second and third quarter [2022]," if another deleterious variant doesn't emerge.

But Osterholm is making no predictions. "If there was ever a time for humility among scientists and policymakers, it's now," he says. "We are in totally uncharted territory from the perspective of understanding what a pandemic is, how it starts, how it unfolds, and how it ends." ∎

Africa battles out-of-control polio outbreaks

Cases tumble in Pakistan and Afghanistan but African outbreaks now threaten eradication

By **Leslie Roberts**

On 17 February, Malawi's Ministry of Health announced a nasty surprise: A 3-year-old girl who was paralyzed in November 2021 was infected with the wild poliovirus, which Africa officially vanquished in 2020. The sequence of the virus showed it had somehow made the leap from Pakistan, one of the last two holdouts of the wild virus. A week later came bad news from Afghanistan: Gunmen killed eight polio workers in the country's northeast.

The incidents are the latest setbacks on the long, bumpy road to global polio eradication. Yet Pakistan has "exported" wild poliovirus before, sparking outbreaks that were quickly snuffed out, and the situation in Afghanistan and Pakistan improved dramatically last year, with polio cases tumbling to a historic low.

Instead, perhaps the biggest threat to the effort now is an explosion of vaccine-derived polio outbreaks in Africa that affected almost two dozen countries last year and paralyzed more than 500 children in 2020 and again in 2021. Vaccine-derived strains emerge where children are un- or underimmunized, allowing the live, weakened virus in the oral polio vaccine (OPV) to circulate and accumulate enough mutations to revert to its neurovirulent form and paralyze kids. These outbreaks—which almost always emerge from type 2 poliovirus, one of the three vi-rus strains—are "very worrying" and "front burner" at the Global Polio Eradication Initiative (GPEI), says John Vertefeuille of the U.S. Centers for Disease Control and Prevention (CDC), a partner in the initiative.

A big part of the problem is that countries don't view vaccine-derived strains as an emergency, says Simona Zipursky, an adviser to the World Health Organization's (WHO's) polio program, even though they behave just like the wild virus. "It is not like there is a milder variant as there is with COVID-19," she says. Nigeria's widely lauded victory over the wild virus—it was the last African country to achieve that feat—fed a sense that "the job was done," says WHO's Aidan O'Leary, who directs GPEI. The quality of Nigeria's polio program, once among the best in the world, slipped, and today the country is "the most important generator" of vaccine-derived polioviruses, says Jay Wenger of the Bill & Melinda Gates Foundation, another partner in GPEI. Nigeria accounted for more than half of all vaccine-derived polio cases globally last year and exported the virus to 18 countries.

Other factors have contributed, too. Lately, many African countries have been slow to respond to new outbreaks. They are frustrated with the vaccine currently used to quash outbreaks, monovalent OPV2 (mOPV2), because it, too, occasionally reverts and seeds new cases. Instead they often choose to wait for a recently developed vaccine that is meant to solve the problem.

Known as novel OPV2 (nOPV2), the new vaccine was engineered to be as effective as mOPV2 but more genetically stable, greatly lessening the chance it will revert (*Science*, 13 November 2020, p. 751). The vaccine, funded by the Gates Foundation, was rolled out in a few countries in March 2021 under an emergency use authorization. Pending its arrival, Senegal waited for almost 1 year before responding to a virus detected in late 2020, instead of using readily available supplies of mOPV2. "If the virus gets a head start for such a long time it is harder to stop," says Mark Pallansch, who recently retired from CDC but remains involved in GPEI.

Although early data suggest nOPV2 is indeed less likely to trigger outbreaks, Pallansch thinks its promise has been oversold. "Governments thought, if I can just get it, things will be fine," he says. But countries ran nOPV2 campaigns of poor quality, reaching just a fraction of the target population. Nigeria has burned through about 184 million nOPV2 doses, out of 255 million used so far, and still hasn't stopped many of its outbreaks. The new vaccine "is not a magic bullet," Zipursky says.

GPEI and other international bodies are hammering home that countries should respond to any detection of the virus—whether it's a polio case or a positive environmental sample—immediately with whatever type 2 vaccine is available. The mantra is "faster, better, bigger," O'Leary says: Be quicker to detect and respond to outbreaks, improve the vaccination campaigns, and broaden them. "We need to conduct them not where you think the virus is, but, based on migration patterns, where you think it will be," he says.

The Africa campaign is also suffering from a "self-inflicted wound," Pallansch says. GPEI has long planned to put itself out of business once polio is gone. As part of a transition plan, many of GPEI's substantial assets and staff would be integrated into existing WHO programs, such as those dealing with other infectious diseases—GPEI has already helped with Ebola and COVID-19—and to boost routine immunization. WHO planned to complete this transition in nonendemic countries, including all of Africa, by January.

Accordingly, in February 2021, WHO's Africa office sent pink slips to all GPEI staff. Unfortunately, the office was slow to say who would be kept on, and some people got ner-

Tale of two threats

In Pakistan and Afghanistan, the last two countries where the wild poliovirus is endemic, cases dropped sharply last year (left). Vaccine-derived outbreaks are now a big threat to polio eradication—especially in Africa (right).

Wild polio cases
● Afghanistan ● Pakistan ● Rest of world

Vaccine-derived polio cases
● Afghanistan and Pakistan ● Africa
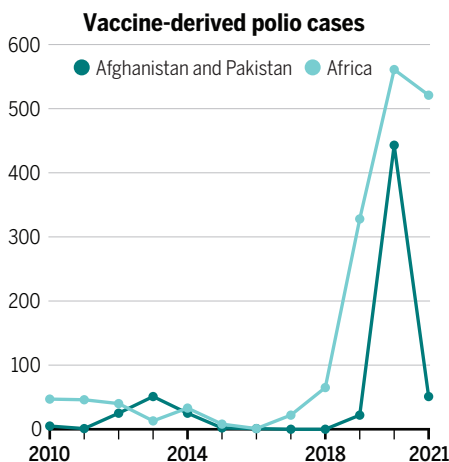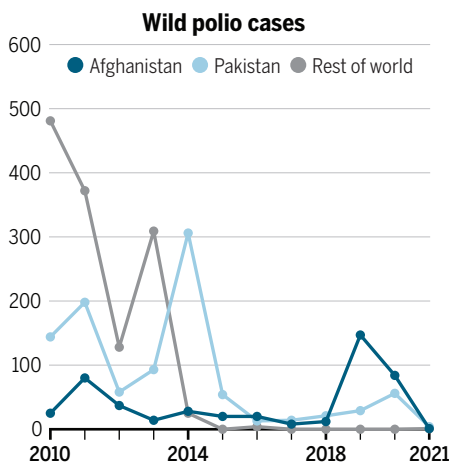
vous and quit, officials say. GPEI soon realized the Africa situation was "too hot right now" to proceed with the plan, Wenger says, and decided to continue to fund the 10 highest risk countries for another 2 years. But the damage had been done. "Things didn't have to happen this way," Pallansch says. "They could have done it in a different sequence and not have viruses all over the continent."

The new worries come as Pakistan and Afghanistan, the last two endemic countries, are doing surprisingly well, with just five reported cases of wild poliovirus last year, down from 140 in 2020. Pakistan has just gone an entire year without a case. (Vaccine-derived cases in both countries are way down as well.) "It looks better than it ever has," Wenger says. The low numbers are "absolutely not" an artefact, says Hamid Jafari, who heads the eradication program in the region; surveillance remains "really, really good." Some of the gains stem from very favorable epidemiology. Polio resurged in both countries in 2019 and 2020, and "after a peak we always see a trough," Jafari says, in part because of increased population immunity. Reduced travel during the COVID-19 pandemic helped.

In Pakistan, vaccination drives already cover most of the target population, and they are improving, Jafari says. Imran Khan, Pakistan's prime minister, is actively involved. Bill Gates just visited the country to bolster enthusiasm. In Afghanistan, too, "we've made more progress than we could have anticipated," O'Leary says. After resuming power in August 2021, the Taliban rescinded its ban on house-to-house polio vaccination in its strongholds, which had left 3.5 million children out of reach. (In some areas GPEI is still restricted to vaccinating in mosques.) Vaccination campaigns in November, December, and January reached 8.5 million out of 9.9 million children, Jafari says, including 2.6 million who were inaccessible for 3.5 years.

But future campaigns could be hobbled if last week's killings are a harbinger of further violence. And Jafari suspects the virus may survive in small populations that move back and forth across the border between the two countries. A couple of positive environmental samples detected in December 2021 in the south of Pakistan's Khyber Pakhtunkhwa province show the virus is still lurking there. Jafari worries it could resurge when the weather warms and people begin to travel for Ramadan and Eid al-Fitr. The recent spread to Malawi underscores the risk of further delays, he says: "We want to kill the virus now." ∎

Leslie Roberts is a science journalist in Washington, D.C.

New class of killer T cells may prevent autoimmune diseases

Study finds human version of mouse immune regulators

By **Mitch Leslie**

The price for a vigilant immune system that can pounce on tumor cells or pathogens is occasional friendly fire—an autoimmune attack. Scientists have now identified a new type of human T cell that quells assaults on healthy tissues, a finding that could suggest treatments for conditions as diverse as lupus and cancer. "It's a major step forward in understanding how the immune response and autoimmunity are regulated," says immunologist Harvey Cantor of the Dana-Farber Cancer Institute, who wasn't involved in the work.

Immunologists already knew mice and people deploy one type of regulatory T cell—a subset called Tregs that sports the protein CD4 and suppresses autoimmune attacks. The newer enforcers belong to a category of T cells distinguished by a different surface protein, CD8. CD8 T cells are best known for killing infected or cancerous cells, but in mice, some of them also kill T cells that orchestrate autoimmune reactions. Although researchers have long suspected humans have similar cells, nobody had confirmed their existence.

One obstacle was that humans don't make the distinctive receptors that mark the subset of CD8 cells in mice. However, some human CD8 T cells flaunt comparable receptors, the KIR proteins. To determine whether these human cells are immune inhibitors, Jing Li, a postdoc in the lab of immunologist Mark Davis at Stanford University's School of Medicine, and colleagues measured their abundance in patients with autoimmune diseases such as multiple sclerosis, lupus, and celiac disease. The cells were more common in blood from patients than in blood from healthy people, the team reports online this week in *Science*. Tissue samples revealed they congregated in parts of the body damaged by the autoimmune attack, such as the joints in people with rheumatoid arthritis and the small intestine in people with celiac disease.

The researchers detected similar surges of the KIR-producing T cells in people fighting infections, especially the pandemic coronavirus. In 56 COVID-19 patients, "We saw the KIR-positive cells going through the roof,"

> ## "These cells exist in humans."
> **Nu Zhang,**
> University of Texas

Davis says. These new regulatory cells may calm the immune system after severe infections, a crucial function in COVID-19, where many deaths are attributed not to the virus, but to an immune overreaction to it.

To investigate the cells' role in autoimmunity, the scientists homed in on celiac disease, which is triggered by the gluten proteins in bread and other grain-based foods. In the condition, so-called helper T cells recognize gluten proteins such as gliadin and then spill molecules that promote inflammation. But in cell culture studies, Li and colleagues found, human CD8 T cells carrying KIR proteins killed the gliadin-detecting helper T cells. "That really opened up a window for us to understand the biology of these [KIR+] cells," Li says.

To find out how much protection the cells provide against autoimmunity, the team analyzed genetically altered mice that have 50% to 75% fewer of the suppressive CD8 cells than normal. After exposure to certain viruses that can trigger autoimmune disease, the rodents developed signs of damage such as kidney inflammation. In contrast, control mice with a full complement of suppressive CD8 T cells didn't show evidence of autoimmune disease after infections.

Other scientists are convinced the team has fingered the long-sought human counterparts to the rodent immune regulators. "The paper provides really solid data that these cells exist in humans," says immunologist Nu Zhang of the University of Texas Health Science Center, San Antonio.

Immunologist Stephen Jameson of the University of Minnesota Medical School says approaches that increase the cells' abundance might soothe difficult-to-treat autoimmune illnesses such as celiac disease. It's also possible, he adds, that the cells are "sitting in tumors" and shielding them from immune attacks, in which case reducing the cells' numbers could unleash a person's immune system to fight cancer. Researchers have attempted to harness the traditional, CD4-carrying Tregs for therapies, but no treatments have been approved, Cantor notes. "The hope is that with this new set of regulatory cells, we can use them more efficiently." ∎

HEALING VENEZUELA

A network of young scientists and doctors aims to rebuild the country's devastated public health system   *By* **Richard Stone**

The reconnaissance trip in western Venezuela was going smoothly—until a gunman took aim at their windshield. It was March 2019, and two infectious disease specialists, Alberto Paniz Mondolfi and Carlos Hernandez, were driving back to their headquarters from villages in Venezuela's disease-ridden Portuguesa state, where they hoped to set up a campaign to collect health data.

On a desolate stretch of highway, a young man leapt from the bushes and took up a shooting stance in their lane. "I even remember the song on the radio. *Money for Nothing.* Dire Straits," Paniz Mondolfi says. "I said, 'Carlos, put your head between your legs, I'm going to try to go around this guy.'" He swerved into the opposite lane and pushed hard on the gas pedal. A shot rang out. And then more shots. Luckily, the would-be carjacker missed.

Rebuilding public health in chaotic, poverty-stricken Venezuela is not for the faint of heart. The oil-rich country once supported a robust science establishment. But a yearslong economic crisis exacerbated by the COVID-19 pandemic has devastated Venezuela and its

public health infrastructure, allowing infectious disease to flourish. Malaria cases, for instance, rose nearly 20-fold from 2001 to 2017, Paniz Mondolfi and colleagues reported in *The Lancet Infectious Diseases* in May 2019. Now, he is on a mission to fight disease in his native country, one harrowing step at a time.

In 2016, the deteriorating conditions in Venezuela prompted Paniz Mondolfi, who is the director of molecular microbiology at Mount Sinai Hospital in New York City,

IVC founder Alberto Paniz Mondolfi directs a team in Venezuela from Mount Sinai Hospital in New York City.

An IVC team member collects a kissing bug, which can transmit Chagas disease, from a gap in the wall of a mud house in rural Lara state in Venezuela.

to launch the nonprofit Venezuelan Science Incubator (IVC), a national network of 60 young physicians, biologists, veterinarians, and others who are diagnosing and treating infections and educating the public on hygiene and sanitation. The incubator has also enlisted an army of citizen-scientists to serve as disease sentinels, collecting basic epidemiological data and alerting IVC to outbreaks that the Venezuelan government's destitute health agencies now largely ignore. "They're doing amazing things," says Martin Llewellyn, a molecular ecologist at the University of Glasgow.

With help from the Rotary Foundation, IVC has just opened what co-leader Isis Mejías, an environmental consultant in Houston, bills as Venezuela's "first state-of-the-art molecular diagnostics lab" in Barquisimeto, IVC's home base. It will help detect pathogens responsible for everything from Chagas disease and leprosy to leishmaniasis, Zika, Mayaro, and malaria. "Venezuela has just a tremendous diversity of infectious

diseases," notes Emilia Sordillo, an infectious disease specialist at Mount Sinai.

That diagnostic firepower has allowed IVC to embark on its most ambitious project yet: a pilot for a nationwide program to distribute drugs in communities hard-hit by parasitic worms.

Along the way, IVC's volunteer corps—many of whom are still earning their college degrees—battle daily deprivations and obstacles. They camp out with hundreds of others at gas stations, waiting for sporadic deliveries to buy rationed fuel, and endure shakedowns at traffic checkpoints. "These kids are so brave," says Sordillo, who is not affiliated with IVC. They are animated by the mission Paniz Mondolfi has set for IVC: "We need to create and construct a new country, a new culture."

A GENERATION AGO, science flourished in oil-rich Venezuela. "There were never problems with the research budget," Paniz Mondolfi says. But former Venezuelan strongman Hugo Chávez expressed a disdain for science soon after taking power in 1999 and railed against intellectuals on his Sunday talk show *Aló Presidente*. Many academics stuck it out—at least until after Chávez's successor, Nicolás Maduro, took power in 2013. Soon afterward, oil prices cratered, collapsing Venezuela's economic house of cards. Ever since, "There's been an enormous brain drain," says Llewellyn, who did his Ph.D. on Chagas disease at the Central University of Venezuela (UCV).

Scientists marooned in Venezuela are mostly idle. Labs have been stripped of anything of value. Criminals pillaged the Institute of Tropical Medicine more than 20 times in 2016 alone, says IVC Academic Coordinator Juan Carlos Gabaldón Figueira, who left Venezuela in 2019 and is now pursuing a Ph.D. in medical parasitology in Barcelona, Spain. "Basically, nothing is happening in science now," says Ismardo Bonalde, a physicist at the Venezuela Institute for Scientific Research and president of the Venezuelan Academy of Physical, Mathematical and Natural Sciences. As one of the nation's top scientists, Bonalde earns less than $20 per month—"a joke," he says. "How can I live on that?"

Venezuela once had enviable health statistics. Now, "Venezuelans are in survival mode," Mejías says. Most lack access to an uninterrupted water supply and basic sanitation services. On top of that, "They lose track of things they need to do to maintain health, like washing their hands and washing food." Medicine is scarce, many hospitals are dysfunctional.

Meanwhile, the government has choked off data collection. In November 2014, the

Ministry of Health stopped publishing *Boletín Epidemiológico*, its weekly and monthly epidemiological reports. In 2018, it abolished the Venezuelan Center for the Classification of Diseases, which for 63 years had provided disease data to the Pan American Health Organization and the World Health Organization (WHO). "Venezuela is a black box," Llewellyn says. Against that dispiriting backdrop, "We're trying to help reignite research in Venezuela," Paniz Mondolfi says.

Paniz Mondolfi spent his childhood in Kenya, where his grandfather, a zoologist,

Melting pot of maladies

Venezuela's economic woes have sparked a resurgence of infectious diseases that the Venezuelan Science Incubator in Barquisimeto is combating, with a focus on Lara and Portuguesa states. The separate, long-running Maracaibo Aging Study is charting the toll of social dysfunction on the elderly.

served as Venezuela's ambassador to the United Nations Environment Programme. "I was always dissecting animals with him." He wanted to be a marine biologist or an astrophysicist, but after moving to Venezuela for college he earned a medical degree and joined UCV's vaunted Institute of Biomedicine. There, Paniz Mondolfi studied the immunology of leishmaniasis, hoping to embark on a Ph.D. in immunology. But Venezuela's universities were withering, and in January 2007, he got a terrifying phone call. A national guardsman claimed his wife's car registration was forged, had detained her at a checkpoint, and was demanding a ransom of about $1000. "That was more than I'd make in a million years in public

health in Venezuela," Paniz Mondolfi says. His father came up with the cash, and his wife was freed.

Days later, the couple was on a plane to New York City. Paniz Mondolfi did a pathology residency at St. Luke's–Roosevelt Hospital Center and a series of medical fellowships at U.S. universities. He and his family returned to Venezuela in 2014. He tried to make a go of it as a dermatopathologist in Barquisimeto, his wife's hometown. His family stayed with his in-laws for what was meant to be a 6-month transition, but "we were so poor we never made it out of their house."

Paniz Mondolfi negotiated a deal to house IVC in a private hospital whose elite clientele safeguards it from power and water outages. The "inventive" arrangement, Llewellyn says, "jibes with how things work in Venezuela." At the outset, the institute's primary focus was on Zika virus, which caused a massive epidemic in Latin America in 2015 and 2016. Reports from Brazil indicated pregnant women infected with Zika were giving birth to infants with abnormally small heads and underdeveloped brains. IVC saw the same in Venezuela and also reported neurological symptoms not previously linked to Zika, including rapid eye movements—ocular flutter—from damage to the brainstem

Using a mobile lab, volunteers for IVC examine a sample from a lesion of a patient with leishmaniasis.

and cerebellum, and Alice in Wonderland syndrome, in which people experience distorted visual perception of objects.

Such revelations haven't sat well with the government. In early 2019, two health officials showed up at the private hospital and delivered a blunt warning: If Paniz Mondolfi continued to publish data without their consent, they'd set the country's intelligence police on him. Soon afterward, he and his family left for New York City, taking some of the heat off IVC. "They assumed they'd severed the head of our organization," he says. As an added safety measure, IVC personnel cite joint affiliations with Colombian institutions when publishing papers.

AS IVC'S ASPIRATIONS to tackle other diseases began to outstrip its technical capacity, assistance arrived from an unexpected source. In the summer of 2018, Mejías, who was born in Valencia, Venezuela, was hiking in the country's rugged Canaima National Park with a group that included a cousin of Paniz Mondolfi, who bent her ear about the nascent science incubator. Mejías had a deep connection with the global network Rotary, which had funded her work to improve sanitation at primary schools in Uganda.

Mejias soon joined the IVC team, and in 2020 she helped secure $82,700 from Rotary and other sources for the new molecular diagnostics lab in Barquisimeto. U.S. economic sanctions on Venezuela deterred U.S. firms from selling them scientific equipment, which was "so heartbreaking," Paniz Mondolfi says, but European dealers came

through, and now the lab has a biosafety cabinet for handling risky pathogens, a benchtop DNA sequencer, and a miniature immunoassay reader for detecting peptides, antibodies, and other substances in blood samples. Llewellyn calls it "high-impact science on a shoestring."

IVC is now targeting Chagas, a tropical malady resurging in Venezuela. The disease is caused by *Trypanosoma cruzi*, a protozoan transmitted by triatomine insects, known as kissing bugs for their vampiric penchant for biting the face and sucking blood. Chagas ravaged Venezuela in the 1960s and '70s before nationwide campaigns to spray pes-

Fever line

Annual malaria cases in Venezuela have risen steeply since 2010. Other infectious diseases have surged as well, but the government has choked off data collection.

ticides and build bug-resistant dwellings. "A lot of it is education. How close you build your chicken hutch to your house is quite crucial," Llewellyn says. Those interventions faded under Chávez. Anecdotal reports suggest the rural scourge is making inroads in Venezuela's urban slums. But with scant data, Llewellyn says, "we don't quite know what's happening."

When IVC gets wind of a Chagas outbreak, it dispatches a team to collect kissing bugs, which are examined for *T. cruzi*. They educate people on measures to reduce exposure to the bugs. They also train people in hot spots to snap cellphone pictures of suspicious insects and upload them to a real-time geographic reference website called #TraeTuChipo, or "Bring Your Bug," designed by Lourdes Delgado Noguera, a medical student who started volunteering for IVC in 2018 when she was 17 years old. Entomologists chime in to identify and flag risky species. "Citizen science like this is an effective way of addressing public health in failing states," Llewellyn says.

Paniz Mondolfi huddles with IVC volunteers over Zoom every weekend, plotting out logistics and keeping the team up to speed on the science with an online journal club. When the team is in the field analyzing samples, he can access the software platforms in real time to confer with them on what they are finding. Late last month, he got "Mayday!" texts from team members who had encountered outside Barquisimeto a rash of leishmaniasis cases, including a rare outbreak of the disease in housecats.

NOW, IVC IS GEARING up for Deworming Venezuela. The inspiration was the 2019 research campaign in Portuguesa, which the team launched a week after the attempted carjacking. When they got to the villages, they encountered rampant malnutrition and intestinal parasitosis. "Basically, every single kid was infected," Paniz Mondolfi says. In a pilot project that kicked off in late January, IVC is studying the link between worm infections and poor sanitary conditions in Portuguesa and Lara and administering the antiparasitic drug ivermectin.

IVC hopes to scale up Deworming Venezuela into a national program, although that would require partnering with a major international organization such as WHO. Meanwhile it forges ahead on meager funding—and the dedication of its tight-knit cadre of volunteers. "It's quite a feeling to know you have found a family. Friends willing to persevere on the same journey," says Nina Pagani, who was born in Caracas and joined IVC as a high school junior in 2017. Thanks to Pagani and others of her generation, Paniz Mondolfi says, "we're keeping the scientific spirit alive in the middle of devastation." ∎
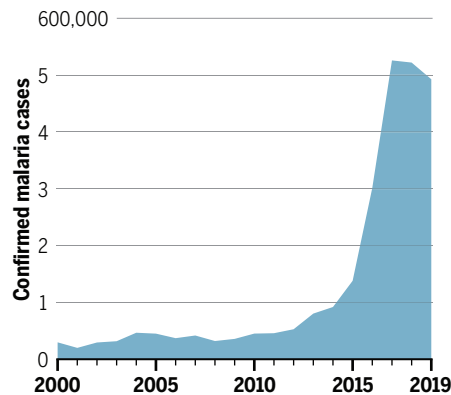
Long-running aging study charts health impacts of Venezuela's collapse

Times were good in Venezuela when Gladys Maestre and her colleagues set out to study memory disorders in Maracaibo, a coastal city near the border with Colombia. That was a quarter-century ago, when the country enjoyed a decent standard of living and universal health care. Since then, the long-running Maracaibo Aging Study (MAS) has acquired iconic status—and a tragic new role: charting the toll of social dysfunction, economic collapse, and hunger on the elderly.

Launched in 1997, the MAS went hunting for Alzheimer's disease in a cohort of 2453 people ages 55 or older in Maracaibo's Santa Lucia parish. Its goal was to test the notion that Alzheimer's is largely a phenomenon of the developed world. "Back then, I'd go to scientific meetings and people would say, 'We don't need to worry about Latin America, Alzheimer's is not there,'" says Maestre, a neuroscientist born in Maracaibo who directs the Alzheimer's Disease Resource Center for Minority Aging Research at the University of Texas, Rio Grande Valley. "In my heart, I was really hoping that was true." Her first visit to a nursing home in Santa Lucia dispelled that notion: "It was obvious that many people there were severely demented."

Maestre's team discovered that compared with U.S. studies, Alzheimer's incidence in Santa Lucia was twice as high in the 55 to 65 age group and 50% higher in people over age 65. And they were among the first to probe the link between cardiovascular health and cognition. "Today, everybody says what's good for the heart is good for the brain. But that wasn't an obvious connection 25 years ago," says Sudha Seshadri, a senior investigator with the famous Framingham Heart Study, now in its 74th year.

When Venezuelan government support for the MAS dried up in the late 2000s, the U.S. National Institutes of Health (NIH) stepped in. In 2011, NIH backed a major expansion of the MAS to probe the genetics of aging in some 2500 members of an extended family in Santa Rosa de Agua, an impoverished and isolated neighborhood on the shores of Lake Maracaibo. Intriguingly, dementia is less common in that community, says Columbia University geneticist Joseph Terwilliger, who spearheads the study. One factor, he says, is Santa Rosa de Agua's high mortality rate: Violence and accidents cause many early deaths. "But those who make it to 40 tend to live long, healthy lives," Terwilliger says.

A novel set of measurements led to an unexpected finding. Blood pressure readings during physical exams tend to be high because people are stressed during an examination, so Maestre's group pioneered the use of 24-hour blood pressure monitoring in an aging study. "No other study in the world has that kind of data," says Seshadri, director of the South Texas Alzheimer's Disease Research Center. About 10 years ago, the MAS team found that about 30% of the Santa Rosa de Agua cohort—"an unbelievably high rate," Maestre says—suffers from an eye condition named glaucomatous optic neuropathy that can cause blindness if left untreated. In 2018, they correlated the condition with extreme and still-unexplained dips in nighttime blood pressure.

The MAS entered a grim new phase when Venezuela's economy nosedived in the mid-2010s. Food and drugs became scarce, frosty U.S.-Venezuelan relations bred distrust of the NIH-funded effort, and the team lost a handful of study subjects who feared being seen as "collaborating with the enemy," Maestre says. Others left the country, as did several key MAS personnel. As power outages grew frequent and lengthy, Maestre scrambled to move frozen tissue samples out of Maracaibo to overseas labs. In 2018, hundreds of samples vanished when highway bandits seized a DHL truck en route to Caracas. They've since used a specialized courier service that flies samples out of Maracaibo.

Then the pandemic arrived, cutting a devastating swath through both the Santa Lucia and Santa Rosa de Agua cohorts, Maestre says. Many study subjects have developed clinical depression as they cope with the loss of loved ones to COVID-19 and separation from family or friends who emigrated, says Carlos Chávez, a physician at the University of Zulia who works with the project and lost his wife to the virus.

In September 2021, a National Survey of Living Conditions found 76.6% of Venezuelans live in "extreme poverty." Exorbitant prices allow most families to buy only a few days' worth of food every month, Maestre says. The MAS team has enlisted local shops to provide free meals during evaluations, for example, and commodi-

Gladys Maestre (left) examines a patient's carotid for signs of atherosclerosis in Maracaibo's Santa Lucia parish.

ties such as coffee and chickens when things were especially dire a couple years ago. It's a balancing act, as such interventions can skew data. Still, "We have to step out of our comfort zone and impact lives as best as we can," Maestre says.

This spring, her team is gearing up for another round of evaluations in Santa Rosa de Agua. They hope to learn whether the residents' meager diets are reducing the incidence of diabetes, although Maestre notes that any health gains may be offset by insufficient protein and deficiencies of some vitamins and minerals. Seshadri sees Venezuela's humanitarian crisis as "a horrible natural experiment" akin to the Dutch Hunger Winter study, which tracked the long-term health effects of a famine during the final months of World War II. By tracing the impacts of Venezuela's plight, Seshadri says, the MAS "might tell us something important for the world."

She adds that the study has already provided a lesson in resilience: "It's amazing they're able to keep it going." —*R.S.*

INSIGHTS

LETTERS

NEXTGEN VOICES

A pandemic education

To reflect on the 2 years since COVID-19 was declared a global pandemic, we asked young scientists to imagine that they could go back in time and create a university course that would help scientists navigate the years 2020 to 2022. Read their pandemic course catalog below. Follow NextGen Voices on Twitter with hashtag #NextGenSci. —**Jennifer Sills**

Communications

COM 145: Identification, analysis, and communication of scientific evidence
This course focuses on developing the skills required to translate scientific evidence into accessible information for the general public, especially under circumstances that lead to the intensification of fear and misinformation. Discussions will cover the principles of the scientific method, as well as its theoretical and practical relevance in counteracting the dissemination of pseudoscience, particularly on social media.

This course discusses chapters from Carl Sagan's book *The Demon-Haunted World*, certain peer-reviewed and retracted papers, and materials related to key science issues, such as the anti-vaccine movement. For the final project, students will comprehensibly communicate a scientific topic to the public.

Camila Fonseca Amorim da Silva
University of Sao Paulo, Sao Paulo, SP 03828-000, Brazil. Email: camilafonsecaamorim@usp.br

COM 198: Everyday science communication
As scientific discoveries become increasingly specialized, the lack of understanding by the general public undermines trust in scientists and causes the spread of misinformation. This course will be taught by scientists and communication specialists who will provide students with a toolset to explain scientific concepts, as well as their own research projects, to the general public. Upon completion of this course, students will be able to explain to their grandparents that viruses exist even though they can't see them, convince their neighbors that vaccines don't contain tracking devices, and explain the concept of exponential growth to governmental officials.

Anna Uzonyi
Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel. Email: anna.uzonyi@weizmann.ac.il

COM 232: Introduction to talking to regular people
Communicating science is difficult. Many scientists, having immersed themselves in the language of their field, have completely forgotten how to talk to regular people. This course hones introductory science communication skills, such as how to talk

about scary things without generating mass panic, how to calmly discourage the hoarding of paper hygiene products, and how to explain why scientific knowledge changes over time. The final project will include cross examination from law school faculty, who are otherwise completely uninvolved with the course and possess minimal scientific training. Recommended for science majors who are unable to discuss impactful scientific findings without citing a *P* value.

Joseph Michael Cusimano
Bernard J. Dunn School of Pharmacy, Shenandoah University, Winchester, VA 22601, USA. Email: joecusi@hotmail.com

Biology

BIO 156: Design strategies for epidemic containment
As population and global travel increase, public and private facilities should consider the threat of recurrent epidemics. This course will cover contagion dispersal rates through air and surface contact and information about the various means of countering each type. Examples include surface "traffic" and ventilation rates per occupant density. Students will explore nonmedical interventions such as the germicidal effects of ultraviolet radiation for both surface and airborne pathogens.

Thomy Nilsson
Department of Psychology, University of Prince Edward Island, Charlottetown, PEI C1A 4P3, Canada. Email: nilsson@upei.ca

BIO 274: Reflexes save lives
With globalization increasing the threat of disease, all preventive options must be explored. Although many focus on vaccines and drugs, quick reflexes may offer better protection. In this course, we will learn how to react when someone is coughing (duck), when someone claims that they have "done their own research" (scoff), or when a flock of antivaxxers approaches (hide). We will exercise our reflexes in simulated real-world situations. Finally, we will learn a little bit of science and history, as knowledge can save (your) lives.

Nikos Konstantinides
Institut Jacques Monod, 75013 Paris, France. Email: nikos.konstantinides@ijm.fr

Computer science

CPS 189: Introduction to online meetings
In this course, scientists will learn how to create meetings and waiting rooms, how to welcome invited guests, how to avoid uninvited guests, and how to remove unwanted filters that could make you look like a cat or a pickle. Most important, scientists will learn how to unmute themselves when speaking. Finally, this course will teach strategies to combat screen fatigue.

Fernanda Oda
Department of Applied Behavioral Science, University of Kansas, Lawrence, KS 66045, USA. Email: oda@ku.edu

CPS 276: Virtual conferences and networking
Navigating your way around a virtual conference platform can be more daunting and aggravating than a busy venue hall. And good luck connecting with peers and employers! This course will teach scientists of all ages and levels of experience how to engage online in the rare event that a conference is rendered virtual.

Ahmed Al Harraq
Cain Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA 70803, USA. Email: aahme22@lsu.edu

Anthropology

ANT 254: Stranded! The archaeology of shipwrecks, first settlements, and survival
This course will examine life in isolation, strategies for survival, identification of resources, cultivation of hidden talents, the art of staying sane, and the skills required to make any environment work for you. We will draw on real-life experiences from "first" settlements, shipwrecks, and preparations for colonies in space. To start, imagine yourself alone and stranded in an unfamiliar place, your homeland preoccupied with threats to its sovereignty, your supply line cut, and no rescue in sight—survival depends on ingenuity and perseverance. This class is for those who find themselves in isolated circumstances—astronauts, scientists in Antarctic labs, solo sailors circumnavigating the globe, and maybe you!

Felicia Beardsley
Department of Anthropology, University of La Verne, La Verne, CA 91750, USA. Email: fbeardsley@laverne.edu

Psychology

PSY 108: Resilience in research: How to collaborate in a virtual world
Research collaborations can be tricky enough to navigate in person, but how do you maintain effective networks and collaborate with colleagues beyond your institution when travel is not an option? This course will help you answer this question through the exploration of topics like leadership, group dynamics, and mentoring in virtual research settings, as well as interactive class sessions such as workshops on collaborative research software and networking opportunities. The topic of resilience in research will be interwoven throughout the course. Required reading: Angela Duckworth's *Grit*.

Ashley Barbara Heim
Department of Ecology & Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA. Email: abh229@cornell.edu

PSY 216: New academic hotspots: Opportunities and challenges
Scientists are often attracted by the emergence of important research hotspots and needs. Whether it is appropriate for a given scientist to move into a new field, however, is often a matter of debate. In this course, we will discuss how to evaluate your research interests and capabilities to see if it is worth the effort to move into a new field and under what circumstances you should do so.

Jiazhen Jiang
Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230026, China. Email: markpen@mail.ustc.edu.cn

PSY 279: Critical thinking in physical isolation
This course will bring together the elements of mental focus, networking, task prioritization, and critical thinking to perform and document scientific thinking while in physical isolation. Without physical access to your colleagues or your laboratory, you will hone your skills of solo analysis, hypothesize possible solutions and their associated risks, and become an expert on remote work. Whether you are completing a paper for publication or trying to complete your thesis, this course will enable you to perfect your virtual networking, online scientific resources, and lab work to optimally communicate your science—all from the comfort of your home.

E. Loren Buhle Jr.
DNAnexus, Mountain View, CA 94040, USA. Email: lbuhle@dnanexus.com

Engineering

ENG 205: Apocalypse bunker design
This course covers historical and practical perspectives on building survival shelters. Topics include design and permitting, food preservation, utilities and sanitation, and security. May be taken concurrently with Microbiology 104: Sourdough for novices and Psychology 102: How pets increase happiness. Substantial overlap with Engineering 201: Zombie apocalypse preparation; cannot earn credit for both.

Katie Burnette
Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, Riverside, CA 92521, USA. Email: katie.burnette@ucr.edu

CORONAVIRUS

Protecting the herd with vaccination

How much do COVID-19 vaccines reduce transmission? The answer is a moving target

By **Natalie E. Dean**[1] and **M. Elizabeth Halloran**[2,3]

Based on studies of households in Israel, up-to-date parental vaccination against COVID-19 not only protects the parent but also reduces the infection risk of vaccine-ineligible children.

V accines directly protect the vaccinated but also indirectly protect the unvaccinated and immunocompromised by increasing population-level immunity. On pages 1155 and 1151 of this issue, Hayek *et al.* (*1*) and Prunas *et al.* (*2*), respectively, report measurements of how COVID-19 vaccines indirectly protect members of the same household from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Both studies leverage large health care databases in Israel that are linked to national testing data. Using the close-knit setting of households, these studies isolate how well vaccines prevent infection and how well vaccines reduce infectiousness of those who do get infected. They find reduced infections paired with variable evidence of reduced infectiousness. Studies demonstrating indirect effects increase the value of vaccines and can guide decisions about which vaccine regimens to prioritize to reduce transmission. Yet, these indirect effects are a moving target, reflecting new variants, vaccine waning, boosters, and our evolving daily lives.

How much does parental vaccination with Pfizer/BioNTech BNT162b2 indirectly protect vaccine-ineligible children? Hayek *et al.* address this question by studying two-parent households with at least one vaccine-ineligible child, excluding households with older, vaccine-eligible children. They compare the vaccination status of parents with rates of SARS-CoV-2 infection in children,

[1]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA. [2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [3]Department of Biostatistics, University of Washington, Seattle, WA, USA. Email: nataliedean@emory.edu

adjusting for household characteristics that may influence infection. Between January and March 2021 in Israel, Alpha (B.1.1.7) was the dominant SARS-CoV-2 variant, and individuals aged 16 years and older were vaccine eligible. Children with one fully vaccinated parent were 26% less likely and children with two vaccinated parents were 71.7% less likely to be infected compared with children with neither parent vaccinated. In July through September 2021, Delta (B.1.617.2) was the dominant variant, and individuals aged 12 years and older were vaccine eligible. Third (booster) doses for adults were rolled out during this time. With few unvaccinated adults in their study, Hayek *et al.* instead compared households according to whether parents had received a booster or not. Children with two boosted parents were 58.1% less likely to be infected than children with double-vaccinated parents.

These estimates of indirect protection demonstrate how up-to-date parental vaccination reduces the average infection risk to a child, regardless of where that infection came from—be it from household members, at school, or in the community. The detailed household data allow researchers to dissect the indirect protection (*3*). Parental vaccination reduces the parents' risk of getting infected, making them less likely to expose their children. Hayek *et al.* find 94.4% fewer documented infections in doubly vaccinated parents during the Alpha period and 86.3% fewer infections during the Delta period in boosted parents. Even if a fully vaccinated parent becomes infected, they may spread the virus less if the vaccine reduces their viral load, duration of infection, or virus-spreading symptoms. In households with one infected parent, the odds of at least one child becoming infected during the Alpha period were 72.1% lower if the infected par-

ent was fully vaccinated and 79.6% lower during the Delta period.

Prunas *et al.* apply a different modeling approach to their data, considering all pairs in the household where transmission could occur (child-to-parent, child-to-child, and pairings involving possibly vaccinated older children). To study these pairs, they use a model that captures uncertainty in who was the first person infected. The model also distinguishes between infections likely to come from household members versus those from the community. This study covers the period between June 2020 and July 2021 (before the booster rollout in Israel). Estimated vaccine effectiveness against documented infection before June 2021 (pre-Delta) was initially high at 89.4%, waning to 58.3% after 3 months. After the Delta variant emerged, short-term effectiveness of two doses for preventing infection was 72.0%, waning to 40.2% after 3 months.

Prunas *et al.* find less evidence that a vaccine reduces infectiousness given infection than Hayek *et al.* Using data from the pre-Delta period, they estimate that vaccines reduce infectiousness by 23% but with substantial uncertainty. The estimate is even lower for the Delta period. One contributing difference may be that Hayek *et al.* focus on time periods when vaccination (with second or third doses) was very recent, so the protective benefits may be strongest. Another difference is the type of transmission pairs included. Despite similarities in the underlying datasets, it is not clear how much of the difference in effects is the result of the analytical methods used. Running parallel analyses on both datasets would help resolve these questions.

Measuring indirect effects, such as how parental COVID-19 vaccination protects vaccine-ineligible children, is an important way to quantify the value of a vaccine beyond

protective efficacy (*4*). Although the focus of regulatory agencies is how well vaccines directly protect the vaccinated against disease—as measured in individually randomized trials—benefits for transmission multiply their impact. Estimated indirect effects also inform policy decisions, such as whether groups of vaccinated individuals should wear masks or whether exposed vaccinated individuals need to quarantine.

To understand the impact on transmission, some individually randomized trials measure vaccine efficacy against infection, regardless of symptoms, requiring either frequent testing or specialized antibody testing. Yet, individually randomized trials cannot capture indirect effects or reduction in infectiousness given infection without add-on studies, such as testing participants' household members. In a randomized trial of pertussis (whooping cough) vaccination in Sweden, researchers demonstrated that vaccinating infants decreased whooping cough risk in siblings and parents (*5*). Estimating indirect effects is typically done after a vaccine is licensed, in either observational studies or cluster randomized trials. Clusters larger than households are common for estimating indirect effects. In a trial in Kolkata, India, communities received either a vaccine to prevent typhoid fever or a vaccine targeting another disease (*6*). Unvaccinated members in the typhoid vaccine communities were 44% less likely to develop typhoid fever than unvaccinated members of the comparison communities. Cluster randomized trials have also demonstrated that vaccinating nursing home staff for influenza indirectly protects residents (*7*).

The household studies of Hayek *et al*. and Prunas *et al*. illustrate how indirect protection against SARS-CoV-2 infection is driven by multiple factors, making it a moving target. For example, vaccine effectiveness against infection decreases with time since vaccination and against evasive variants. In January 2022, Omicron became the dominant variant globally, exhibiting higher transmissibility and immune evasion (*8*). As seen in preliminary real-world studies, vaccine effectiveness against Omicron infection is lower than that against previous variants, although a recent booster dose helps recover protection (*9*). Data from a household study in Denmark suggest reduced Omicron infectiousness from infections in vaccinated and boosted individuals (*10*). Yet, large outbreaks among highly vaccinated groups were early warning signs of Omicron's transmissibility (*11*).

Other changes in communities affect indirect protection. Risk of SARS-CoV-2 infection is driven by both household and nonhousehold contacts. Parental vaccination reduces household transmission to children but not transmission from the wider community, such as schools. Compared with when these studies were conducted, individuals may have more-frequent interactions with nonhousehold members, lessening the indirect effects of household vaccination. Yet, indirect protection occurs in community settings too. Children benefit from the vaccination of teachers and, increasingly, of their classmates. Since these studies were conducted, vaccine eligibility has expanded to children 5 to 11 years old in Israel, and vaccines for children 6 months and above may be forthcoming. Alongside these changes, a large Omicron wave has increased population immunity. Key future questions to address include the durability of protection from vaccination alone, Omicron infection, reinfections, and combined vaccination and Omicron infection.

Detailed databases in Israel and research in other countries are critical for ongoing assessment. Goals include quantifying indirect effects during the Omicron wave; measuring the waning of third-dose boosters; and assessing fourth-dose boosters, which Israel deployed to high-risk populations in January 2022. These studies could also evaluate the transmission reduction of next-generation vaccines. It is hoped that intranasal COVID-19 vaccines, some of which are currently in clinical trials (*12*), could confer high protection against severe disease along with higher impacts on transmission if strengthened mucosal immunity can prevent infection.

COVID-19 vaccines have provided exceptional protection against severe disease, and they have reduced transmission. Notably, indirect protection is not all or nothing but rather incrementally increases with each newly immunized person. Ensuring our communities are well vaccinated is a major priority as SARS-CoV-2 becomes endemic. ∎

REFERENCES AND NOTES

1. S. Hayek *et al*., *Science* **375**, 1155 (2022).
2. O. Prunas *et al*., *Science* **375**, 1151 (2022).
3. M. E. Halloran, *Epidemiol. Methods* **1**, 83 (2012).
4. A. Wilder-Smith *et al*., *BMC Med*. **15**, 138 (2017).
5. B. Trollfors *et al*., *Pediatr. Infect. Dis. J.* **17**, 196 (1998).
6. D. Sur *et al*., *N. Engl. J. Med*. **361**, 335 (2009).
7. A. C. Hayward *et al*., *BMJ* **333**, 1241 (2006).
8. R. Viana *et al*., *Nature* 10.1038/s41586-022-04411-y (2022).
9. UK Health Security Agency, "SARS-CoV-2 variants of concern and variants under investigation in England: Technical briefing 34" (2022); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1050236/technical-briefing-34-14-january-2022.pdf.
10. F. P. Lyngse *et al*., medRxiv 10.1101/2021.12.27.21268278 (2021).
11. L. T. Brandal *et al*., *Euro Surveill*. **26**, 2101147 (2021).
12. A. Alu *et al*., *EBioMedicine* **76**, 103841 (2022).

The right shoe for the job

Natural dynein protein motors are reengineered to walk on specific artificial DNA tracks

By **Dhanush Gandavadi**[1] *and* **Rizal F. Hariadi**[1,2]

During the Summer Olympics, sprinters wear running shoes to ensure adequate traction, and in the Winter Olympics, speed skaters use ice skates for their gliding motion. At the molecular scale, protein motors, such as kinesins and dyneins, move along microtubules, whereas myosins walk on actin networks. These motors need to interact with their respective tracks with the right balance of specificity, affinity, and processivity. Their biomolecular tracks are structurally different, just like a running track and a frozen ice rink. This difference raises the question of whether the specificity of molecular motors can be tailored by replacing the part of the protein motor that binds to its track. On page 1159 of this issue, Ibusuki *et al*. (*1*) report the construction of a protein motor that walks on a completely different track by engineering an existing protein motor, dynein, to bind to and walk along synthetic DNA tracks.

Many molecular motors of the cytoskeleton are dimeric proteins that move directionally on a specific track by "race walking"—at least one of the two proteins is bound to the track at all times, until the dimeric motors dissociate from the track. Movement occurs when one protein detaches and subsequently triggers the power stroke. As a result, the motor binds to the next attachment site on the track. In the cellular milieu, groups of protein motors autonomously transport molecules over long distances on these tracks. These motor proteins have been explored as active components for manipulating molecules at the micrometer scale, which could lead to applications in medicine, materials science, and molecular programming (*2*).

During the past several decades, attempts to use protein motors and cytoskeletal tracks for nontrivial tasks have been hindered by the constraints imposed by biological systems that differ from those of microdevices. The biological cytoskeleton

is naturally complex and constantly reorganizes itself. Despite their emergent robustness, these cytoskeletal tracks are not suitable for applications that require reproducibility or exact placement. Another critical challenge is embedding information into the tracks and the motor proteins to program the dynamics of the system, such as guiding motor-protein movement. In general, cytoskeletal tracks do not contain navigation cues. Instead, intricate reaction networks involving subtle interactions of signaling proteins choreograph the movement of motor proteins along their tracks.

To overcome these design challenges, scientists have explored self-assembled DNA structures, such as DNA origami nanostructures (3, 4) and micrometer-long DNA nanotubes (5), as tracks for molecular motors. DNA nanostructures can be designed to form arbitrary structures and patterns and can be arranged on lithographic-patterned microarrays with defined positions and orientations (6). Prior studies have used DNA walkers (7, 8) that move directionally along DNA tracks. However, most DNA walkers destroy their tracks to achieve unidirectional motion because of a lack of power strokes, and most also move several orders of magnitude slower than motor proteins.

Ibusuki et al. converted dynein into motors that could traverse and perform predetermined tasks on synthetic DNA tracks by replacing the microtubule-binding domain of a monomeric human cytoplasmic dynein with 19 types of DNA binding domains that recognize nonpalindromic sequences. The speed of these motors is close to that of cellular motors. The orientation of the recognition sequence on the track (see the figure) ensured unidirectional motion and enabled highly multiplex control of hybrid motors because only motors fitted with the right shoe can walk on a specific track.

Dimeric motors were constructed by linking two hybrid motors with a rigid protein linker. Although the artificial dimer exhibited directional movement on DNA nanotubes, the movement contained a diffusive component and was not highly processive. The run length was 700 nm, which is insufficient for micrometer-scale systems but is comparable to the run lengths of many processive protein motors (9). By linking 15 hybrid motor monomers on artificial DNA scaffolds, the run length increased to 6 μm, analogous to how inter-

¹Biodesign Institute, Arizona State University, Tempe, AZ, USA. ²Department of Physics, Arizona State University, Tempe, AZ, USA. Email: rhariadi@asu.edu

Bespoke shoes for walking DNA tracks

Ibusuki et al. created hybrids of biological motor protein and nonmotor DNA binding domains (light blue and orange) that walk unidirectionally on synthetic DNA tracks containing the recognition sites (dark blue and orange) for the DNA binding domains.

The orientation and position of the recognition sites along DNA tracks determine the walking direction of hybrid motors.
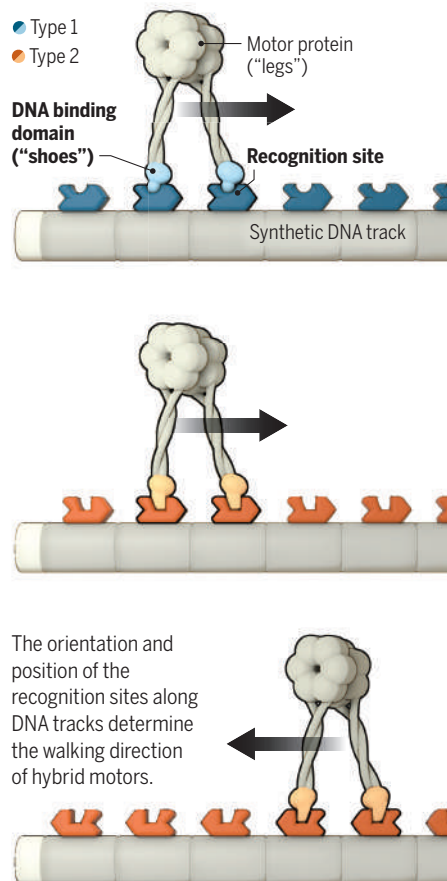
actions between protein motors on a common chassis increase their collective run lengths (9, 10).

Finally, the authors exploited the orthogonality of the DNA binding domain of the hybrid motor by designing four different types of Y-shaped molecular transporters (a disperser, an aggregator, a sorter, and an integrator) by tuning the placement and orientation of recognition sites. In the disperser, Y-shaped tracks have the same recognition sites throughout, and the motors move away from the junction. In the aggregator, the directions of the recognition sites are flipped, so motors move toward the junction. The sorter, composed of two orthogonal recognition sequences that branch out into separate tracks with their respective recognition sequences, sorts different motors with minimal cross-talk. The integrator combines two orthogonal motors onto a single track containing mul-

tiple recognition sites. Collectively, these primitives can function as modules for executing complex molecular robotic tasks in a molecular factory (11).

The apparent modularity of the approach has only been demonstrated for dynein (1, 12). Previous studies argued that dynein is an ideal engine for molecular engineering because of its well-separated modules, such as its catalytic and track-binding domains (12). By contrast, myosins and kinesins are more close packed, and the demarcation between their modules is less distinct than that in dynein. Tracks and environments that these shoes could walk on beyond DNA could be explored. The same strategy can theoretically be extended to the construction of motor proteins that walk on one-dimensional carbon nanotubes or two-dimensional graphene sheets in abiological environments. In an opposite extreme environment, scientists may be able to design new cellular functions by coexpressing hybrid motor constructs and RNA nanostructures in cells.

De novo design of motors and filamentous tracks that rival their natural counterparts has been a challenging route to gaining insights into individual motor proteins and their self-organization. The authors circumvent the challenge by developing chimeras of DNA binding proteins and natural motors that move on specific DNA tracks. The work of Ibusuki et al. brings us closer to understanding the complex cytoskeletal system, which is a network of interconnected modules composed of accessory proteins, motor proteins, and their tracks. In this regard, the reengineering of motors and programmable cytoskeleton tracks could provide a controlled experimental framework for studying other complex, self-organized biomolecular systems. ∎

REFERENCES AND NOTES

1. R. Ibusuki et al., Science 375, 1159 (2022).
2. M. G. L. van den Heuvel, C. Dekker, Science 317, 333 (2007).
3. P. W. K. Rothemund, Nature 440, 297 (2006).
4. K. Lund et al., Nature 465, 206 (2010).
5. P. Yin et al., Science 321, 824 (2008).
6. A. Gopinath et al., Science 371, eabd6179 (2021).
7. W. B. Sherman, N. C. Seeman, Nano Lett. 4, 1203 (2004).
8. J.-S. Shin, N. A. Pierce, J. Am. Chem. Soc. 126, 10834 (2004).
9. R. F. Hariadi, M. Cale, S. Sivaramakrishnan, Proc. Natl. Acad. Sci. U.S.A. 111, 4091 (2014).
10. N. D. Derr et al., Science 338, 662 (2012).
11. N. Stephanopoulos, E. O. P. Solis, G. Stephanopoulos, AIChE J. 51, 1858 (2005).
12. A. Furuta et al., Nat. Nanotechnol. 12, 233 (2017).

GRAPHIC: V. ALTOUNIAN/SCIENCE

Climate change and biospheric output

Large changes in global ecosystem productivity are set in motion by carbon dioxide rise

*By **Corinne Le Quéré** and **Nicolas Mayot***

The response of terrestrial and marine ecosystems to rising carbon dioxide (CO_2) concentrations has serious implications for projections of climate change in the coming decades. Ecosystems store vast amounts of carbon, which, if destabilized, could amplify climate change (*1*). They also provide multiple services to society, from food and shelter to recreation and well-being. Changes in ecosystems and their productivity at the global scale could have fundamental implications for society's future. On page 1145 of this is-

from the regrowth of vegetation after agricultural abandonment and wood harvesting (*4, 5*). By contrast, there is no widespread CO_2 fertilization effect in the oceans. This is because marine ecosystems are primarily limited by nutrients. Although ecosystems do respond to ocean acidification that results from rising CO_2 in ways that are not fully understood (*6*), the ocean storage of carbon is primarily controlled by ocean circulation (*7*). Terrestrial and marine ecosystems are both affected by climate change (*8, 9*), but because it is difficult to quantitatively assess the relative impact of all the processes (*10–12*), there remain large uncertainties in the study of

photosynthesis by the oceanic and terrestrial ecosystems combined. Photosynthesis captures preferentially lighter oxygen isotopes because of their mass, and, therefore, changes in productivity leave a specific signature behind.

Yang *et al.*'s analysis suggests large decreases in productivity during glacial periods, estimated between 55 and 87% of present global GPP. This observation of reduced productivity during glacial periods is for the terrestrial and marine productivity combined. However, the timing of the decrease in global productivity cannot be strictly explained by processes that were operating in the ocean at that time. During glacial periods, some marine photosynthesizers were limited by a reduction in the supply of essential nutrients that was triggered by reduced ocean ventilation. Meanwhile, the increase in atmospheric dust deposition was a large source of iron—a productivity-limiting nutrient—to the ocean, although this was limited to the Southern Ocean and occurred in the later stages of the glaciations (*13*). By using several local records of past marine biosphere productivity, Yang *et al.* confirmed the diversity of responses of the oceanic

CO_2 concentration, ecosystem productivity, and climate interactions

During glaciations, global productivity decreased because of the effect of low carbon dioxide (CO_2) fertilization on terrestrial ecosystems. The ongoing rising CO_2 from human activities fertilizes global ecosystems but has limits, whereas climate change reduces global carbon storage and amplifies warming.
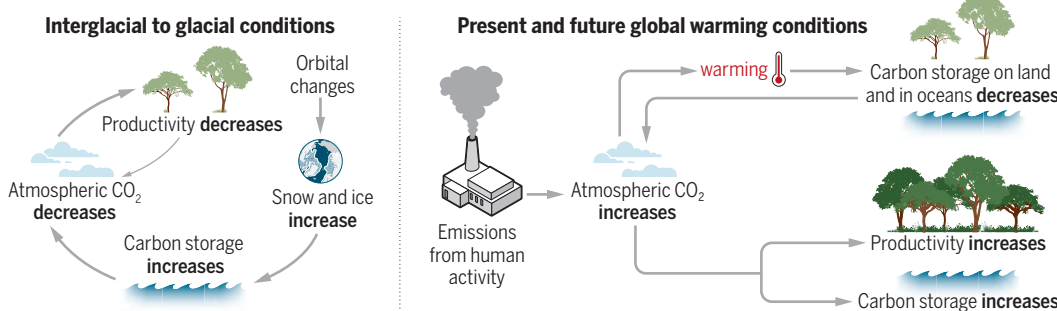
sue, Yang *et al.* (*2*) reconstruct changes in global biosphere productivity during the past eight glaciations over about 800,000 years and provide insights into the sensitivity of global ecosystems to CO_2 concentrations and climate change.

Several processes influence the productivity of terrestrial and marine ecosystems (see the figure). For terrestrial ecosystems, the concentration of CO_2 in the atmosphere is a major limiting factor (*3*). Higher CO_2 concentration directly enhances plant photosynthesis and also makes plants more water efficient as their pores (stomata) need to open less to capture CO_2 for photosynthesis. This effect, known as CO_2 fertilization, is directly linked to changes in carbon storage on land, but it has been challenging to quantify because of the difficulty of separating the effect

how rising CO_2 and climate change will affect the storage of carbon in the land and oceans in the future.

Glacial cycles offer a natural laboratory to study the large-scale response of ecosystems to changing CO_2 concentrations and climate. Much of the data on glacial cycles is based on measurements of air trapped in ice cores and from the analysis of marine sediments. These natural archives provide distinctive opportunities for understanding the processes that are responsible for the complex matter exchanges between the natural environment, including the storage of carbon in the land and ocean, and climate.

Yang *et al.* analyzed the isotopic signature of oxygen in ancient air bubbles trapped in Antarctic ice cores covering eight glacial cycles. They used the oxygen isotopes to reconstruct past changes in global gross primary productivity (GPP), which reflects the amount of atmospheric CO_2 absorbed during

biosphere to the glacial climates, and they exclude the ocean as the main cause of the large decreases in GPP.

The changes in productivity inferred by Yang *et al.* occurred systematically in response to changes in atmospheric CO_2 concentration, pointing to the effect of CO_2 fertilization on the terrestrial biosphere. The coinciding changes in productivity and atmospheric CO_2 between glacial and interglacial periods were also observed within glacial periods owing to the high–temporal resolution index of the past global GPP. During glaciations, CO_2 fertilization acted as negative (damping) feedback because productivity decreased as atmospheric CO_2 concentrations decreased. This led to less uptake of CO_2 by the biosphere and left more CO_2 in the atmosphere, where it dampened the initial effect. Yang *et al.* tested this CO_2 fertilization hypothesis using GPP reconstructed with a process model and showed that the inferred GPP changes

School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. Email: c.lequere@uea.ac.uk

during glaciations are consistent with estimates based on current knowledge of plant physiology and their responses to changes in CO_2 concentrations (*1*, *4*). Moreover, the CO_2 fertilization effect on terrestrial biosphere productivity during glacial periods was supported by pollen records from Europe and southeastern Africa.

Since the beginning of the industrial period, CO_2 emissions from human activities are causing a rise in atmospheric CO_2. The fertilizing effect of CO_2 leads to an increase in terrestrial productivity and carbon storage (*4*), which provides a well-known negative feedback that slows down the rate of climate change (*1*). The analysis of Yang *et al.* confirms the global importance of CO_2 fertilization and provides constraints on the size of its effect. Deforestation, in addition to releasing the carbon stored in existing biomass, also reduces the carbon sink potential from this important CO_2 fertilization effect (*14*).

The comparison of glacial and future changes in productivity has its limits. Glaciations do not provide information on the concentration at which CO_2 fertilization saturates, which depends on the limitation by water and nutrients and on plant physiology. Of greater importance, the growing and pervasive warming of the climate system caused by human activities creates an amplifying feedback as the terrestrial biosphere and the ocean respond to a changing climate (*1*, *4*). Although many uncertainties remain, it is clear that the damping feedback of CO_2 fertilization will decrease in the coming decades while the amplifying feedback of climate on carbon storage continues to grow (*15*). ∎

REFERENCES AND NOTES

1. J. G. Canadell *et al.*, in *Climate Change 2021*: The *Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte *et al.*, Eds. (Cambridge Univ. Press, 2021).
2. J.-W. Yang *et al.*, *Science* **375**, 1145 (2022).
3. P. J. Franks *et al.*, *New Phytol.* **197**, 1077 (2013).
4. P. Friedlingstein *et al.*, Earth System Science Data essd-2021-386 [Preprint] (2021).
5. D. Schimel, B. B. Stephens, J. B. Fisher, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 436 (2015).
6. C. L. Hurd, A. Lenton, B. Tilbrook, P. W. Boyd, *Nat. Clim. Chang.* **8**, 686 (2018).
7. J. Hauck *et al.*, *Front. Mar. Sci.* **7**, 571720 (2020).
8. M. B. Osman *et al.*, *Nature* **569**, 551 (2019).
9. Y. Malhi *et al.*, *Philos. Trans. R. Soc. London Ser. B* **375**, 20190104 (2020).
10. X. Morin *et al.*, *Sci. Rep.* **8**, 5627 (2018).
11. L. Kwiatkowski *et al.*, *Biogeosciences* **17**, 3439 (2020).
12. L. Bopp *et al.*, *Biogeosciences* **10**, 6225 (2013).
13. K. E. Kohfeld, C. Le Quéré, S. P. Harrison, R. F. Anderson, *Science* **308**, 74 (2005).
14. W. A. Obermeier *et al.*, *Earth Syst. Dyn.* **12**, 635 (2021).
15. The Royal Society, "The carbon cycle: Better understanding carbon-climate feedbacks and reducing future risks" (Climate Change: Science and Solutions, Briefing 7, The Royal Society, 2021).

Growing polymers, caught in the act

Researchers show the polymerization of ethylene at the active centers of a catalyst

By **Joost Wintterlin**[1,2]

Polyethylene is a common plastic with a production rate of ~100 million tons every year, which corresponds to about one-third of all plastics (*1*). Most of the polyethylene is created with the help of solid catalysts. These catalysts help link together ethylene molecules, the monomers that make up polyethylene. To better understand the complex reactions that connect the monomers, researchers have mainly focused on the measurements of the formation rate and the composition of the resulting polymer. They have also used spectroscopy to analyze the catalysts, but extracting information about the processes on the individual connection sites, known as active centers, can be challenging. On page 1188 of this issue, Guo *et al.* (*2*) present a way to view the processes on the active centers microscopically by using scanning tunneling microscopy (STM) and a model catalyst.

Industrial polyethylene producers mainly use two solid catalysts: the Ziegler-Natta catalyst, which uses titanium (Ti) (*3*), and the Phillips catalyst, which uses chromium (Cr) (*4*). The Ziegler-Natta catalyst consists of silica-supported, microcrystalline magnesium chloride ($MgCl_2$) particles, the surfaces of which are covered by Ti atoms. During the manufacturing process, this catalyst is activated by adding AlR_3, a compound of aluminum and organic groups R, usually ethyl groups ($-CH_2-CH_3$). Upon activation, the Ziegler-Natta catalyst forms Ti–R units on its surface where the Ti atoms, which are reduced to an oxidation state of +2 or +3, additionally bind to several Cl atoms. The Ti–R units act as the active centers. The Phillips catalyst contains Cr atoms adsorbed on amorphous silica particles. In contrast to the Ziegler-Natta catalyst, the Philips catalyst does not rely on a separate activation agent and is activated directly by the ethylene monomers used for the polymerization. The operating catalyst is believed to contain Cr atoms bound to an organic group R, where the Cr–R units act as active centers. In these units, the Cr atoms are reduced to an oxida-

tion state of +2 or +3 and are additionally bound to several O atoms.

This somewhat simplified picture of the active centers for the two catalysts is the result of hundreds of studies (*5*). In detail, the structure of the active centers as well as the exact oxidation numbers of the metal atoms have remained elusive. This is mainly caused by the active centers not being uniform, a result of the microcrystalline or amorphous morphology of the catalysts, which is reflected by the observation that only a few percent of the centers are actually catalytically active.

The polymerization process is assumed to proceed according to a mechanism proposed for the Ziegler-Natta catalyst by Cossee and Arlman in 1964, shortly after the industrial catalytic synthesis of polyethylene was first introduced (*6*, *7*). The mechanism was put forward as a hypothesis and was entirely based on considerations taken from metal-organic chemistry about the binding of ethylene molecules to transition metal atoms and on considerations about the structure of defects on

Growing monomers into polymers

Researchers use an iron carbide film (Fe_3C) with staggered rows of iron trimers (Fe_3) that act as connection sites—known as active centers—for polyethylene formation on the surface of this model catalyst. As resolved by using a scanning tunneling microscope, chain initiator groups (=CH–CH_3) are observed to bind to these centers, which allow the formation of longer polymer chains by inserting further ethylene molecules.

science.org **SCIENCE**
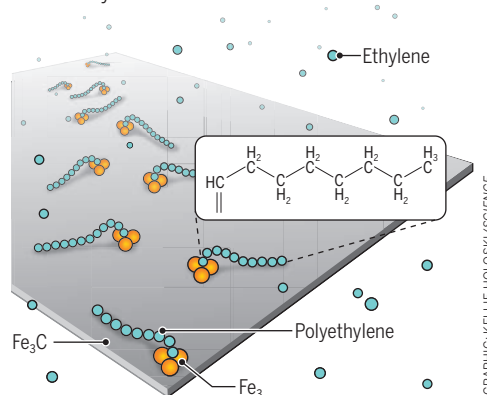
GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

the surface of the MgCl$_2$ particles—the possible bonding sites of the Ti atoms on the Ziegler-Natta catalyst. An analogous mechanism is assumed for the Phillips catalyst. In the first step of this mechanism, an ethylene molecule binds to the Ti or Cr atom of an active center facilitated by the electrons in the ★ orbital of the ethylene molecule. In the second step, the molecule forms a covalent bond to the group R—the so called "initiator group"—and a second covalent bond to the metal atom. The ethylene molecule thus becomes inserted between the metal atom and the group R, which in this way is expanded to a –CH$_2$–CH$_2$–R group. These two steps are repeated many thousands of times, leading to a chain consisting of –CH$_2$–CH$_2$– elements. It is assumed that during the entire growth process, the chain is bound at one end to the same ac-

> **"These observations confirm the Cossee-Arlman mechanism, the central hypothesis about the formation of polyethylene, through direct observation."**

tive center until the molecule is finally released. However, alternative theories have suggested, among other ideas, that a polymer chain may be attached at both ends to an active center (4).

The model catalyst used by Guo *et al.* is a thin iron carbide film on an iron single crystal. The film is flat and crystalline, making it ideal for STM to obtain images with atomic-level resolution. On the iron carbide (Fe$_3$C) film are one-dimensional rows of iron trimers (Fe$_3$) whose structures and charges are different from the Fe atoms on the rest of the film. These Fe$_3$ sites form the active centers of the model catalyst, where ethylene monomers (C$_2$H$_4$) can bind to and connect to polyethylene chains (see the figure).

The STM data of Guo *et al.* provide insight into the molecular processes through microscopic imaging, which resolved almost the entire sequence of processes. In the experiments, the sample is kept in a vacuum system, so that it can be exposed to ethylene in a controlled way. Upon dosing the Fe$_3$C film with ethylene, the ethylene molecules are first adsorbed as intact CH$_2$=CH$_2$ units that are highly mobile on the Fe$_3$C film, before binding to the active

centers and becoming immobilized. During this process, the ethylene molecules transform into initiator groups, most likely as –CH–CH$_3$ units. This step is equivalent to the formation of the initiator groups by ethylene on the Phillips catalyst. Upon further ethylene dosing, short polymers, known as oligomers, begin to grow, and chains consisting of up to 16 to 20 carbon atoms are observed. During growth, the chains remain attached to the active centers at one end, whereas the other end is mobile. These observations confirm the Cossee-Arlman mechanism, the central hypothesis about the formation of polyethylene, through direct observation. Alternative mechanisms with ring-like configurations, with both ends of the chains attached to the active center, can be ruled out.

What the authors propose is a variation of the Cossee-Arlman mechanism. According to their proposal, the first element of each chain is a CH rather than a CH$_2$ group, so the oligomers bind with nominal double bonds to the active centers. Such a variation has been discussed in the literature but was found less likely because it requires a shift of an H atom along the chain each time a monomer is inserted (4). However, it appears that the STM data would also be consistent with the classical Cossee-Arlman mechanism, which assumes a CH$_2$ group at the active center. It would be worth extending such STM experiments to flat models that more closely resemble the compositions of the industrial catalysts. Such models exist and have been investigated with spectroscopic methods (8, 9).

These molecular processes at the active centers are not only of academic interest; they can have important ramifications for industrial applications. These processes determine the average lengths of the molecules, the distributions of the lengths, and when molecules with side groups are added during growth, the formation of branches. In this way, these processes determine the physical properties and the quality of the synthesized polymer. ∎

REFERENCES AND NOTES

1. J. O. Bühler-Vidal, in *Handbook of Industrial Polyethylene and Technology*, M. A. Spalding, A. M. Chatterjee, Eds. (Wiley, 2017), pp. 1297–1330.
2. W. Guo *et al.*, *Science* **375**, 1188 (2022).
3. J. P. Claverie, F. Schaper, *MRS Bull.* **38**, 213 (2013).
4. M. P. McDaniel, in *Advances in Catalysios*, B. C. Gates, H. Knözinger, Eds. (Academic Press, 2010), vol. 53, 123–606.
5. E. Groppo *et al.*, *Catal. Sci. Technol.* **3**, 858 (2013).
6. P. Cossee, *J. Catal.* **3**, 80 (1964).
7. E. J. Arlman, *J. Catal.* **3**, 89 (1964).
8. E. Magni, G. A. Somorjai, *Surf. Sci.* **377-379**, 824 (1997).
9. T. Risse, J. Schmidt, H. Hamann, H.-J. Freund, *Angew. Chem. Int. Ed.* **41**, 1517 (2002).

[1]Faculty of Chemistry and Pharmacy, Ludwig-Maximilians-Universität München, 81377 Munich, Germany. [2]Center for Nanoscience, 80799 Munich, Germany. Email: wintterlin@cup.uni-muenchen.de

The modular cell gets connected

Integrative molecular cell biology can be used to interpret networks beyond modules

By **Stephen W. Michnick**[1] *and* **Emmanuel D. Levy**[2]

To understand living cells and the transfers of mass, energy, and information underlying living processes, spatiotemporal relationships among networks of genes, their product RNA, and protein molecules need to be defined. It is these spatiotemporal relationships that will allow us to grasp how variations in the genome manifest cellular characteristics and how cells interact with their environments. There have been enormous efforts to bridge the abstract structures of biomolecular networks and the spatiotemporal relationships of their component molecules. On page 1143 of this issue, Cho *et al.* (1) describe "OpenCell," a data resource and analysis roadmap that takes us closer to this aim. They provide a glimpse of the complex and surprising spatial organization of a living human cell, including the existence of a cellular space in which new functions evolve.

Cho *et al.* engineered ~1300 genes in human HEK293T cells using CRISPR-Cas9 to integrate DNA that encodes a green fluorescent protein (GFP) fragment next to the genes. The same protein fusions and cell lines were used to map out two types of intracellular networks: a physical protein-protein interaction network (interactome) through affinity purification-mass spectrometry and a protein subcellular localization similarity network through imaging with confocal fluorescence microscopy. For the latter, a self-supervised neural network analyzes the subcellular localization data to reduce the high-dimensional patterns into low-dimensional encodings. The

[1]Département de biochimie, Université de Montréal, Montréal, Québec, Canada. [2]Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel. Email: stephen.michnick@umontreal.ca; emmanuel.levy@weizmann.ac.il

authors compare the encodings of protein pairs to compute relationships between their localizations and reveal a continuous space that encompasses molecular and cellular scales.

In this continuous space, Cho *et al*. find that the spatial distributions of proteins partitioned into three global domains. Notably, these three domains were mirrored in the physical interaction network. The first two groups of protein-protein interactions correspond to well-understood soluble and membrane-associated proteins, but the third is unforeseen: an RNA binding protein group, which suggests that RNA itself may have more to do with organizing the cell than ever suspected. This concept is supported by the recent finding of hundreds of nuclear territories with high concentrations of noncoding RNAs (*2*) and by RNA being an active player in the formation of biomolecular condensates (*3*).

> "…a glimpse of the complex and surprising spatial organization of a living human cell, including a space where new functions evolve."

Ideas of spatial and functional modularity of protein networks were postulated long ago and were observed in protein interactomes (*4*, *5*). But do these modules communicate with each other, and if so, how? A key variable in defining how proteins are organized relative to one another is the stoichiometry of their interactions (*6*, *7*). For a pair of proteins, the stoichiometry is the ratio of their cellular concentrations, which are typically similar among stable complexes (stoichiometry ~ 1). Cho *et al*. identify two modes in the distribution of stoichiometry values: a narrow mode of protein pairs with concordant stoichiometry (around 1:1) and a larger mode with discordant stoichiometry (around 1:100) (see the figure). Protein pairs such as proteasome subunit α type-2 (PSMA2) and PSMA4 show concordant stoichiometries and similar localization patterns. At the same time, these two proteins interact with proteasome-assembly chaperone 3 (PSMG3), which exhibits a sevenfold lower abundance and distinct localization pattern. The dominance of such discordant pairs is observed both at the level of cellular concentrations and localizations, highlighting that only a small part of the protein interactome conforms to the concept of spatial and functional modularity.

Others have described the low stoichiometry region as "glue" that holds the interactome together (*7*), and now Cho *et al*. extend this concept to this region that connects (through connector interactions) cellular compartments and orchestrates communication between a myriad of cellular processes. It is also in this low stoichiometry space that new and sometimes aberrant protein-protein interactions can evolve, such as those that occur with pathological fusion proteins associated with cancers or interactions that change properties of mesoscale assemblies (*8*, *9*).

A low-stoichiometry pair is also a signature of nonfunctional interactions. These are interactions that arise at random and that, if disrupted, do not affect cellular fitness (*10*). These interactions are also a vast reservoir for innovation, in which a once nonfunctional interaction may serve a new function under new conditions (*11*, *12*). The acquisition of new or enhanced stoichiometry interactions can occur with single-point mutations (*13*), and therefore, many such new interactions are likely to form during the course of evolution. Indeed, Cho *et al*. have discovered a fourth supercompartment, the "evolvosome." In essence, this domain is to protein interaction networks what intrinsically disordered regions are to protein structures.

Cho *et al*. chart a path to systematic and quantitative human cell biology, yet much remains to be done. The OpenCell collection of cells only covers about 7% of protein-coding genes. There may be many new cellular territories and structures to be discovered. The identified features of protein localization reflect physical principles of how biomolecules are organized in cells, which will need exploration. For example, the enrichment for proteins with intrinsically disordered regions in the RNA compartment points directly to the possibility that organization of RNA-mediated subcompartments may involve liquid-liquid phase separation. Other assembly principles may exist. The most exciting prospective applications of the OpenCell collections will be the spatiotemporal analyses of responses of cells to environmental perturbations and genetic variation, as has been documented in yeast (*14*). This work and other studies (*15*) that integrate scales of spatiotemporal relationships between molecules will provide insight into how matter, energy, and information flow through living cells and may reveal mechanisms for drugs or mutations in cellular processes. These conceptual advances will also provide design principles for creating synthetic cells with desired traits. ∎

Most protein-protein interactions are not in modules

The density distribution (left) of protein pairs according to stoichiometry and localization similarity, from Cho *et al*. (*1*), reveals a minority of module-like interactors exhibiting concordant stoichiometry and localization (gray). The schematic network (right) shows modules such as the proteasome, which includes the subunits PSMA2 and PSMA4, that are wired to the connector, proteasome-assembly chaperone 3 (PSMG3). The dominance of connectors stresses the need for new concepts on modes of biomolecular organization to understand networks.

Stoichiometry and localization
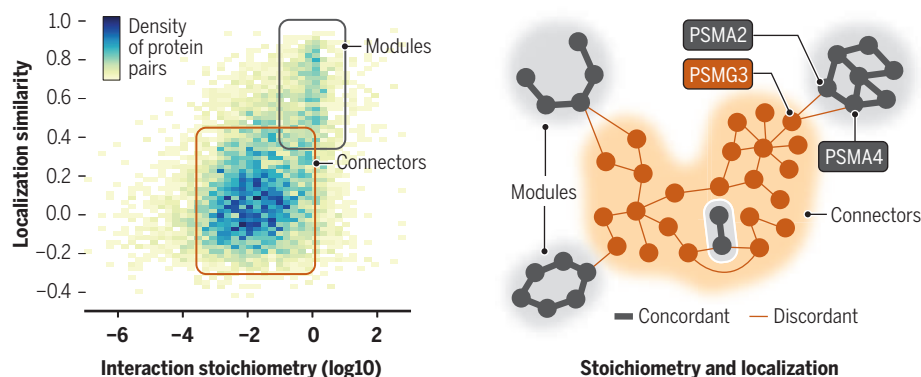
REFERENCES AND NOTES

1. N. H. Cho *et al., Science* **375**, eabi6983 (2022).
2. S. A. Quinodoz *et al., Cell* **184**, 5775 (2021).
3. D. W. Sanders *et al., Cell* **181**, 306 (2020).
4. L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, *Nature* **402**, C47 (1999).
5. A. C. Gavin *et al., Nature* **440**, 631 (2006).
6. O. Matalon, A. Horovitz, E. D. Levy, *Curr. Opin. Struct. Biol.* **26**, 113 (2014).
7. M. Y. Hein *et al., Cell* **163**, 712 (2015).
8. T. Mittag, A. Z. Ansari, *Nat. Struct. Mol. Biol.* **28**, 543 (2021).
9. S. Alberti, A. A. Hyman, *Nat. Rev. Mol. Cell Biol.* **22**, 196 (2021).
10. E. D. Levy, C. R. Landry, S. W. Michnick, *Sci. Signal.* **2**, e11 (2009).
11. J. Kuriyan, D. Eisenberg, *Nature* **450**, 983 (2007).
12. C. R. Landry *et al., Cell* **155**, 983 (2013).
13. H. Garcia Seisdedos *et al., Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.2101117119 (2022).
14. Y. T. Chong *et al., Cell* **161**, 1413 (2015).
15. Y. Qin *et al., Nature* **600**, 536 (2021).

10.1126/science.abo2360

Mildred Dresselhaus works in her office at MIT in November 2006.

BOOKS *et al.*

SCIENCE LIVES

All hail the Queen of Carbon

A new biography solidifies the legacy of a pioneering and prolific nanoscientist

By **Vijaysree Venkatraman**

"Dr. Mildred S. Dresselhaus, of Lincoln Laboratory, who has achieved prominence as a solid-state physicist, has been appointed Abby Rockefeller Mauzé Visiting Professor at the Massachusetts Institute of Technology," read a small news item in the *Boston Globe* on 8 October 1967. The appointment was for a single year, but it was still notable because it was the first time MIT's electrical engineering department had hired a female professor. Before that year was out, the department would offer Dresselhaus a full professorship with tenure.

In *Carbon Queen*, the first full-length biography of the history-making scientist, science writer Maia Weinstock traces Dresselhaus's exceptional career, which included foundational research on various forms of carbon that has enabled other scientists and engineers to make tremendous advances with nanoscale structures "on the order of one-hundred-thousandth the width of a human hair." Such materials—which include cylindrical nanotubes, iconic buckyballs, and two-dimensional graphene—have applications in energy storage, medical research, and quantum computers.

It all began with graphite—the familiar "lead" of pencils, which is neither a metal nor a semiconductor but, nevertheless, conducts electricity. Dresselhaus's investigations into the electronic structure of graphite contradicted the theorized energy level spacing within the material's valence and conduction bands and led to a standout publication in 1968, her first year at MIT (*1*). In her six-decade-long career, she would publish ~1700 peer-reviewed articles as well as eight books on the fundamental properties of carbon, earning her the moniker "Queen of Carbon," a title she apparently hated at first but later embraced.

Weinstock writes evocatively of Dresselhaus's improbable journey from a tough neighborhood in New York City in the 1940s to a world-renowned research institution. In the Depression era, we learn, her family struggled to make ends meet. Fortunately, the musically talented "Millie"—as she was widely known—won free lessons at the Greenwich House Music School. There, she learned of a nearby school for girls, where, after passing the difficult entrance examination, she excelled in mathematics and science. Later, at Hunter College, Dresselhaus was preparing to become a math teacher until Rosalyn Yalow, who would go on to win the Nobel Prize in Physiology or Medicine in 1977 and briefly taught nuclear physics at Hunter, insisted that her protégé apply to graduate school.

At the University of Chicago, where Millie completed her graduate studies and briefly enjoyed the informal mentorship of Nobelist Enrico Fermi, she met and married postdoc Gene Dresselhaus, a rising star in theoretical physics. Gene would become his wife's lifelong cheerleader and collaborator (he introduced her to graphite), and together they raised four children. He celebrated his wife's accomplishments, but Millie regretted the fact that she got far more credit than her husband for work they had done together, notes Weinstock.

Only the Nobel Prize, which was awarded to other researchers for discovering buckyballs and carbon nanotubes, respectively, eluded Millie. "In both cases, they had ideas I missed, and they did great work," she told the *New York Times*.

Keenly aware that a philanthropic grant made it possible for her to enter academia, Dresselhaus used her prominence to further the same vision. Not content with being a distant role model, she regularly met with young women on campus and sought to increase their numbers in technical courses. Her research into the undergraduate admissions process revealed that "it was harder for women to get into MIT than for men," and over the years she made useful recommendations to level the playing field for women in science at MIT and elsewhere.

In *Carbon Queen*, Weinstock has pieced together Dresselhaus's story using decades of profiles, print interviews, oral histories conducted with the scientist herself, and new interviews with her contemporaries. Using lively metaphors, she makes complex scientific concepts accessible, comparing, for instance, the bandgap—which determines whether a material can conduct electricity—to a bouncer at a popular nightclub. Readers are also left with vivid images of the woman herself, as a child on her way to music school; as a high-spirited teen, sneaking friends into the Hayden Planetarium; and finally as a trailblazing scientist who, politely but always with great effect, gave the academy hell for its dismal track record with women. ∎

Carbon Queen: The Remarkable Life of Nanoscience Pioneer Mildred Dresselhaus
Maia Weinstock
MIT Press, 2022. 320 pp.

REFERENCES AND NOTES
1. P. R. Schroeder, M. S. Dresselhaus, A. Javan, *Phys. Rev. Lett.* **20**, 1292 (1968).

The reviewer is a science journalist in Boston, MA 02144, USA. Email: v.vijaysree@gmail.com

ANTHROPOLOGY

Sanctifying work in Silicon Valley

What is lost when spiritual practices are secularized for the workplace?

By **J. A. English-Lueck**

Patrick Sullivan is an ambitious executive at a tech firm that was recently acquired by a multinational corporation. When his team struggled to adapt to changing company culture, Sullivan enlisted the help of an executive coach and Buddhist dharma teacher, who—with the blessings of upper management and the human resources department—developed a Buddhist-inspired program for workers that incorporated meditation, mindfulness, and self-examination exercises designed to increase job satisfaction and allow employees to connect to their "authentic selves" at work. The program was an unequivocal success, from the company's perspective. But what happens when spiritual practices move from faith-based institutions to the workplace? Sociologist Carolyn Chen explores this question vividly in her new book, *Work Pray Code*, providing readers with a rich ethnography of Silicon Valley's elite alongside the stories of those who service this workforce's souls.

The reviewer is at the Department of Anthropology, San José State University, San José, CA, USA; is a distinguished fellow at the Institute for the Future, Palo Alto, CA, USA; and is the author of *Cultures@SiliconValley* (Stanford Univ. Press, 2017) and *Being and Well-Being: Health and the Working Bodies of Silicon Valley* (Stanford Univ. Press, 2010). Email: jan.english-lueck@sjsu.edu

Chen collected her observations between 2013 and 2019, a time of corporate reinvention in Silicon Valley. During this period, competition for elite talent was escalating, and the demands of work were intensifying. Perks, at least for elite workers, proliferated, and meals, massages, and errand runners kept employees anchored to the workplace.

The corporate culture of this era did little to encourage traditional religiosity, but companies soon began secularizing spiritual practices, couching them in terms such as "self-discipline." With these tools, they promised, practitioners would be better able to control their emotions and sustain concentration. Spiritual practices, in other words, could be used to cultivate human capital.

But what changes when corporate actors incorporate spiritual disciplines, stripped of their original contexts, into temples of work? Not surprisingly, the meaning of many terms morphs into something less religious. "Sangha," for example, becomes simply "community," not a community of Buddhist believers. And although Asian workers make up a considerable slice of high-tech workers, the forms of religiosity borrowed from their cultures are often secularized and reimagined in order to be made accessible to white employees.

The juxtaposition of religious sentiment and high-tech work in Silicon Valley is not as unusual as one might suppose. In their book *Seeing Silicon Valley*, Mary Beth Meehan and Fred Turner suggest that the work ethic of the elite of Silicon Valley is not so different from the Protestant spiritual mission of colonial New England, which posited that hard work is the path to building a "better world." That same sentiment, Chen notes, permeates the rhetoric of high-tech companies, especially start-ups, in Silicon Valley. When bringing Buddhist luminaries to corporate campuses or advocating regular mindfulness practices increases productivity, such practices conveniently align with the company's mission. But there are inherent contradictions in employing Buddhist practices to advance material growth.

Chen sprinkles her book with compelling stories of workers who live with these tensions. Some, such as Cecelia Lau, feel compelled to justify meditation with studies that demonstrate that such programs increase productivity. While she struggles with the idea that this corrupts the teachings, she recognizes the need to drop the more overtly spiritual features to attract more participants.

While *Work Pray Code* centers on the corporate mindfulness initiatives for elite tech workers, Chen also takes moments to explore the impacts of these programs on larger issues. Early in the book, she demonstrates that developing a meaningful, ethical corporate culture, one aligned with workers' personal values, revitalizes employees and sustains them as they experience the inevitable economic and emotional fluctuations of high-tech work. But when the ethical components of Buddhist teachings are dropped, workplaces lose an opportunity to become authentic sources of meaning to their workers and the communities around them.

Strategic conversations that invoke ethical principles could drive innovation. For example, good stewardship could translate into research to mitigate environmental damage. Cultivating compassion could drive empathy and improve design. Chen thoughtfully considers a wide range of possibilities.

Corporate leaders and technologists may learn to control their attention and emotions with secularized spiritual practices, but not how to become better humans. It is this component that is critical to developing technologies with heart. ∎

In the absence of ethical context, spiritual practices such as meditation are devoid of greater meaning.

A TIME TO REFLECT

Lessons from 2 years of the COVID-19 pandemic

By **Caroline Ash**, **Gemma Alderton**, **Priscilla Kelly**, **Seth Thomas Scanlon**, **Valda Vinson**, *and* **Brad Wible**

Two years ago, on 11 March 2020, the World Health Organization declared COVID-19 to be a pandemic, caused by a virus new to humankind. As 2020 unfolded, events moved catastrophically fast. Epidemiologists showed us that asymptomatic transmission permits severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to escape scrutiny and that age makes people vulnerable. It was clear that national borders would not contain this virus, however much governments wished otherwise.

Even so, the virus sequence was swiftly solved, and the spike protein structure emerged. With this information, scientists worked day and night to develop vaccines and therapeutics. Astonishingly, several vaccines were developed and billions of doses were distributed in less than a year.

So much has been achieved, even without a full understanding of immune responses to natural infections and vaccines. Although we do not know why some people succumb to acute disease and Long Covid, our focus on the spike protein has paid off. Despite some work on viral replication and processing, other virus protein properties are largely mysterious. So far, spike remains the virus's greatest weakness, even though it has undergone a rapid series of structural changes.

Our greatest failing is inequitable distribution of vaccines, therapeutics, and their technologies. We also cannot escape the realization that misinformation and misgovernance during the pandemic killed people as surely as global war. We must reflect: How do we translate good science into good policy and benign understanding? How do we ensure global disease surveillance and universal health care? We must learn to bridge these gaps to avoid millions of fatalities and countless families torn apart in the next pandemic.

ILLUSTRATION: STEPHAN SCHMITZ/FOLIO ART

Workers take samples from a sewer in Nice, France, in June 2021.

SIGNALS FROM THE
SEWER

Measuring virus levels in wastewater can help track the pandemic. But how useful is that?

By **Gretchen Vogel**

n March 2020, the Austrian ski town of Ischgl—known for 239 kilometers of uninterrupted runs and an exuberant après-ski scene—suddenly became infamous as the site of the one of the first COVID-19 superspreading events. Hundreds of infected skiers took the virus home and seeded outbreaks all around Europe.

As the pandemic progressed, however, Ischgl was on the vanguard for a more positive reason: Health officials and scientists in the state of Tyrol were among the first to monitor levels of the pandemic coronavirus in sewage—and base health policy decisions on them. Because the region is so dependent on tourism, officials were eager to know whether the virus was truly on the decline so they could lift key restrictions. They also wanted to catch the earliest possible signals that it might be coming back. Wastewater analysis, which picks up fragments of virus shed in feces, was invaluable, says Stefan Wildt, a wastewater expert at the state's department of water management. Following Tyrol's lead, a national program has recently expanded to cover more than half of Austria's population.

Although wastewater monitoring has been used to track polio and other pathogens for decades, the COVID-19 pandemic has led to an explosion of interest. The technique takes advantage of the fact that SARS-CoV-2 replicates in the digestive system and is shed in high quantities, often before symptoms appear. (The virus is also detected in urine, though not as consistently.) That provides an inexpensive way to monitor infections in thousands or even millions of people without pesky nose or throat swabs, or to predict where cases might be about to surge and hospitals risk getting overburdened. The genetic sequences of the shed virus can also provide hints about how it is evolving.

Scientists in the Netherlands, which has had a nationwide network of wastewater monitoring for decades, were among the first to show fragments of SARS-CoV-2 virus in wastewater samples could accu-

A sewage treatment facility in Austria's Tyrol state, which launched an ambitious wastewater monitoring program early in the COVID-19 pandemic.

rately reflect its levels in the community (see graphic, p. 1103). Since then, monitoring projects for SARS-CoV-2 have sprung up in at least 58 countries, according to a dashboard set up by Colleen Naughton and colleagues at the University of California (UC), Merced. The European Union recommended all member countries establish monitoring systems for SARS-CoV-2 by October 2021, and 26 of 27 have complied, says Bernd Manfred Gawlik, who is helping coordinate efforts through the European Commission. In the United States, the National Wastewater Surveillance System includes 400 sites in 19 states. Last month, the U.S. Centers for Disease Control and Prevention added a national dashboard of wastewater data, and on 2 March, President Joe Biden's administration said the monitoring system will be part of the effort to detect new variants. In India, a successful project in Bengaluru is expanding to half a dozen new cities.

Still, the jury is out on just how useful the technology is. Reliably determining viral levels in wastewater has posed logistical and technical challenges, and interpreting the data can be difficult. (For one, a good downpour will send virus concentrations in sewers plummeting.) Establishing collection, testing, and reporting systems can be time consuming and expensive as well. And although policymakers have welcomed the results of wastewater monitoring, few have used them to take action; typically, they have waited for cases to rise and intensive care units to fill up.

Shelesh Agrawal of the Technical University of Darmstadt, who has been analyzing water samples from sites across Germany since 2020, says it has been a struggle to convince policymakers the data are useful. "We are the information delivery guys. We deliver to your doorstep, but can't make you eat." Even in the Netherlands, which boasts one of the world's most sophisticated monitoring systems, researchers acknowledge it has had little impact on national policies. Local officials have made use of the Dutch data, however—for example, by ramping up testing in neighbor-

Waste flowing through the century-old sewers of St. Louis yielded unusual SARS-CoV-2 sequences.

Mysterious strains show power—
and limitations—of wastewater monitoring

In March 2021, a team led by virologist Marc Johnson of the University of Missouri, Columbia, made an unsettling discovery in the sewers of St. Louis: a version of SARS-CoV-2 that no one had ever seen before, with half a dozen mutations in the protein that helps the virus enter cells. It was unclear where the virus came from, how it had evolved, or whether it might spark a new wave in the pandemic. "I didn't sleep much that month. I was obsessed with it," Johnson recalls. "I couldn't figure out what was going on."

One year later, Johnson is still trying to solve the riddle, together with John Dennehy, an expert in viral evolution and ecology at the City University of New York who has found similar strains in wastewater from New York City. Now, they're hoping RNA samples from half a dozen countries, sent to them by more than 30 researchers after a February call for help on Twitter, will provide answers.

The finding is an example of the power and limitations of wastewater monitoring, a SARS-CoV-2 surveillance technique that is burgeoning worldwide (see main story, p. 1100). In search of new variants, both teams extracted and sequenced RNA coding for SARS-CoV-2's receptor-binding domain, a rapidly mutating part of the spike protein that helps the virus lock onto and infect cells. A few weeks after his discovery, Johnson heard Dennehy describe his own finding on a virology podcast and got in touch. "We realized both of us were seeing weird stuff, and that it was possibly related," Dennehy says.

The unusual sequence in St. Louis suddenly disappeared from the samples after 6 weeks. "We've never seen it again," Johnson says. But the New York sequences, seen in three of the city's 14 main sewersheds, are still showing up, though they still haven't been identified in human patients. Finding their source is not straightforward. Water in the sewersheds contains the feces of more than 4.4 million people in all of New York City's five boroughs.

Dennehy and Johnson found the wastewater also contains RNA from dogs, cats, and other animals, all potential hosts for the variant. The researchers ruled out three of them—cows, sheep, and pigs—as potential hosts, but Johnson says rats are a likely suspect. Yet there's very little rat RNA in one of the sewersheds with a persistent cryptic strain.

"Nothing makes perfect sense" with the data so far, Dennehy says. He thinks the strains are most likely coming from immunocompromised patients. "It could be people who are getting chronically infected, who can't eliminate the virus, but who do have some sort of immune response" forcing the virus to evolve, Dennehy says. Johnson agrees: "It looks exactly like it's running from the immune system." So far, the variants aren't successful at spreading; they have mostly stayed put in the watersheds where they were discovered.

Because wastewater monitoring only yields virus fragments, not full genomes, figuring out where the cryptic strains might fit in the evolutionary tree of SARS-CoV-2 is difficult. They all lack a key mutation that emerged in summer 2020, suggesting they derived from the strain that emerged in Wuhan, China, and caused the pandemic's first wave. But they have genetic changes at many of the same places as the Omicron variant, suggesting a case of convergent evolution. "They seem to have been subjected to the same selective pressures," Dennehy says.

The researchers hope the RNA samples they are now receiving from colleagues will provide more clues. Already, similar strains have shown up in California and Wisconsin. "It makes me somewhat concerned that if we can't identify the source of the cryptic variant that it's continuing to evolve," Dennehy says. "If it manages to acquire some kind of combination of genetic changes that makes it more transmissible, we could be looking at the origin of Pi"—the Greek letter after which the next variant of concern will be named. —*G.V.*

hoods where wastewater suggested cases were being missed.

As the pandemic shifts, however, wastewater could start to play a bigger role in shaping policy. Many countries are not only lifting pandemic restrictions, but abandoning widespread testing of the population, and more people now rely on home tests that aren't reported in official statistics. That makes wastewater a key remaining tool to understand the course of the pandemic, says Heather Bischel, a wastewater expert at UC Davis: "It provides a bigger picture snapshot."

IN THEORY, WASTEWATER testing is straightforward. Like standard clinical tests, it uses a polymerase chain reaction (PCR) assay to search for specific snippets of viral RNA in a sample, which are then copied repeatedly to amplify the signal. The number of cycles, or rounds of copying, needed to detect a signal in a sample is a rough measure of how much virus is there.

But whereas a throat or nose swab contains about the same amount of material from person to person, wastewater samples contain different amounts of feces, depending on the day and time a sample is taken, recent rainfall, and whether the toilets upstream are in homes, offices, or other buildings. All such variables have to be factored in to be able to accurately "read" a sample. How the water is collected, stored, and processed also affects the results. All of these variables make it very difficult to compare data from different sites.

"Some of the stories you read make it sound like you scoop some water out, dip a test stick in, and get your answer," says Hannah Safford, a former student in Bischel's lab at UC Davis and a policy expert at the Federation of American Scientists. "But it's so much harder than that."

Trial and error has helped scientists and technicians refine their techniques during the pandemic. Multiple groups have tested the best way to concentrate samples, comparing, for example, centrifuge times and filtration techniques; they have also identified reference viruses common in wastewater that can help calibrate samples. One unlikely sounding but popular reference is the pepper mild mottle virus (PMMoV). Harmless to humans, it attacks spicy and bell peppers and is ubiquitous in wastewater, passing through the digestive system when we eat infected produce. Because the concentration in human feces stays relatively stable year-round, scientists use it as a proxy for the amount of feces in a sample, reporting results as a ratio of PMMoV levels to SARS-CoV-2 levels.

In Sweden, Zeynep Cetecioglu Gurol and her colleagues at the KTH Royal Institute of Technology discovered that freezing samples, necessary for a while because PCR reagents were in short supply, made virus levels plummet. In Kansas City, Missouri, researchers were puzzled that the viral level in wastewater appeared to increase in a region where clinical cases seemed stable. They found out that repair work had diverted millions of liters of additional wastewater into the sewers from outlying suburbs, skewing their calculations.

Angela Chaudhuri, a public health expert at Swasti Health Catalyst who helped coordinate Bengaluru's monitoring project, says her team had to figure out how levels in open drain systems, common in Bengaluru, compare with closed sewers.

Increasingly, scientists are making their results available directly to the public. The Bengaluru project, for example, has had an online dashboard since May 2021 that shows where virus levels are increasing, decreasing, or remaining stable. "There is mistrust of the government when it comes to COVID," Chaudhuri says. "People think they might be hiding cases. This dashboard gives an independent signal" that can corroborate the official numbers. Marc

Johnson, a virologist at the University of Missouri, Columbia, who has helped lead the state's extensive wastewater monitoring, says media and the public have welcomed the data. When the dashboard couldn't be updated on Christmas Eve, he says, the team spoke with journalists eager for information about the appearance of the Omicron variant in sewersheds. "It gives people peace of mind. If you feel like you know what is going on, you have more control," Johnson says.

In addition to tracking the spread of variants, wastewater monitoring has also enabled Johnson and others to spot odd SARS-CoV-2 strains in St. Louis and New York City, never seen in patients, that pose an epidemiological and evolutionary mystery (see sidebar, p. 1102).

SCIENTISTS SAY perhaps the clearest case for wastewater monitoring is in places where there is little virus at all. Australia and New Zealand, for example, used wastewater monitoring as a key part of their zero COVID strategy: As soon as a positive sample appeared, health officials ramped up testing, alerted the public, and, if active cases were confirmed, swiftly imposed restrictions to nip the outbreak in the bud.

Tracking the pandemic's waves

In the Netherlands, which boasts one of the world's most sophisticated monitoring systems, SARS-CoV-2 levels in wastewater have correlated fairly closely with reported COVID-19 cases throughout the pandemic.

(Although monitoring continues, both countries have recently given up on zero COVID.) A similar approach has worked on a small scale as well: In fall of 2020, wastewater testing on U.S. college campuses identified infections in residence halls before anyone had tested positive. Testing all the residents enabled health authorities to identify the infected person before the virus spread.

But in regions where cases are high and individual testing common, wastewater may not add significant information. "If you have

testing capacity or people's reduced willingness to get tested. And as the pandemic begins to fade into the background in many places, the role of data from wastewater monitoring is likely to grow.

To make the data as reliable as possible, researchers are continuing to improve and standardize their techniques. Especially promising, says Gertjan Medema, an expert on pathogens in wastewater at the KWR Water Research Institute in the Netherlands, are new collecting devices consisting of a container housing mag-

identify outbreaks early and help identify the pathogens at work. Other researchers are adding tests to keep an eye on influenza, rotavirus, norovirus, adenoviruses, respiratory syncytial virus, and antibiotic-resistant bacteria.

Researchers would also like to sequence more of the viruses they catch, because standard PCR assays usually can't distinguish between variants. New York City's program, for example, has discovered Omicron in samples taken on 21 November 2021, several days before scientists in South Africa and Botswana announced they had identified the variant. But the New York City team didn't recognize its find until a week later, after they had sequenced the sample—and knew what to look for.

Sequencing wastewater samples is a challenge because the virus particles are usually degraded and they can come from hundreds or thousands of sources, Johnson says: "You can't figure out which ones fit together" in a whole genome. And because of the way sequencing software works, the most interesting parts of the genome—the regions that change most frequently—can easily be missed.

One way around that is to tailor the sequencing process to capture those fast-changing regions. That's how Johnson's team and another one in New York City identified the mysterious strains: They designed sequencing primers to match the beginning and end of the viral gene that encodes the receptor-binding domain (RBD), the part of the virus that helps it lock onto and infect cells. The technique fishes those RBD sequences out of the genetic soup, allowing the researchers to identify unique patterns of mutations that could provide clues about how the virus is continuing to evolve.

More surprises are likely. The Omicron variant, for example, seems to result in much less virus shedding in feces. In multiple countries, Medema says, wastewater levels lagged slightly behind the explosion in cases. After comparing results, researchers have concluded that, after Omicron took over, wastewater measurements have been underestimating cases—perhaps by a factor of three or four. The reduced shedding might be due to the changes in the virus—or the higher levels of immunity in the population, Johnson says.

Case numbers and wastewater levels in most countries are plummeting in tandem—at least for now. "We all feel it's finally sort of over. Though of course it's not really over, and the virus will continue to be there," Medema says. "We will continue to use wastewater as a sentinel to see what this virus is doing in the population." ∎

Caracas, Venezuela, participated in an international pilot program for SARS-CoV-2 monitoring in wastewater.

limited resources, it's important to think about where to best deploy them," Safford says. If wastewater is simply confirming the trends already seen in clinical tests, she says, it may not be the best investment.

And in some cases, politicians simply aren't very interested. In December 2021, wastewater analysis in Florida's Orange county—home to Walt Disney World—showed Omicron, not yet detected in patients, was already the dominant virus. That meant a huge spike in cases was coming—but Florida's hands-off policies meant the find made little difference in public health policies.

ON THE OTHER SIDE of the Omicron wave, however, wastewater has provided reassurance that the decrease seen in tested cases is real, and not just an artifact of maxed-out

netic beads or cotton "tampons" to trap the virus. The device is submerged in a sewershed and can collect a sample over hours or days, helping eliminate some of the irregular swings in viral concentration that can result from simply dipping a bottle into the water once a week. Ultimately, Gawlik says, researchers would like to converge on a standard protocol for collection and analysis.

Standardization could also benefit surveillance for other pathogens. Screening for poliovirus has been in place in many countries for decades, and several regions have kept tabs on illicit drugs in wastewater, but the pandemic has increased interest in looking for other diseases. In Bengaluru, Chaudhuri notes, diarrheal diseases are major killers. Wastewater monitoring might allow health authorities to

Workers leave a garments factory in Gazipur, Bangladesh, 3 February 2022.

POLICY FORUM

End COVID-19 in low- and middle-income countries

Vaccines are changing the course of the COVID-19 pandemic, but in grossly uneven ways. Low- and middle-income countries (LMICs) face considerable obstacles in both receiving and distributing doses. To limit virus transmission, its devastating impacts, and opportunities for further mutations, this must change. Until it does, nonpharmaceutical interventions such as masking must remain a priority. *Science* invited global experts to highlight research and innovations aimed at quickening the end of COVID-19 in LMICs. **—Brad Wible**

Contributions to COVID-19 research and innovation

By **Amrita Ahuja**[1] and **Ahmed Mushfiq Mobarak**[2]

Although it is common knowledge that LMICs have suffered severe pandemic-related economic consequences, less well known is that many of their pioneering research and innovation efforts have helped mitigate pandemic impacts and shape policy globally, including in high-income countries (HICs). Prominent examples include genome sequencing in South Africa, which led to early identification of the Omicron variant; vaccine development of both injectable and intranasal Covaxin in India; and trials of fluvoxamine, an existing drug repurposed for treatment of COVID, in Brazil. Models integrating economic and epidemiological concerns for cost-benefit analyses of lockdown policies were first developed for LMIC contexts (*1*), and household survey data collected in

LMICs highlighted the large losses to income, employment, market access, and food security during the pandemic (*2*). These influenced discourse around the nature and length of economic and social restrictions globally.

Masking guidelines from the World Health Organization (WHO) and US Centers for Disease Control and Prevention (CDC) remained unclear months into the pandemic. A randomized controlled trial involving 347,000 adults in rural Bangladesh that successfully encouraged consistent mask use showed substantial declines in symptomatic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission. This has since informed masking policy across contexts, from South Asian villages to US schools, and influenced revisions to WHO and CDC masking guidelines. Other studies in LMICs have generated insights on how to encourage adherence to public health behaviors such as social distancing and vaccination. Many frugal innovations to contain COVID-19 spread—such as methods to reach and persuade remote populations regarding new public health behaviors—were originally tested in West Africa during the Ebola crisis (*3*). One of the first population-wide, representative-sample pooled polymerase chain

reaction testing efforts was done in Punjab, Pakistan, resulting in algorithms for efficient large-scale pooled testing (*4*). Targeted lockdowns based on color-coded viral spread risk factors were implemented in Pakistan starting in April 2020 and then adopted elsewhere (*5*). Studies in LMICs alerted policy-makers about the special risks the pandemic posed to migrants, women, and children; to mental health; and to routine health care (*6*).

These innovations are products of a strong research infrastructure emerging in many LMICs, exemplified by the abilities to (i) conduct trials and surveys in community settings at massive scale and rapid pace and (ii) combine tools from public health and social sciences to understand the interplay of the disease and social environments. These advantages could be leveraged further—for example, to provide evidence on issues like disease transmission pathways and externalities, and individual and community impacts of alternate vaccine doses, testing approaches, and therapeutics for early treatment of disease.

Different environments lead researchers to ask different questions. Supporting high-quality research in a range of environments addresses a wider set of critical questions and creates a broader and stronger array of tools and strategies for pandemic management. Research investments in LMICs would facilitate cost-effective provision of global public goods for ending the pandemic not only in LMICs, but everywhere.

Lessons from India in April 2021

By **Gagandeep Kang**[3]

In April 2021, the second wave of the SARS-CoV-2 pandemic in India moved inexorably from the west and north to the east and south. Hospitals in northern India had no beds, oxygen was in short supply, and health systems were unable to cope. The scale of infection was documented by a survey showing seropositivity increasing from just over 20% in January 2021 to 67% in June–July 2021 (*7*). Because vaccine supply was a constraint until July, the bulk of the antibody acquisition was likely to be from infection.

The cities of Mumbai and Delhi struggled during the first wave in 2020, but the magnitude, severity, and speed during the second wave in April 2021 were on a different scale. Cases had started to increase a few weeks earlier, and scientists had ascribed the rise to a new variant. The government acknowledged the "double mutant" detection, later named the Delta variant, in March but in mid-April 2021 stated that the variant was not established to be more transmissible (*8*). Despite public health researchers emphasizing the need for continued vigilance and genomic surveillance, there were no restrictions on the rallies conducted in key states in preparations for elections. Further, the Kumbh Mela, the world's largest religious gathering, which stretches several weeks, began in April, drawing pilgrims across India to Haridwar to participate in prayers and ritual bathing. There were a million attendees on some days. Testing was advised but not implemented or monitored, and newspapers reported high positivity in returning pilgrims.

As infections exploded, polypharmacy and the inappropriate use of steroids, antibiotics, antivirals, and anti-inflammatories created both shortages and a second epidemic of cases of mucormycosis, not seen at the same scale anywhere else in the world (*9*). SARS-CoV-2 cases and deaths continued to climb, and hospitals converted the greater proportion of their beds to COVID-19 wards. Care for other conditions, such as cancer, could not be accessed by the bulk of the population, and the consequences of the gaps that were created remain to be fully measured (*10*).

Vaccination was initiated in January 2021 backed by electronic registration systems, but supply constraints restricted vaccination to 1 to 3 million doses a day, insufficient to protect the bulk of the population. Unlike the excellent tracking of vaccination, the ability to accurately estimate the scale of morbidity and mortality during the pandemic has been a major gap. Available mortality analyses show that socioeconomic deprivation led to greater excess deaths during the second wave (*11*), but these data are from states with stronger health and data systems, and the full impact on India may only be available with census data on a longer time scale.

India's second wave offers lessons that point the way for the future for LMICs—early signal detection, acknowledgment, and analysis; preparedness of the health system and its supply chains to not only treat the emergent situation but also protect the ability to handle care provisions for other conditions; and the ability to collect, collate, integrate, analyze, and interpret data in real time so that resources can be directed appropriately when needed, and health systems can build an increased understanding of what drove poor and good outcomes.

The economic impact of COVID-19

By **Edward Miguel**[4] *and* **Ahmed Mushfiq Mobarak**[2]

The COVID-19 pandemic is the largest macroeconomic shock the world has seen since the cataclysm of the Great Depression and World War II, with a global decline in gross domestic product of 3% (*6*), far larger than the shock caused by the Great Recession of 2008–2009 or the 1998 Asian financial crisis. LMICs have been especially hard-hit, with an outpouring of quantitative evidence from household surveys documenting adverse effects of the pandemic on economic outcomes and living standards. The majority of households in most LMIC country samples reported sharp drops in incomes, employment, and consumer spending (*2*, *12*) in the early months of the pandemic. Although unemployment rose and earnings dropped in nearly all countries—rich and poor—only among the poorest populations does this translate into widespread hunger with attendant deleterious effects on long-run child growth and cognitive development, whereas in rich countries social safety nets kick in to provide some measure of food security. Fifty-six percent of Rwandans and Sierra Leoneans reported either missing meals or reducing food portion sizes in nationally representative surveys conducted in May–June of 2020 (*2*).

Researchers have documented even broader adverse effects on education, health care access, mental health, and increases in domestic violence (*6*). Partly in response to growing concerns about food security and meeting basic family needs, the first 2 years of the pandemic saw new public social safety net programs like cash transfers introduced in over 200 countries (*13*). Reaching beneficiaries quickly required innovations in delivery, such as machine-learning–based targeting, and low-cost mobile money transfers (*14*).

The adverse effects of pandemic-related lockdowns indicate that it was important to think carefully about balancing disease risk with the risk of hunger and other unintended socioeconomic consequences in deciding on the stringency of lockdowns in LMICs, where the reach of social safety nets is more limited. In addition, epidemiological modeling indicates that benefits of lockdown policies were arguably far smaller in LMICs, given that generally younger populations face lower COVID-19 mortality risk (*1*).

Although the optimal design of lockdown policies may have been less clear for LMICs, there is no doubt that there remains an urgent need to promote vaccine take-up, masks, and other preventive behaviors in all countries—both poor and rich.

Doses of COVISHIELD vaccine manufactured by Serum Institute of India are administered in Mumbai, India, 3 May 2021.

Scale up production of COVID-19 vaccines in LMICs

By **Gregg S. Gonsalves**[5,6,7] *and* **Saad B. Omer**[8,9]

As of 15 February 2022, 4.27 billion people around the world had been fully vaccinated against COVID-19. However, vaccine distribution has been highly inequitable, with 45 countries having vaccinated less than 10% of their population, another 105 nations having offered a primary series (i.e., the initial two doses without boosting) to less than 40%, and 20 countries not having enough doses to vaccinate even their elderly citizens and health care workers (15). The COVID-19 Vaccines Global Access (COVAX) facility, founded in April 2020, was meant to ensure equitable vaccine access in LMICs, but though it had promised to deliver more than 2 billion doses globally to the neediest nations by the end of 2021, it struggled to supply even less than half of the shots promised for LMICs (16). COVAX initially suffered from a lack of financial support from HICs, and by the time it obtained sufficient resources, a substantial number of vaccine supplies had been claimed by HICs through direct agreements with manufacturers. Even before the trials of the current COVID-19 vaccines had been completed, many were warning about this kind of vaccine nationalism, in which HICs were signing advance purchase agreements with manufacturers for hundreds of millions of doses, restricting future access to potential vaccines by poorer countries (17). Although health systems constraints and vaccine hesitancy have been raised by some as key barriers in improving access to COVID-19 vaccines, these notions have been challenged by others (18). Other actors, including major nongovernmental organizations, the Joint United Nations Programme on HIV/AIDS, WHO, individual scientists, clinicians, and public health experts, made alternative proposals to expand access to COVID-19 vaccines, by waiving intellectual property protections, sharing technology, and expanding manufacturing capacity for COVID-19 vaccines through production hubs supported by WHO or by government-owned, contractor-operated facilities (19). As 2021 came to a close, these proposals were stalled, with little support from governments in HICs and resistance from the key vaccine companies that would be the required partners in technology transfer. This evident policy paralysis in the quest for global vaccine access is imperiling hundreds of millions of lives, risking the development of new variants of SARS-CoV-2, and delaying the worldwide recovery from this pandemic. As we enter the third year of COVID-19, unless there are new commitments to scale up production and access to vaccines to all who need them, we may be back again in 2023, with numerous additional and preventable deaths to answer for. Although Afrigen Biologics and Vaccines in Cape Town, part of the WHO COVID-19 vaccine technology transfer hub for the region, recently announced it has been able to copy and produce Moderna's messenger RNA (mRNA) vaccine, this was done without support from the company. However, scientists from around the world, including those from the US National Institutes of Health who were involved with the initial work on these immunogens, did assist on the project, which sped progress along (20). This is a small step forward. Meanwhile, BioNTech, the co-manufacturer of Pfizer's mRNA vaccine, has been accused of undermining the efforts by Afrigen and the WHO vaccine hub (21). If HIC vaccine manufacturers are unwilling to help scale up COVID-19 vaccines, at the very least, they have to get out of the way of others who are trying to do so.

Is vaccine hesitancy a problem?

By **Arjun Kharel**[10] *and* **Shana Warren**[11]

COVID-19 vaccine acceptance rates are generally higher in LMICs than in HICs. Studies conducted in 2020 found average acceptance rates across 24 LMICs in Asia, Africa, and South America significantly higher (80%) than in the United States (65%) and across seven HICs in Europe (74%) (18, 22, 23). This acceptance gap is consistent with attitudes toward childhood immunizations prior to the pandemic; 95% of respondents in South Asia and 92% in Africa believe vaccines to be safe, in contrast to 72% for North America, 73% for Northern Europe, and only 59% for Western Europe (24).

Yet LMICs are not a homogeneous bloc; there is substantial variation in vaccine acceptance rates between and within LMICs (18, 24–27). Studies in sub-Saharan Africa, South Asia, and Latin America and the Caribbean find a higher degree of COVID-19 vac-

cine hesitancy among people with less education, and those who are elderly and have lower levels of trust in the health care system (*25–27*). National governments and international agencies must develop context-specific strategies to reach out to different segments of the population within and across countries in the LMICs.

Building trust and mitigating misinformation will remain pivotal to translating vaccine acceptance into uptake, and convincing hesitant segments of the population. Vaccine acceptance rates may fluctuate with every bit of misinformation and media reports of side effects or potential side effects, however small the odds of those side effects may be (*18*). Coherent messaging through trustworthy sources like local health care professionals will remain important (*18*).

As COVID-19 vaccination progresses in LMICs, there are good reasons to be optimistic about translating acceptance to uptake; 79% of Brazilians, 67% of Indians, and 64% of Indonesians have received at least one dose (*28*). Recent surveys in 19 sub-Saharan African countries have found that 78% of respondents had received or planned to receive a COVID-19 vaccination (*25*). These successes demonstrate that hesitancy is not a widespread problem in LMICs when supply is readily available.

Although vigilant efforts to track and counter misinformation to address hesitancy must continue, the data suggest that broadly speaking, vaccine acceptance rates in LMICs are sufficiently high that hesitancy should not be used as an excuse to delay or downsize vaccine shipments. If vaccine doses are being wasted within LMICs, the solution likely requires investments in the supply chain logistics, because LMIC populations are likely to take vaccines when they have access.

Overcoming last-mile vaccine delivery challenges

By **James Dzansi[12], Niccolo Meriggi[13,14], Ahmed Mushfiq Mobarak[2], Maarten Voors[15]**

Making the COVID-19 vaccine easily accessible to everyone, everywhere is the most promising solution to end this pandemic (*29*). Beyond distributing vaccines to every country based on their needs, ensuring access for every individual within each country requires investing in infrastructure for domestic distribution. Many LMICs face substantial challenges in last-mile delivery of vaccines, especially to people living in more remote, rural, low-density areas. These are the countries with deficiencies in both health system capacity and in transportation infrastructures, which jointly make it more difficult for citizens to access vaccination centers. Data that we collected in Sierra Leone show that a trip to a vaccination center for a person residing in a rural community is $6 and 1.5 hours, on average, each way. In a place where over 56% of the population lives hand-to-mouth with less than $1.25 per day, such a transportation cost is prohibitive. Thus, what may at first appear to be vaccine hesitancy and explain why doses get wasted in fact reflects the real constraints on accessibility.

Leaving remote populations unvaccinated exposes the entire world to the risk of new virus mutations. We should therefore prioritize development of creative solutions that enhance access for all LMIC citizens. The good news is that domestic distribution capabilities can be fast to establish, affordable, and cost-effective. LMICs have considerable prior experience with mass immunization campaigns that have led to high levels of childhood vaccine coverage (*18*).

Several countries, including Ghana, Liberia, India, Pakistan, and Sierra Leone, have started experimenting with the concept of "mobile vaccination teams" that take batches of vaccines closer to where people live, to make getting jabs more convenient (*30, 31*). These can involve nurses visiting remote villages with doses, backed by community mobilizers to sensitize the local population and leaders, and gather people for efficient vaccine administration. Such models have been applied successfully in the past: "outreach clinics" to provide immunization services to hard-to-reach subpopulations are cited as an important factor in eliminating measles in The Gambia (*32*).

Such frugal innovations can be cost-effective solutions to the accessibility issue, but public health officials, researchers, and nongovernmental organizations need to further experiment with these models, to understand whether small financial incentives or other complementary services can improve efficacy (*33*). To achieve vaccine equity, supplying doses of COVID-19 vaccines to LMICs needs to be supplemented with creative efforts to reach remote, underserved areas within each country.

Optimizing vaccine dosing in pandemics

By **Denise Garrett[16], Michael Kremer[17], Helen Rees[18], Babatunde Salako[19], Firdausi Qadri[20], Witold Więcek[17]**

COVID-19 highlighted that accelerating vaccine rollout in LMICs during pandemics requires parallel, complementary investments, including vaccine distribution systems, vaccine manufacturing capacity (*34*), and more equitable systems for allocating doses based on health need. We argue that optimizing vaccine dosing can also substantially increase supply, thereby increasing vaccine equity. The strategy has also been used during global vaccine shortages; for example, one-fifth doses were recommended by WHO Strategic Advisory Group of Experts on Immunization (SAGE) and used for yellow fever (*35*) and inactivated polio (*36*) vaccines.

Vaccine developers typically optimize dosage by trading off efficacy with possible side effects, both of which may increase with higher doses. The COVID-19 pandemic created pressure to evaluate preclinical candidates and then produce high-efficacy vaccines quickly, leading to large doses being tested and adopted. Once vaccines were approved, developers faced overwhelming commercial incentives to stick with the approved formulations. Yet from a public health standpoint, during a pandemic and a vaccine shortage, there are large potential benefits from using lower

[1]Douglas B. Marshall, Jr. Family Foundation, Houston, TX, USA. [2]Yale University, New Haven, CT, USA. ahmed.mobarak@yale.edu [3]Christian Medical College, Vellore, Tamil Nadu, India. gkang@cmcvellore.ac.in [4]Department of Economics and Center for Effective Global Action, University of California, Berkeley, Berkeley, CA, USA. emiguel@berkeley.edu [5]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA. [6]Yale Law School, New Haven, CT, USA. [7]Yale Global Health Justice Partnership, Yale University, New Haven, CT, USA. gregg.gonsalves@yale.edu [8]Yale Institute for Global Health, Yale University, New Haven, CT, USA. [9]Department of Medicine, Section of Infectious Diseases, Yale University School of Medicine, New Haven, CT, USA. [10]Tribhuvan University, Centre for the Study of Labour and Mobility, Kirtipur, Kathmandu, Nepal. [11]Innovations for Poverty Action, Washington, DC, USA. swarren@poverty-action.org [12]International Growth Centre, Accra, Ghana. [13]International Growth Centre, Freetown, Sierra Leone. [14]Development Economics Group, Wageningen University, Wageningen, Netherlands. niccolo.meriggi@theigc.org [15]Development Economics Group, Wageningen University, Wageningen, Netherlands [16]Sabin Vaccine Institute, Washington, DC, USA. [17]University of Chicago, Chicago, IL USA. kremer.m@gmail.com [18]Wits Reproductive Health and HIV Institute, University of the Witwatersrand, Johannesburg, South Africa. [19]Nigerian Institute of Medical Research, Lagos, Nigeria. [20]International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh. [21]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, USA. [22]John F. Kennedy School of Government, Harvard University, Cambridge, MA, USA. marcella_alsan@hks.harvard.edu [23]Department of Economics, Harvard University, Cambridge, MA, USA. [24]Department of Economics, Stanford University, Stanford, CA, USA. [25]School of Management, Yale University, New Haven, CT, USA. [26]Yale School of Management, Yale University, New Haven, CT, USA. jason.abaluck@yale.edu [27]Center for Effective Global Action, University of California, Berkeley, Berkeley, CA, USA. [28]North South University, Dhaka, Bangladesh. [29]Vyxer Remit Kenya, Busia, Kenya.

Maasai elders don face masks at their homestead in Narok County, Kenya, 10 August 2020.

doses to save supplies, thus allowing more doses to be distributed. This could greatly accelerate vaccination, particularly benefiting those at the end of the queue (mostly people in LMICs). If further booster doses are required or if a new variant emerges against which some vaccines are ineffective, the benefit of fractional dosing would be even higher.

Several published and ongoing studies suggest that lower doses of the more effective vaccines lead to a robust immune response for both primary vaccination and boosters [for a regularly updated review of evidence on fractional dosing and list of ongoing trials see (37)]. For some vaccines, doses at one-quarter to one-half the level currently used may be superior to full doses of the less effective vaccines being used in some LMICs (38). Lowering doses may also reduce both mild side effects and the risk of serious, rare adverse events seen with these vaccines, thereby increasing vaccine acceptance. Some upper middle- and high-income countries have already authorized reduced doses for boosters and for children, and interviews with decision-makers and researchers in several LMICs suggest great interest in optimizing dosing if research confirms they are beneficial.

International and national regulators and policy-makers should review evidence from lower-dose vaccine trials and decide whether reoptimizing dosing levels (for both primary series and boosters) would likely improve public health in their context. For future pandemics, research on dosage optimization should receive early support as a global public good.

Mobile phone messaging to promote preventive behaviors

Abhijit Banerjee[21], Marcella Alsan[22], Emily Breza[23], Arun G. Chandrasekhar[24], Esther Duflo[21], Paul Goldsmith-Pinkham[25], Benjamin A. Olken[21]

Populations vulnerable to COVID-19 due to conditions of poverty and marginalization often have less access to timely, accurate, and credible information. For the first time in history, most poor people had direct or indirect access to a mobile phone during a global pandemic. This was both an opportunity and a danger: Mobile phones could be used to transmit useful public health messages to the most remote corners of the world. At the same time, individuals were potentially overwhelmed by messaging. In a survey we conducted in West Bengal, India, in May 2020, the average person had received about 20 messages on COVID-19 in the previous 2 days.

In this context, is it possible to use mobile phone messaging to convey information and promote prevention? How best to do it? Using a series of mobile phone messaging interventions across developed and developing countries, we have found that trusted messengers can induce preventive behavior change during a pandemic and are effective even in an information-rich environment or polarized climate.

Building upon prior research in Indonesia demonstrating that messages delivered by celebrities improved vaccination rates, we worked with the government of West Bengal to develop and disseminate videos that featured Nobel Laureate and West Bengal native Abhijit Banerjee discussing the importance of reporting symptoms to local health workers and social distancing (39, 40). The video messages were sent via text to millions of mobile phone users across West Bengal in randomly selected small geographical areas, and preliminary results suggest the intervention had immediate and lasting effects on reports of symptoms to health workers, self-reported preventive behavior, and mobility. We then used a similar strategy in the United States, partnering with 40 physicians at the Massachusetts General Hospital's Center of Diversity and Inclusion, who recorded video messages on masking and distancing. We tested these videos in experiments involving thousands of participants recruited on online survey platforms and found that the messages affected knowledge of effective prevention, willingness to pay for masks and obtain more information, and self-reported prevention behavior (41, 42).

Encouraged by these results, we sent millions of messages (by text in West Bengal, and via Facebook ads in the US) to encourage people to stay put prior to major holidays in both countries. In the US, we randomly assigned counties to different message penetrations via Facebook. We documented meaningful declines in movement (recorded by mobile phones) and COVID-19 cases in treated areas, underscoring the power of actionable, well-timed messages targeting specific behavior (43).

Notably, the exact content of the message, which was also ran-

domized in the first set of experiments, did not make a considerable difference. Furthermore, even in a polarized climate, impacts were similar for all types of recipients, suggesting that specific messages by trusted messengers can still lead to meaningful change at scale.

Promoting mask wearing to reduce COVID-19 infections

By **Jason Abaluck**[26], **Aleksandra Jakubowski**[27], **Muhammad Maqsud Hossain**[28], **Carol Nekesa**[29], **Edward Miguel**[4]

Masks alone will not eliminate COVID-19 and they are not a substitute for vaccines. But when COVID-19 mortality is high and health systems are strained, masks can potentially save many lives at low cost.

A growing body of laboratory and field studies indicates that masks reduce the public health burden of COVID-19 (44). A randomized trial in Bangladesh found that a 29 percentage-point increase in mask use led to a 9% reduction in symptomatic infections over a 10-week period (45). The reductions were especially large among the elderly, among whom infections fell by 15 to 35%. But despite this evidence, some people have adopted the viewpoint that everyone will ultimately be infected—especially with highly infectious variants like Omicron—so masks are an unnecessary nuisance that will have no impact on the long-term infection rate. There are two reasons why this view is not right. First, delaying infections gives people more time to become vaccinated, a point especially critical in LMICs where vaccine distribution has been slow. Second, by reducing viral load at transmission, masks may reduce the severity of illness and risk of death when infections do occur (46). More generally, masking by symptomatic people may be warranted to reduce transmission of respiratory diseases other than Covid (47).

The costs and benefits of masks are not identical at all times and places. The benefits are larger in indoor, crowded areas with poor ventilation. Masking in schools remains complex; though children bear lower morbidity and mortality burden from Covid, masks also prevent secondary infections in the elderly. The long-term impacts of wearing masks in schools on learning are not well-understood, in addition to discomfort costs. Both in schools and elsewhere, the benefits of masking are larger during surges when many people are hospitalized or dying of Covid (and thus where fewer people are adequately vaccinated).

Given masks' low cost and relative ease of use, how can we get people to wear masks when their use is warranted? A preliminary point is that direct observation is necessary to measure mask use due to social desirability bias (48): in Kenya, 88% of survey respondents said they wear masks to public places but only 10% of people were observed with masks (49), and gaps were even larger in Uganda (50). People say they wear masks but promoting adoption at appropriate times is a more serious challenge.

How can we increase actual observed mask use rather than just self-reports? Preliminary findings from Africa suggest that mask distribution alone barely moved the needle on masking in Kenya (50) and that pairing mask distribution with education increased take-up by 3 percentage points in the short-term, with this effect fading over time (51). Asking people directly to wear masks in public areas led to substantially larger increases in masking in the aforementioned Bangladesh study (29 percentage points). The effectiveness of such reinforcement will of course vary across contexts due to the underlying political situation; in places where there is not active political resistance to mask-wearing, approaching people in public and asking them to put on masks is the most effective strategy to date, but important questions remain about whether this approach will continue to work after multiple waves.

REFERENCES AND NOTES

1. Z. Barnett-Howell, O. J. Watson, A. Mushfiq Mobarak, *Trans. R. Soc. Trop. Med. Hyg.* **115**, 807 (2021).
2. D. Egger et al., *Sci. Adv.* 10.1126/sciadv.abe0997 (2021).
3. N. F. Meriggi, A. M. Mobarak, "This Country Fought Ebola. It May Beat Another Disease," *The New York Times*, 16 June 2020.
4. F. Majid, S. B. Omer, A. I. Khwaja, *Lancet Microbe* **1**, e101 (2020).
5. Smart Containment with Active Learning: A Graded & Data-Responsive Approach to COVID-19, Center for International Development (CID), Harvard University Special Brief: COVID-19, May 2020, https://www.hks.harvard.edu/sites/default/files/centers/cid/files/publications/CID%20Special%20Brief_COVID19_May20.pdf.
6. E. Miguel, A. M. Mobarak, "The economics of the COVID-19 pandemic in poor countries," NBER Working Paper 29339, National Bureau of Economic Research, October 2021.
7. N. Jahan et al., *Int. J. Infect Dis.* **116**, 59 (2021).
8. Press Information Bureau, Government of India, www.pib.gov.in/PressReleasePage.aspx?PRID=1712312.
9. U. Arora et al., *J. Infect.* 10.1016/j.jinf.2021.12.039 (2021).
10. P. Ranganathan et al., *Lancet Oncol.* **22**, 970 (2021.)
11. J. A. Lewnard et al., *Lancet Infect. Dis.* 10.1016/S1473-3099(21)00746-5 (2021).
12. T. Bundervoet, M. E. Dávalos, N. Garcia, "The Short-Term Impacts of COVID-19 on Households in Developing Countries: An Overview Based on a Harmonized Data Set of High-Frequency Surveys" (Policy Research Working Paper 9582, The World Bank, 2021).
13. U. Gentilini, M. Almenfi, I. Orton, P. Dale, "Social Protection and Jobs Responses to COVID-19: A Real-Time Review of Country Measures" (The World Bank, 2020); https://openknowledge.worldbank.org/handle/10986/33635.
14. E. Aiken et al., "Machine learning and mobile phone data can improve the targeting of humanitarian assistance," NBER Working Paper 29070, National Bureau of Economic Research, July 2021.
15. M. K. Patel, *N. Engl. J. Med.* **385**, 2476 (2021).
16. "Covax promised 2 billion vaccine doses to help the world's neediest in 2021. It won't even deliver even half that," Washington Post, www.washingtonpost.com/world/2021/12/10/covax-doses-delivered/.
17. K. Kupferschmidt, *Science* 10.1126/science.abe0601 (2020).
18. J. S. Solís Arce et al., *Nat. Med.* **27**, 1385 (2021).
19. M. M. Kavanagh, L. O. Gostin, M. Sunder, *JAMA* **326**, 219 (2021).
20. A. Maxmen, *Nature* **602**, 372 (2022).
21. M. Davies, *BMJ.* **376**, o304 (2022).
22. S. Neumann-Böhme et al., *Eur. J. Health Econ.* **21**, 977 (2020).
23. AfricaCDC, COVID 19 Vaccine Perceptions: A 15 country study (10 March 2021); https://africacdc.org/download/covid-19-vaccine-perceptions-a-15-country-study/.
24. Wellcome Global Monitor, Wellcome (2018); https://wellcome.ac.uk/reports/wellcome-globalmonitor/2018.
25. Prevent Epidemics, Responding to COVID-19 in Africa: Finding the Balance (2021); https://preventepidemics.org/covid19/perc/.
26. M. Abedin et al., *PLOS One* **16**, e0250495 (2021).
27. A. J. Rodriguez-Morales, O. H. Franco, *Lancet Regional Health–Americas* **3**, 100073 (2021).
28. Our World in Data, Coronavirus (COVID-19) Vaccinations (2022); https://ourworldindata.org/covid-vaccinations.
29. N. D. Paula, C. Brown, *Lancet Planet. Health* **5**, e758 (2021).
30. L. Bloxham, Covid-19 vaccine rollout in Liberia and Sierra Leone helps reach vulnerable people. Concern Worldwide (2021); www.concern.org.uk/news/covid-19-vaccine-rollout-liberia-and-sierra-leone-helps-reach-vulnerable-people.
31. WHO Regional Office for Africa, Emerging lessons from Africa's COVID-19 vaccine rollout (2021); www.afro.who.int/news/emerging-lessons-africas-covid-19-vaccine-rollout.
32. O. Wariri et al., *Lancet Global Health* **9**, E280 (2021).
33. A. Banerjee et al., "Selecting the most effective nudge: Evidence from a large-scale Experiment on Immunization," NBER Working Paper 28726, National Bureau of Economic Research, April 2021.
34. A. Ahuja et al., *AEA Pap. Proc.* **111**, 331–35 (2021).
35. World Health Organization, *Vaccine* **35**, 5751 (2017).
36. H. Okayasu, *J. Infect. Dis.* 216 (suppl. 1), S161 (2017).
37. Development Innovation Lab, Trial landscape of fractional dosing studies of COVID-19 vaccines: A living review of evidence (2021); https://bfi.uchicago.edu/development-innovation-lab/fractional-dosing-trials/.
38. W. Więcek et al., "Testing fractional doses of COVID-19 Vaccines," NBER Working Paper 29180, National Bureau of Economic Research, August 2021.
39. A. Alatas et al., "When celebrities speak: A nationwide Twitter experiment promoting influenza vaccination," NBER Working Paper 25589, National Bureau of Economic Research, February 2019.
40. A. Banerjee et al., "Messages on COVID-19 prevention in India increases symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their community," NBER Working Paper 27496, National Bureau of Economic Research, August 2020.
41. M. Alsan et al., *Ann. Intern. Med.* **174**, 484 (2021).
42. C. Torres et al., *JAMA Network Open.* **4**, e2117115 (2021).
43. E. Breza et al., *Nat. Med.* **27**, 1622 (2021).
44. J. Howard et al., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014564118 (2021).
45. J. Abaluck et al., *Science* **375**, 6577 (2021).
46. M. Gandhi et al., *J. Gen. Intern. Med.* **35**, 3063 (2020).
47. C. Feldman, R. Anderson, *Pneumonia* **13**, 5 (2021).
48. P. Grimm, "Social desirability bias," Wiley International Encyclopedia of Marketing (2010).
49. A. Jakubowski et al., *JAMA Network Open.* e2118830 (2021).
50. A. Jakubowski et al., "Evaluation of a national program to distribute free masks for COVID-19 prevention in Uganda: Evidence from Mbale District," CEGA Working Paper WPS-193, Center for Effective Global Action, 2022.
51. D. Egger et al., medRxiv 10.1101/2022.02.16.22270815 (2022).

10.1126/science.abo4089

Gender-responsive social protection post–COVID-19

Investment in gender-responsive social protection systems and evidence is key to a more equal future post–COVID-19

By **Maja Gavrilovic[1], Monica Rubio[2], Francesca Bastagli[3], Roopa Hinton[4], Silke Staab[5], Ruth Graham Goulder[4,6], Charlotte Bilo[7], Ruby Khan[8], Amber Peterman[9,10], Bobo Diallo[11], Laura Alfers[12], Aroa Santiago[13], Zehra Rizvi[14], Rebecca Holmes[3], Juan Gonzalo Jaramillo Mejia[15], Constanza Tabbush[5]**

COVID-19 has reaffirmed that in the face of crises, social and economic fallout is gendered. From the risk of job loss and economic instability to rising care responsibilities and the experience of violence inside the home, gender inequalities have tended to widen during the pandemic (1, 2). While countries focus on health and mortality impacts of the disease, a mounting, damaging gendered social and economic crisis threatens to roll back decades of development progress, exposing the fragility of equality gains. Social protection has been a key policy response to address pandemic-related social and economic crises; however, attention to gender has been insufficient. Less than one in five global social protection measures during COVID-19 has addressed gender, such as supporting women in informal employment, mitigating risks of violence, and confronting the unequal distribution of care work. Policy priorities (see the box) must include closing gendered research gaps in the COVID-19 recovery.

In early 2020, researchers called attention to evidence from previous infectious-disease outbreaks, such as Ebola and Zika, and predicted that COVID-19 impacts would follow similar gendered trends (3). Two years later, multidisciplinary evidence from diverse settings shows that disease and control measures have had detrimental—and often disproportionate—effects on economic security, health, well-being, protection, and opportunities for women and girls (1, 2). For example, a study of 40 countries found that women were more likely to stop working between April and June 2020 as compared to men (36 versus 28%) (2). Similarly, a study of 46 countries conducted in June and July 2020 found that girls were more likely to report increases in household chores as compared to boys (63 versus 43%) (2). Women have also been underrepresented in decision-

making on COVID-19 yet are crucial voices to mitigate the adverse gendered impacts of the pandemic and achieve an equitable recovery for families and society (1).

STRENGTHEN SOCIAL PROTECTION

The World Bank and partners estimate that by May 2021, more than 3300 social protection and labor market measures had been planned, adapted, or implemented in more than 220 countries and territories in response to COVID-19 (4). More than $2.9 trillion dollars, or about 3% of global gross domestic product, in 2021 was devoted to these measures. Despite this unprecedented response, measures are typically temporary, cover a fraction of those in need, and are implemented in predominantly high-income countries. Calls for establishing universal social protection, in line with United Nations (UN) Sustainable Development Goal 1.3, acknowledge that expanding coverage, ad-

equacy, and comprehensiveness also requires that systems address the needs of different populations. COVID-19 has reestablished the importance of making social protection work for women and girls.

The call to address gendered poverty and risks through social protection is not new. A robust body of evidence shows that social protection measures, including cash transfers, are promising tools to improve women's basic consumption, socioeconomic status, agency, mental and physical health, and schooling outcomes for girls and to reduce intimate partner violence, among others (5–7). Global stakeholders have called for greater action to use and strengthen social protection systems to promote gender equality and empower women and girls (8). This investment in "gender-responsive" or "gender-transformative" social protection explicitly responds to the differentiated needs of women and girls, alongside those of men and boys, to tackle the root causes of gender inequality, including discriminatory gender norms and practices. A new conceptual framework that is guiding investment in gender-responsive social protection advocates for a systematic integration of gender into all stages of intervention—from design and implementation to monitoring and evaluation—to optimize gender equality gains and ensure that programs and systems do not reinforce inequalities (9). Nonetheless, there is mounting evidence that the social protection pandemic response

Little attention to gender in government social protection measures

The figure includes data from 3099 social protection and labor market responses, with the total number of measures per region given in each corresponding bar. Because of rounding, percentages may not add up exactly to 100%. Social protection refers to both noncontributory social assistance (e.g., cash transfers, food, and in-kind support) and contributory social insurance (e.g., unemployment benefits, paid family leave, and health insurance). Labor market measures include, for example, reduced work time, wage subsidies, and training programs. Attention to gender is categorized as either targeting women's economic security or supporting unpaid care. Data are from (10).

Legend: ● Other measures ● Targeting women's economic security ● Supporting unpaid care

Region	Other measures	Targeting women's economic security	Supporting unpaid care
World n = 3099 measures	80%	12%	7%
Latin America and the Caribbean n = 766	78%	18%	4%
Europe, North America, Australia, and New Zealand n = 930	80%	5%	15%
Sub-Saharan Africa n = 418	80%	17%	2%
Northern Africa and Western Asia n = 380	81%	15%	4%
Central and South Asia n = 183	81%	15%	4%
Eastern Asia, Southeastern Asia, and Oceania n = 422	86%	9%	5%

has overlooked gender considerations, a reflection of existing systems that are poorly prepared to tackle gender within shock-responsive programming.

The COVID-19 Global Gender Response Tracker was launched in September 2020 by the United Nations Development Programme (UNDP) and UN Women to track response measures by governments across key sectors and assess them from a gender perspective. The range of social protection responses and their attention to gender by type of measure and region was analyzed (see the figure). By July 2021, out of 3099 social protection and labor market measures included globally, only 19.6% (606 measures) took gender into account. These include 12.3% that target or prioritize women's economic security and 7.3% that provide support for rising unpaid care demands (*10*). In addition, a complementary analysis shows that out of 262 COVID-19 task forces in 130 countries, only 24% of task force members were women (*10*). Like all policy trackers, there may be gaps or biases due to lack of information (including geographic coverage and scale of the measure), underreporting of measures being announced, overreporting of suspended measures, or the lack of data on key gender components. Even so, the lack of attention to gender highlights the inadequacy of systems and their shock responsivity in planning for and prioritizing the specific needs of women and girls throughout the life cycle and limits their ability to address intersectionality by important dimensions, including disability, race, ethnicity, and nationality.

PROMISING APPROACHES
Not all social protection responses have failed to integrate gender meaningfully. Governments have intentionally designed or adapted social protection responses to mitigate gendered risks across three domains: (i) supporting women in informal employment, (ii) tackling the risk of violence against women and girls, and (iii) confronting the unequal distribution of care work and strengthening care systems. Examples of cross-cutting best practices have also been documented, including undertaking preprogram gender analysis to understand gendered needs, risks, and opportunities; engaging and coordinating responses with local actors; and taking an inclusive approach that leads to broad population (or universal) coverage, with no strict behavioral or eligibility requirements, promoting choice, dig-

nity, and agency (*8*). Moreover, although we focus on women and girls, a gender-transformative approach requires addressing the needs of and working with men and boys, particularly to change discriminatory gender norms in the household and society. For all case studies described below, we draw on the COVID-19 Global Gender Response Tracker (*10*), unless explicitly cited otherwise.

WOMEN IN INFORMAL EMPLOYMENT
It is estimated that 85% of women's total employment in low-income countries is in the informal sector, a sector heavily affected by the global economic contraction consisting of unregistered jobs, enterprises, and economic activities where levels of vulnerable self-employment are high (*6*). This so-called "missing middle" in social protection coverage includes women working as street traders or at home producing goods, who often lack access to the contributory social protection typically offered through formal employment and financed through payroll (e.g., unemployment benefits, paid family leave, health insurance). Yet these populations often do not meet the standard eligibility criteria of tax-financed noncontributory social assistance (e.g., support without pay-in required, including cash or food transfers) because they are income earning and engaged in the labor market. During COVID-19, several countries extended coverage to informal workers, including in Argentina, Brazil, Colombia, Kenya, and Togo, giving priority in targeting or extra benefits to women. In Brazil, the Emergency Grant targeted monthly cash benefits over the course of 2020 to 68.2 million recipients, many of whom were informal workers. Female single-headed households were entitled to receive double benefit values in recognition of their dual role as income earners and caregivers (*11*). The pandemic has highlighted the need for systems to reach informal workers, better harmonize contributory and noncontributory components of the social protection system, and ensure decent and safe work during crises.

VIOLENCE AGAINST WOMEN AND GIRLS
Evidence suggests that economic insecurity and poverty-related stress during COVID-19 have been key drivers of intrafamilial violence (*1*). To acknowledge the multiple ways that economic insecurity intersects with violence risk, measures in Ireland, Malta, and Uruguay provided access to rent supplements and safe housing, which allowed women

and children to temporarily escape violent situations. Similarly, Argentina and some Brazilian states established cash transfers explicitly for survivors of violence. In Sierra Leone, the government and partners provided training on gender-based violence to frontline social protection workers, seeking to mitigate risks and ensure more effective embedded referrals to violence response services (*12*). Intentional integration of violence prevention components in social protection is a promising practice to both safeguard participants and tackle compounding risks of poverty and violence.

UNPAID AND UNEQUAL CARE WORK
Key defining factors that contribute to inequalities during COVID-19 are the increased childcare, homeschooling, and care for sick family members shouldered by women and girls. Transformative social protection can recognize, reduce, and redistribute care work within families and lay the foundation for more equitable care systems and norms in a postpandemic world. Numerous promising examples exist of government schemes that extend or establish new family leave policies, which allows workers to care for dependents, including those in Austria, Canada, the Republic of Korea, Seychelles, Trinidad and Tobago, and Uzbekistan. Teleworking and flexible work arrangements became common, and some countries like Belgium, Chile, and Hungary provided parents with cash-for-care payments to compensate for childcare and school closures. Others have dropped cash transfer conditionalities, provisions that require participants to comply with behavioral requirements to gain or maintain eligibility, which may have inadvertently increased care work for female participants. During the recovery, public investments in care systems—including accessible, affordable, and quality child and long-term care services—could yield a "triple dividend": supporting women's (re)entry into the labor market, promoting child development and care of the elderly, and increasing the availability of decent jobs in the paid care sector (*13*).

Although not directly comparable to government responses, measures implemented by the humanitarian sector have also been critical in settings affected by fragility, conflict, and disaster, where government social protection systems are nascent, nonexistent, or unable to reach the scale of support needed. Although there has been no system-

[1]Social and Economic Policy Unit, United Nations Children's Fund (UNICEF) Office of Research—Innocenti, Belgrade, Serbia. [2]Social Policy Unit, UNICEF, Panama City, Panama. [3]ODI, London, UK. [4]Policy Division, Foreign, Commonwealth and Development Office, London, UK. [5]Research and Data Section, UN Women, New York, NY, USA. [6]Child Poverty and Social Protection Unit, UNICEF, New York, NY, USA. [7]International Policy Centre for Inclusive Growth, Brasilia, Brazil. [8]Save the Children, London, UK. [9]Department of Public Policy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [10]Center for Global Development, Washington, DC, USA. [11]Economic Empowerment Section, UN Women, New York, NY, USA. [12]Social Protection Programme, Women in Informal Employment: Globalizing and Organizing, Makhanda, South Africa. [13]Gender Team, United Nations Development Programme, New York, NY, USA. [14]Social Policy Unit, UNICEF, Amman, Jordan. [15]Social Protection Unit, World Food Programme, Rome, Italy. Email: amberpeterman@gmail.com

Recommended actions for gender-responsive social protection

Actions adapted from (15).

Policy and system-level enablers

- Expand political commitments and fiscal space with earmarked investments for gender-responsive social protection.
- Commit to advancing gender equality in national social protection policies and systems to better respond to covariate shocks, with a focus on rights and dignity.

Operational and implementation-level facilitators

- Eliminate practical barriers that limit women's access to contributory and noncontributory social protection (e.g., constraints in access to information, technology, documentation of identification, financial inclusion, safe transport, and decent work).
- Extend social protection to previously uncovered groups, including informal workers, recognizing that women are often disproportionately concentrated in lower-paying work within the informal economy.
- Invest in linkages between social protection and care systems to adequately recognize, reduce, and distribute care work equitably across society and support care infrastructure.
- Promote systems linkages and social protection "plus" approaches that support women and girls in multiple domains, including advancing women's economic empowerment and labor market participation and changing harmful social norms that perpetuate gendered violence, discrimination, and inequality.
- Prioritize women's active leadership and political voice in decision-making structures, including meaningful partnership with local organizations that champion rights of women, persons with disabilities, and indigenous peoples.

Data, research, and learning

- Invest in data, research, and evidence generation, including monitoring gender equality in all programming regardless of objectives, to inform future generations of equitable policy-making.

---

atic mapping of these responses, humanitarian actors have been working to better promote gender equality in social protection, guided by human rights principles and the importance of accountability to crisis-affected women and girls (14). Coordination between humanitarian sector and government social protection actors will be critical to successfully deliver gender-responsive programming across the humanitarian-development divide.

GENDER-RESPONSIVE SYSTEMS

In a recent call to action, the gender working group of the Social Protection Inter-Agency Cooperation Board (SPIAC-B) proposed eight actions to "build forward better" in the protracted COVID-19 recovery (15). SPIAC-B is the multiagency, multidonor global coordinating body for advancing social protection cooperation and advocacy. We summarize the recommendations across three levels (policy, program, and evidence) (see the box). The final recommendation—to invest in data, research, and evidence generation—is critical to inform future equitable policy-making. Although social protection is one of the most researched policy interventions globally, gender analysis remains insufficient, a trend that has continued during the pandemic. Few

countries collect and report gender-disaggregated data through monitoring efforts, limiting our understanding of gender differences in coverage as well as adequacy and uptake of entitlements. Although the evidence of social protection's positive impact on gender equality is continually growing, we know less about specific design and operational features that can magnify these impacts across settings (7). We also know little about the cost-effectiveness and synergistic effects of social protection "plus" approaches, including whether linking to existing systems in other sectors (e.g., health, education) or embedding additional gender-responsive programming leads to more effective and sustainable investments. Finally, we have only begun to explore the long-term and intergenerational effects of social protection, with implications for the well-being of future generations. Answers to these questions are challenging given the diversity and complexity of gender norms and intrahousehold relationships in different contexts. Without investment in future evidence generation and uptake in policy-making cycles, we will fail to leverage the potential benefits of social protection for both poverty reduction and gender equality.

The world stands at a crossroads: Either walk down the well-trodden path of fiscal

consolidation, resulting in widening gaps in gender and other inequalities, or invest in gender-responsive policies and programming for a more equal future. Such investment will require drawing on past lessons, building the evidence, and fostering global solidarity and cooperation to make and sustain progress. Social protection yields both current and future dividends. Let us set a clear ambition to make social protection gender responsive and boldly transformative: The time is now. ∎

REFERENCES AND NOTES

1. S. Fisseha et al., Lancet 398, 471 (2021).
2. C. de Paz Nieves, I. Gaddis, M. Muller, "Gender and COVID-19: What have we learnt, one year later?" (Policy Research Working Paper 9709, World Bank, 2021).
3. C. Wenham et al., Nature 583, 194 (2020).
4. U. Gentilini et al., "Social protection and jobs responses to COVID-19: A real-time review of country measures" (Working Paper 159043, World Bank, 2021).
5. F. Bastagli, J. Hagen-Zanker, G. Sturge, "Cash transfers: What does the evidence say? A rigorous review of programme impact and of the role of design and implementation features" (ODI, 2016).
6. L. Alfers, R. Holmes, C. McCrum, L. Quarterman, "Gender and social protection in the COVID-19 economic recovery: Opportunities and challenges" (Social Protection Approaches to COVID-19 Expert Advice Service, 2021).
7. United Nations Children's Fund (UNICEF) Office of Research—Innocenti, "Gender-Responsive & Age-Sensitive Social Protection (GRASSP) think piece series" (UNICEF Office of Research—Innocenti, 2019).
8. SPIAC-B, "Social protection to promote gender equality and women's and girls' empowerment" (SPIAC-B, 2019).
9. UNICEF Office of Research—Innocenti, "Gender-Responsive Age-Sensitive Social Protection: A conceptual framework" (Working Paper WP-2020-10, UNICEF Office of Research—Innocenti, 2020).
10. UNDP, UN Women, COVID-19 Global Gender Response Tracker: Factsheets, November 2021 update; https://data.unwomen.org/publications/covid-19-global-gender-response-tracker-factsheets.
11. N. Yamasaki, F. Rodopoulos, in Policy in Focus, vol. 19 (The International Policy Centre for Inclusive Growth, 2021), pp. 20–21; https://ipcig.org/sites/default/files/pub/en/PIF47_What_s_next_for_social_protection_in_light_of_COVID_19.pdf#page=20.
12. UNICEF, "UNICEF's social protection response to COVID-19: Strengthening social protection systems before, during and after crises" (Technical Report, UNICEF, 2021).
13. UN Women, "Gender equality, child development and job creation: How to reap the 'triple dividend' from early childhood education and care services" (Policy Brief, UN Women, 2015).
14. UN Women, "How to promote gender equality in humanitarian cash and voucher assistance: Guidelines for Grand Bargain Cash Workstream" (UN Women, 2019).
15. SPIAC-B, "COVID-19, social protection and gender equality: A call to action" (SPIAC-B, 2020).

PERSPECTIVE

COVID-19—lessons for zoonotic disease

Disease emergence is driven by human–animal contact in a global viral ecosystem

By **Edward C. Holmes**

The COVID-19 pandemic is an unrelenting demonstration of the devastating impact of zoonotic disease, whereby viruses jump from animals to infect humans. Although there is rightly an urgent focus on the development of vaccines and antivirals to limit the spread and severity of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections, it is essential that this once-in-a-generation experience is used to determine the factors that drive zoonotic disease emergence and identify where gaps in our knowledge lie. By understanding why and how zoonotic diseases emerge in humans, as well as the barriers to this process, it is possible to be better prepared to prevent pandemics like COVID-19 from happening again or at least respond more effectively.

Zoonotic diseases have been part of the human experience since the origin of our species. In cases like SARS-CoV-2 or Ebola, the viral jumps from animals to humans occurred recently, whereas others, such as herpesviruses or papillomaviruses, likely occurred in our earliest ancestors. The antiquity of zoonotic disease highlights the intimate relationship between human and animal viruses. The COVID-19 pandemic has led to considerable debate over the identity and role played by "reservoir hosts" (such as bats), "novel hosts" (like humans), and "intermediate hosts" (with suggestions including pangolins and raccoon dogs) that act as a conduit between the former two. Although it is natural to place humans at the end of this chain of emergence, such an anthropocentric perspective is misleading. In reality, viruses are ubiquitous components of global ecosystems that regularly move between interacting species, but usually do not result in overt disease (*1*). Humans are also part of this viral ecosystem, and rather than being the end point of emergence, they can pass their viruses to other species (*2*).

The key issue, then, is not that zoonotic diseases appear in humans, but that their emergence seems to be increasing in frequency (*3*). Major changes in land use, increasing urbanization, and global connectedness are well documented as driving disease emergence through increasing human–animal contacts and accelerating transmission rates, and climate change will similarly accelerate the rate of zoonotic events. Warming global temperatures will result in changing geographic distributions of wildlife as appropriate habitats shrink, perhaps leading to multispecies refugia that will increase the rate of cross-species virus transmission. Those human populations that rely on the animal world will similarly find subsistence increasingly difficult and so may exploit previously pristine areas

A flying squirrel is for sale in a live animal market in Guangzhou, China, in 2004. Exposure to mammals in live animal markets and the fur trade could exacerbate the potential for virus spillover to humans.

or change farming practices, increasing the risk of exposure to animal pathogens. Unless these processes are limited now, with combating global climate change at the forefront, COVID-19 will only be an unsatisfying taste of what is to come.

A core question in understanding the drivers of zoonotic emergence is whether particular animal groups are common sources of zoonotic viruses. If so, can this information be used to establish a watch-and-act list of those species most likely to carry potentially pandemic viruses? It has long been known that most viral infections in humans have their ancestry in mammals (*4*). Birds are the only other probable source of zoonotic diseases, with the various forms of avian influenza virus that occasionally appear in humans (such as H5 subtype viruses) presenting an ongoing pandemic threat. Although viruses are often abundant in other animal groups (such as bony fish), their phylogenetic distance to humans greatly reduces the likelihood of successful cross-species transmission. Within the mammals, a variety of groups have served as reservoirs for zoonotic viruses (*3*), particularly those with which humans share proximity, either as food sources (such as pigs) or because they have adapted to human lifestyles (like some species of rodent), as well as those that are so closely related to humans (such as nonhuman primates) that viruses face little adaptive challenge when establishing human-to-human transmission.

Most of all, since the emergence of SARS in late 2002, there has been intense interest in bats as virus reservoirs, although this may in part reflect biases in both ascertainment and confirmation (*5*). Although bats seemingly tolerate a high diversity and abundance of viruses, the underlying immunological, physiological, and ecological reasons for this are not fully understood (*6*). More pragmatically, the majority of bat viruses have not appeared in humans, and those that have emerged often do so through other host species (i.e., "intermediate hosts") prior to successful emergence (*7*). Bats are important players in disease emergence, but they are only one component of the more complex global viral ecosystem.

Sydney Institute for Infectious Diseases, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW, Australia. Email: edward.holmes@sydney.edu.au

A related question is whether the viruses that are most likely to emerge in humans can be identified. Although metagenomic sequencing is revealing an increasingly large virosphere, with mammals carrying many thousands of different viruses, most of which remain undocumented (5), the greatest pandemic risk is posed by respiratory viruses because their fluid mode of (sometimes asymptomatic) transmission makes their control especially challenging. Three groups of RNA viruses that regularly jump species boundaries best fit this risk profile: paramyxoviruses, influenza viruses, and, particularly, coronaviruses. Hendra and Nipah viruses, both with ultimate bat ancestry, are exemplars of paramyxoviruses that have emerged in humans. Although neither have resulted in large-scale outbreaks, it is possible that more transmissible paramyxoviruses (such as the case of measles virus) lurk in the mammalian virosphere. The documented host range of influenza viruses is growing, including recent reports of avian H9N2 influenza virus in diseased Asian badgers (2), but most human influenza virus pandemics have their roots in those viruses that circulate in waterbirds and poultry, often with the secondary involvement of pigs. Fortunately, birds and humans are sufficiently different in most virus–cell interactions that avian viruses are usually unable to successfully transmit among humans.

By contrast, coronaviruses are commonly found in mammals that often exist at very high population densities, particularly bats and rodents, or that have strong connections with humans, such as pigs and dogs, and are appearing in humans with an increasing frequency (8). SARS-CoV-2 has also highlighted the potential for "generalist" coronaviruses that can transmit in a wide range of mammalian species. Notably, SARS-CoV-2 has been reported in such animals as cats, dogs, lions, tigers, mink (with transmission back to humans), and, most recently, white-tailed deer in the US, where the virus has jumped multiple times from humans, reaching high levels of prevalence in some populations (9). As sampling of wildlife continues, more SARS-CoV-2–like viruses will surely be identified in a broader range of species and their natural ecology revealed.

The viruses belonging to the evolutionary lineage containing SARS-CoV-2—the sarbecoviruses—have a complex history of genomic recombination (10). Does viral recombination elevate the risk of zoonotic emergence? Recombination increases genetic diversity, generating virus variants that might, by chance, be better able to infect humans, although like most point mutations, most recombinants will reduce fitness. There is also no overall association between the ability

to recombine and the ability of zoonotic viruses to emerge in humans (11). For example, of those viral groups perhaps most likely to cause the next pandemic, coronaviruses and influenza viruses recombine (or reassort) abundantly, whereas paramyxoviruses exhibit very low rates of recombination.

Although viruses frequently jump species boundaries, there are a range of host genetic, immunological, ecological, and epidemiological barriers to successful cross-species transmission. Humans must come into contact with infected animals. This human–animal interface forms the central nexus for disease emergence, and modern human lifestyles mean exposure events will be increasingly commonplace. Following exposure, a virus must establish a productive infection, transmitting within the human population. The barriers at this stage are likely substantial, especially as humans are exposed to many more viruses than lead to disease outbreaks. This is apparent in the relatively high frequency with which people are exposed to animal coronaviruses (12), but the rarity of major outbreaks. The intimate relationship between virus and host cell receptor acts as a major obstacle, and those animal viruses that by evolutionary chance can bind sufficiently well to human cell receptors will have an advantage in the game of emergence. Modeling studies of receptor binding have revealed a myriad of mammalian species that could be productive hosts for SARS-CoV-2 (13), and a similar approach could be employed for other viral groups. However, one important lesson from the ongoing evolution of SARS-CoV-2 is that those viruses that initially emerge do not necessarily need to be fully optimized for transmission to have a substantial impact. Compared to the highly infectious Delta and Omicron variants of SARS-CoV-2, the virus first detected in Wuhan in 2019 was far less efficient at human transmission, but still a good enough respiratory pathogen to spread rapidly in a dense, well-mixed, and fully susceptible population.

How might zoonotic diseases be prevented, or limited? More intensive and effective surveillance at the animal–human interface is the simplest way to mitigate future pandemics. Such surveillance should be performed in those living and working at the human–animal interface, including in the wildlife trade and fur farming, and in animal production and slaughter, who work at live animal markets; in people who live near bat roosts; and even in those who work at animal rescue centers or in the veterinary profession (5). There is little doubt that the wildlife trade and its high-risk end point—live animal markets—pose a danger for emerging zoonotic viruses. It is no coincidence that, like SARS before it, as well as multiple

outbreaks of avian influenza, COVID-19 was initially associated with a live animal market. Recent metagenomic surveillance of the animal breeding facilities that supply these markets in China has identified a high diversity of host-jumping viruses, including novel coronaviruses and influenza viruses, sometimes in animals experiencing respiratory disease (2). By contrast, the large-scale surveillance of wildlife species in nature for potentially pandemic viruses does not seem feasible. Wildlife harbor a large, diverse, and continually evolving pool of viruses, and determining whether they can infect human cells requires time-consuming and costly laboratory work.

As physical distancing has been adopted to dampen the spread of COVID-19, to minimize morbidity and mortality, similar approaches should be deployed to better separate ourselves from wildlife. The wildlife trade and the live animal markets they supply must be strongly regulated and monitored, and effort should be devoted to establishing and maintaining suitable and sustainable environments for wildlife, including bats, located away from population centers. It is imperative that some form of global "pandemic radar" is established in which information on sporadic zoonotic events to full-blown disease outbreaks is shared rapidly and freely (14). Such a radar can involve regular immunological surveillance, perhaps using approaches (15) adapted to recognize those groups of viruses, such as coronaviruses, that most often jump species boundaries, combined with ongoing metagenomic surveillance to detect active infections. That humans live in a viral world with an increasingly porous human–animal interface makes future zoonotic outbreaks a reality that must be prepared for. ∎

REFERENCES AND NOTES

1. R. K. French, E. C. Holmes, *Trends Microbiol.* **28**, 165 (2020).
2. W.-T. He *et al.*, *Cell* 10.1016/j.cell.2022.02.014 (2022).
3. K. E. Jones *et al.*, *Nature* **451**, 990 (2008).
4. K. J. Olival *et al.*, *Nature* **546**, 646 (2017).
5. M. Wille, J. L. Geoghegan, E. C. Holmes, *PLOS Biol.* **19**, e3001135 (2021).
6. M. Letko, S. N. Seifert, K. J. Olival, R. K. Plowright, V. J. Munster, *Nat. Rev. Microbiol.* **18**, 461 (2020).
7. A. T. Irving, M. Ahn, G. Goh, D. E. Anderson, L.-F. Wang, *Nature* **589**, 363 (2021).
8. J. A. Lednicky *et al.*, *Nature* **600**, 133 (2021).
9. J. C. Chandler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2114828118 (2021).
10. M. F. Boni *et al.*, *Nat. Microbiol.* **5**, 1408 (2020).
11. J. L. Geoghegan, A. M. Senior, F. Di Giallonardo, E. C. Holmes, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4170 (2016).
12. Centers for Disease Control and Prevention (CDC), *MMWR Morb. Mortal. Wkly. Rep.* **52**, 986 (2003).
13. J. Damas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 22311 (2020).
14. Carbis Bay G7 Summit Communiqué, Our shared agenda for global action to build back better, https://www.g7uk.org/wp-content/uploads/2021/06/Carbis-Bay-G7-Summit-Communique-PDF-430KB-25-pages-3.pdf (2021).
15. G. J. Xu *et al.*, *Science* **348**, aaa0698 (2015).

REVIEW

The changing epidemiology of SARS-CoV-2

Katia Koelle[1]*, Michael A. Martin[1,2], Rustom Antia[1], Ben Lopman[3,4], Natalie E. Dean[3,5]

We have come a long way since the start of the COVID-19 pandemic—from hoarding toilet paper and wiping down groceries to sending our children back to school and vaccinating billions. Over this period, the global community of epidemiologists and evolutionary biologists has also come a long way in understanding the complex and changing dynamics of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes COVID-19. In this Review, we retrace our steps through the questions that this community faced as the pandemic unfolded. We focus on the key roles that mathematical modeling and quantitative analyses of empirical data have played in allowing us to address these questions and ultimately to better understand and control the pandemic.

By February 2020, it was clear that a global coronavirus pandemic had begun to unfold. The rapid rise in COVID-19 cases before lockdown measures went into effect in Wuhan, China, on 23 January 2020 showed how efficient human-to-human transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was. By January 2020, the increasing number of documented cases of SARS-CoV-2 outside of China further indicated that regional containment of this virus was going to be extremely unlikely. At this time, epidemiologists and infectious disease modelers began to characterize the key features of the 2019 novel coronavirus (2019-nCoV)—subsequently renamed SARS-CoV-2 in February 2020—that were critical for assessing its containment potential and for projecting its spread at various geographic scales (Fig. 1).

The ability to contain an emerging pathogen depends in part on its basic reproduction number R_0 (1), defined as the average number of secondary cases arising from a primary case in a completely susceptible population (2). No active containment efforts are required when $R_0 < 1$, but higher values of R_0 make containment more difficult (2). Estimating R_0 for an emerging disease is not straightforward; doing so robustly requires calculating how quickly the virus is spreading through a population and having a measure of the time between one infection and the next, which is known as the generation interval (Fig. 2). Early epidemiological analyses therefore focused on estimating the rate of viral spread from reported Wuhan case data. The generation interval was parameterized using data from previous coronavirus outbreaks [SARS-CoV-1 and Middle East respiratory syndrome coronavirus (MERS-CoV)] and from the small number of transmission pairs that had been identified in early SARS-CoV-2 outbreaks. This led to estimates of R_0 ranging between 2 and 4 before the implementation of lockdown measures in Wuhan. The differences between these estimates were largely the result of differences in how the generation interval was parameterized (3), which underscores the importance as well as the scarcity of high-resolution outbreak datasets to accurately infer the generation interval and, thereby, the value of R_0.

> ## "...the ability to contain an emerging pathogen depends on the capacity to isolate infectious individuals."

Beyond R_0, the ability to contain an emerging pathogen depends on the capacity to isolate infectious individuals (1). This in turn depends on the proportion of infections that are asymptomatic or only mildly symptomatic (and therefore easily missed during surveillance) and their transmission potential. It also depends on the potential for presymptomatic transmission from individuals who are infectious before developing symptoms. Early on, it was unclear how common asymptomatic and mildly symptomatic infections were—i.e., how large the bottom of the infection "iceberg" might be. Epidemiological models were interfaced with reported infection data from China to infer the proportion of SARS-CoV-2 cases that went undocumented before 23 January 2020 (yielding an estimate of 86%) (4). In February and March of 2020, mass screening and surveillance of passengers aboard cruise ships and flights where outbreaks occurred yielded some of the first estimates of the proportion of infections that remained asymptomatic: 18 to 31% (5, 6). It was also apparent from early data that mild disease, most notably in younger age groups, was a common outcome. Crucially, contact tracing studies further revealed high infectiousness around the time of symptom onset, with frequent transmission from presymptomatic and mildly symptomatic individuals (7–9), and transmission also occurring (although less commonly) from fully asymptomatic individuals (10). These findings were sobering in terms of their implications for containing the pandemic. Unlike SARS-CoV-1 and MERS-CoV, SARS-CoV-2 was discerned from the start to have the potential to cause a global pandemic with tremendous public health impacts.

Flattening the curve

SARS-CoV-2 had spread to many regions of China and across the world by late February 2020. Although there were a few countries, such as New Zealand and Vietnam, that achieved almost complete control of the initial spread through rigorous surveillance efforts (11, 12), community transmission quickly took off in many other countries. Mathematical scenario models predicted the devastation that would result if no control measures were implemented (13, 14). Coupled with accounts coming out of northern Italy, London, New York City, and the other earliest-affected regions, projections from these models quickly led to government-imposed restrictions to curb viral spread (15). Nonpharmaceutical interventions (NPIs) included lockdowns, social distancing measures, and, eventually, the use of face masks. Adoption of NPIs during the first wave of the pandemic served to flatten the curve—that is, to extend the period over which cases occurred. Flattening the curve was desirable for three reasons. First, it would prevent the health care system from being overwhelmed because the peak number of beds occupied at any one time would be lower under a flatter curve. Second, it would slow the momentum of the outbreak, reducing the overshoot of cases after the outbreak peak (16). Third, it allowed time to improve clinical care strategies and capacity and to evaluate therapeutics.

With some NPI measures in place, modeling efforts turned toward forecasting the future dynamics of COVID-19. Some of the most public-facing forecasting that was reported in April and May of 2020 were the controversial projections made by the Institute for Health Metrics and Evaluation (IHME). These projections relied on statistical modeling approaches, which can provide accurate short-term projections but are likely to fail for longer-term predictions because they do not incorporate the process of transmission (17). Indeed, IHME projections at this time predicted little viral circulation by midyear 2020. Mechanistic epidemiological modeling studies instead predicted a slowdown of viral spread while tough NPI measures were in place but a resurgence of cases after NPI relaxation (14). As the first wave provided epidemiologists with longitudinal SARS-CoV-2 case data, COVID-19

[1]Department of Biology, O. Wayne Rollins Research Center, Emory University, Atlanta, GA 30322, USA. [2]Graduate Program in Population Biology, Ecology, and Evolution, Emory University, Atlanta, GA 30322, USA. [3]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA. [4]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA. [5]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA.
*Corresponding author. Email: katia.koelle@emory.edu

death data, and virus sequence data, the effect that NPIs had on reducing viral spread started to be quantitatively assessed. These analyses found that NPIs were effective at flattening the curve and even at temporarily reducing reproduction numbers to the extent that circulation levels declined (*18, 19*). These findings mirrored those from studies of historical pandemics, such as the 1918 H1N1 influenza pandemic, which found that NPIs were successful at reducing disease transmission (*20*). Counterfactual models were used to quantify how many cases and deaths could have been averted by different policies, such as earlier social distancing measures (*21*).

Although NPIs were effective at reducing viral spread at the population level, infection risk was not reduced equally across populations. Health care workers and frontline workers (e.g., first responders) were at particularly high infection risk in April and May of 2020 owing to their close contact with SARS-CoV-2–positive individuals. Many other essential occupations were also found to carry increased risk, including public-facing jobs, such as taxi and bus drivers, and jobs in poorly ventilated and crowded settings, such as factories and meatpacking plants. Crowded living settings, where social distancing was not possible, also carried increased risks for infection (*22*). Even with NPIs in place, large outbreaks occurred in communal-living facilities, such as nursing homes, homeless shelters, jails, and prisons (*23*). Risks were exacerbated by poor sanitation and inadequate health care. Sharp differences in case rates also became clearly visible across socioeconomic and racial and ethnic groups (*24*).

As anticipated, the widespread use of NPIs, where implemented, exacted considerable social and economic costs. To facilitate the reopening of the economy, researchers sought to distinguish highly effective NPIs from those that could potentially be discarded. Key to assessing the effectiveness of different NPIs was understanding the routes of viral transmission. How easily did the virus spread between individuals through contaminated surfaces (fomites) compared with direct transmission by aerosols or droplets? Although an early study from April 2020 found that SARS-CoV-2 could remain viable on surfaces, such as plastic, for days (*25*), subsequent studies indicated that transmission occurred primarily by direct transmission from infected individuals rather than through fomites (*26*). Contact tracing studies also revealed the higher risk of direct transmission occurring in closed and crowded places and close-contact settings (*27*), known by Japanese public health officials as the three Cs. Large outbreaks were traced to nightclubs, karaoke bars, and choir practices—all indoor settings characterized by loud vocalization. Statistical analyses found that limitations on public gatherings and workplace closures were positively associated with lower levels of SARS-CoV-2 spread (*28, 29*), corroborating the role of direct transmission in closed areas in facilitating the transmission of SARS-CoV-2.

Large outbreaks detected by contact tracing, coupled with emerging virus sequence data, quickly led to the recognition of substantial transmission heterogeneity between individuals. Studies using data from the first wave found that ~10% of cases were responsible for 80% of secondary infections (*19, 30*), a relatively high level of heterogeneity when compared with other human pathogens (*31*). Explanations for superspreading dynamics ranged from variable contact patterns across individuals (*32*); high-risk settings conducive to spread (*23*); and biological factors, such as high viral load (*32–34*). Notably, if superspreading individuals or events could be systematically identified, then control efforts could plausibly focus on mitigating spread in a more targeted manner. An influential review of the efficacy of face masks also indicated that mask wearing had an appreciable effect on reducing transmissibility per contact by reducing transmission of infected respiratory particles (*35*). Mask wearing, particularly by infected individuals with high viral loads (i.e., source control), therefore had the potential to reduce viral spread among the population.

Epidemiological modeling also pointed toward ways in which surveillance and control efforts could be improved upon to allow for relaxation of broadly applied NPIs. For example, modeling studies showed that traditional contact tracing and digital contact tracing had the potential to slow viral spread in the presence of relaxed lockdown measures (*36, 37*). Another study found that increasing testing frequency and reporting speed for SARS-CoV-2 would be more effective at limiting the spread of the virus than exclusively relying on highly sensitive tests with slower reporting timelines (*38, 39*). Several modeling studies highlighted the trade-offs involved for different timelines of NPI relaxation and how uncertainty in NPI
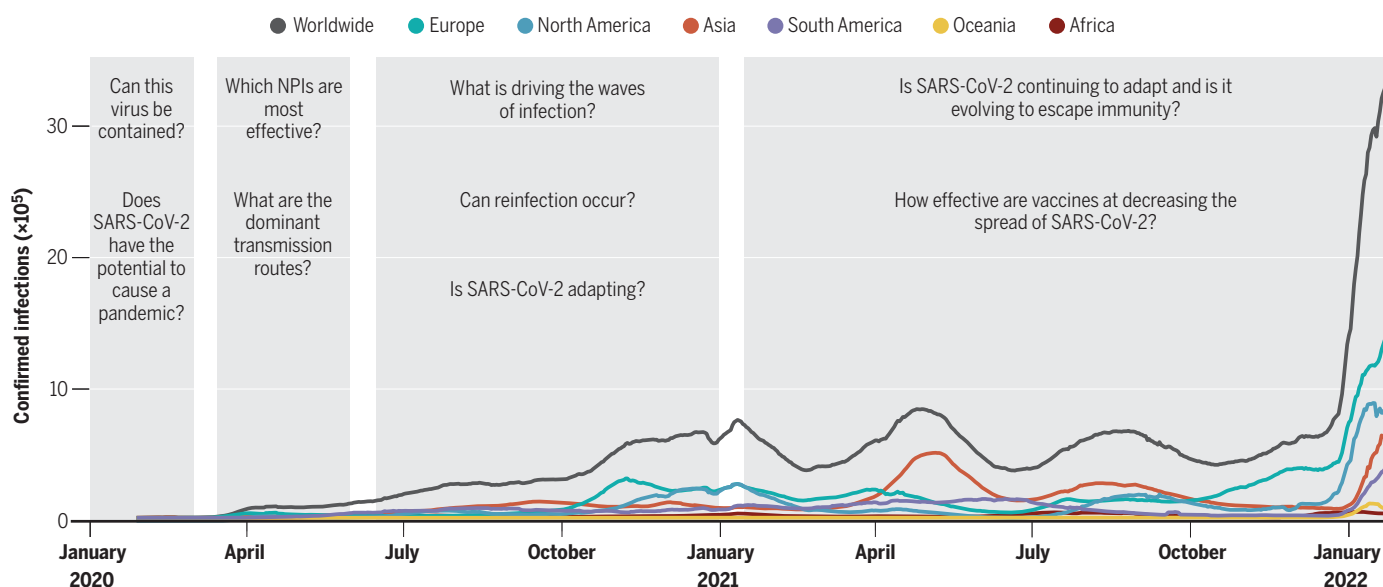

Fig. 1. Timeline of changing epidemiological questions asked as the SARS-CoV-2 pandemic unfolded. Confirmed infections are shown by continent as well as on a worldwide scale. Gray bars indicate the time periods discussed in this Review (the emerging pandemic, flattening the curve, riding out the waves, and vaccines and variants).

effects and implementation errors may thwart our ability to optimally control disease spread through the use of NPIs (*40*, *41*).

Riding out the waves

The initial wave of infections was followed by subsequent waves, reminiscent of those seen in the years after the 1918 influenza pandemic (*42*). Factors that contribute to waves of transmission include evolutionary changes in the virus as well as changes in host immunity and behavior.

Theory would predict that soon after the introduction of a zoonotic pathogen, virus adaptation would result in increased transmission between humans. The rapid spread of the D614G variant [with a glycine residue (G) at position 614 in the spike protein replacing an aspartic acid residue (D)] between March and May of 2020 pointed toward the possibility of virus adaptation occurring (*43*) (Fig. 3A). That this replacement reflected virus adaptation was supported by experiments showing that the variant had enhanced replication in human bronchial and nasal airway epithelial cell cultures as well as increased replication and transmissibility in in vivo infections of hamsters and ferrets (*44*). Quantitative analyses of population-level SARS-CoV-2 sequence data from the UK also detected a transmission advantage of D614G in humans (*45*).

As time progresses during a pandemic, immunity from natural infection builds up in a population. For SARS-CoV-2, the initial hope was that population immunity in areas with large first waves would greatly reduce future transmission. The level of herd immunity required for protecting a population is roughly $1 - 1/R_0$ (*46*). Based on an R_0 of 3, this yielded an estimate of 67% for SARS-CoV-2, but transmission heterogeneity was theorized to significantly lower the herd immunity threshold (*47*). However, as early as August 2020, documented accounts of reinfection indicated that immunity to SARS-CoV-2 may only transiently protect from infection (Fig. 3B) (*48*). Studies of the endemic human coronaviruses (HKU1, OC43, 229E, and NL63) at this time

> ## "Vaccine-related questions broadened to ones regarding the extent to which vaccination could reduce transmission."

also found evidence for repeated reinfections (*49*) and for antigenic evolution (*50*, *51*) (Fig. 3C), which put a damper on hopes that herd immunity would rapidly bring an end to the pandemic.

Throughout the second half of 2020, as we were gaining a better understanding of the virus's transmission characteristics and the duration of immunity, SARS-CoV-2 cases ebbed and flowed, with waves peaking at different times in different countries and even in different regions in the same country. The causes of

these temporal changes in incidence were likely the result of a combination of factors, including changes in NPIs, behavioral changes, seasonal changes (influenced by weather and holiday schedules), and the development of transmission-reducing immunity through natural infection followed by its waning. Quantifying the interactions between these factors and their relative roles remains an important area for study, critical for projecting the longer-term dynamics of SARS-CoV-2 (*52*, *53*).

Vaccines and variants

By December 2020, vaccines against SARS-CoV-2 were developed, trialed, and approved for emergency use—a timeline unparalleled in the history of vaccinology. Epidemiological factors facilitated the accelerated timeline of these pivotal trials. With little acquired immunity to SARS-CoV-2 and a high incidence of COVID-19 in mid-2020 when trials were underway, the studies reached their prespecified end points rapidly and ahead of anticipated schedule. As is widely appreciated now, the trials demonstrated excellent protection against COVID-19 and severe outcomes, like hospitalization and death. But questions remained even after the vaccines were authorized for use (Fig. 4): How long will protection last? Will the vaccines protect against new virus variants? Will they prevent transmission? Although randomized trials continued to accrue data for a period of several months after their completion, observational studies, including cohorts and case-control designs, were critical for monitoring
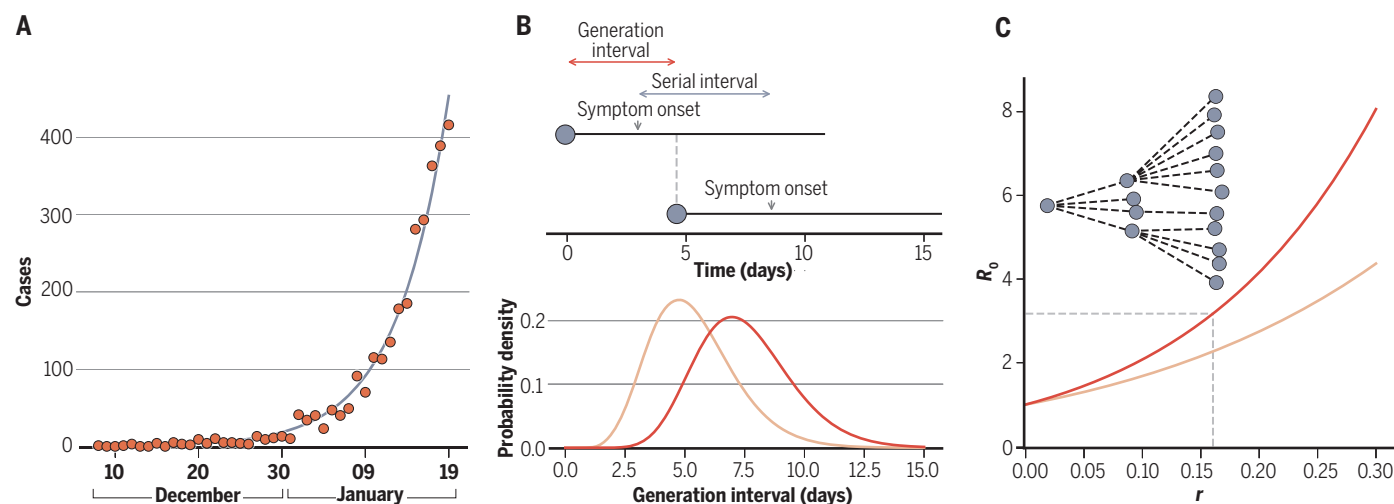
Fig. 2. Estimation of the basic reproduction number R_0 from case data and detailed outbreak data. (**A**) Exponential growth models are fit to observed case data to estimate the speed at which a virus spreads through a population. For SARS-CoV-2 spread in Wuhan, early estimates of the exponential growth rate r fell in the range of 0.10 to 0.20 per day (*3*), yielding epidemic doubling times of ~3.5 to 7 days. (**B**) Outbreak data are used to estimate the viral serial interval, defined as the time between symptom onset of an index case and symptom onset of the index case's contacts. This serial interval is often used as an approximation for the viral generation interval. For SARS-CoV-2, an early Wuhan estimate of the mean serial interval was 7.5 days (*86*) (distribution in red). An example of a generation interval distribution with a smaller mean is shown in light orange. (**C**) The basic reproduction number R_0 can be calculated from the exponential growth rate r and the distribution of the generation interval using an equation derived from demographic analyses (*87*). Red and light orange curves, corresponding to the distributions in (B), show how the R_0 estimates depend on the generation interval. The transmission chain at the top illustrates an outbreak with an $R_0 = 3$ pathogen.

real-world vaccine effectiveness along these different dimensions.

Many decisions about the vaccine rollout, including policy-relevant questions such as which groups to target initially, had to be made before such knowledge was available, so epidemiological modeling served a key role in exploring the potential impact of decisions under different assumptions. One initial question was whether to maximize coverage by offering a single dose of a two-dose series to as many people as possible or to maintain fidelity to the trial protocols of two doses (spaced 3 to 4 weeks apart for the mRNA vaccines). Modeling studies, parameterized with trial data that showed apparent protection starting 10 to 14 days after receipt of a first dose, indicated that in the short term, a one-dose strategy is likely preferable for limiting disease at the level of the population (*54*). However, in the longer term, if vaccine efficacy is substantially lower with one dose than with two, a two-dose strategy might be preferable, depending on how rapidly the virus was infecting people and how quickly the vaccine could be rolled out. Concerns about the potential for SARS-CoV-2 to evolve immune escape in the context of incomplete immunity from a single dose was also addressed through quantitative modeling and reasoning (*54, 55*). On the basis of early trial data and modeling results, the UK promoted a strategy whereby second doses were delayed compared with the trial interval. In doing so, high coverage was achieved rapidly among >50 year olds, and a decoupling became apparent in which high case counts did not result in parallel hospitalization or mortality trends in postvaccination pandemic waves. This observation, predicted by scenario modeling studies (*56*), showed the value of vaccines in preventing severe disease. Indeed, postintroduction evaluations indicated that vaccines offered ~90% protection against hospitalization and death (*57, 58*).

Vaccine-related questions broadened to ones regarding the extent to which vaccination could reduce transmission (Fig. 4). Vaccines may avert transmission by two mechanisms—by preventing infection and by rendering breakthrough cases less infectious. Landmark observational studies from mid- to late 2021 demonstrated protection against infection, although to a lesser degree than against disease or severe outcomes (*58, 59*). Some studies also indicated that vaccinated individuals experiencing breakthrough infections may have lower viral loads or a more rapid viral decline and thus perhaps less potential for onward transmission (*60–62*) [but see (*63, 64*)]. The most holistic understanding of vaccine performance against transmission comes from household studies that offer insight into infection-blocking and transmission-blocking mechanisms. One such study estimated vaccine efficacy against infectiousness of breakthrough cases to be 23% (*65*). When combined with protection against infection, the reduction in transmission was estimated to be 92%.

Complicating the assessment of SARS-CoV-2 vaccine efficacy was the evolution of new variants of concern around the time that vaccines were rolled out. In late 2020, the B.1.1.7 lineage [now known as Alpha in the World Health Organization's (WHO's) variant naming scheme] was detected and rapidly spread in the UK. Several of the mutations harbored by this
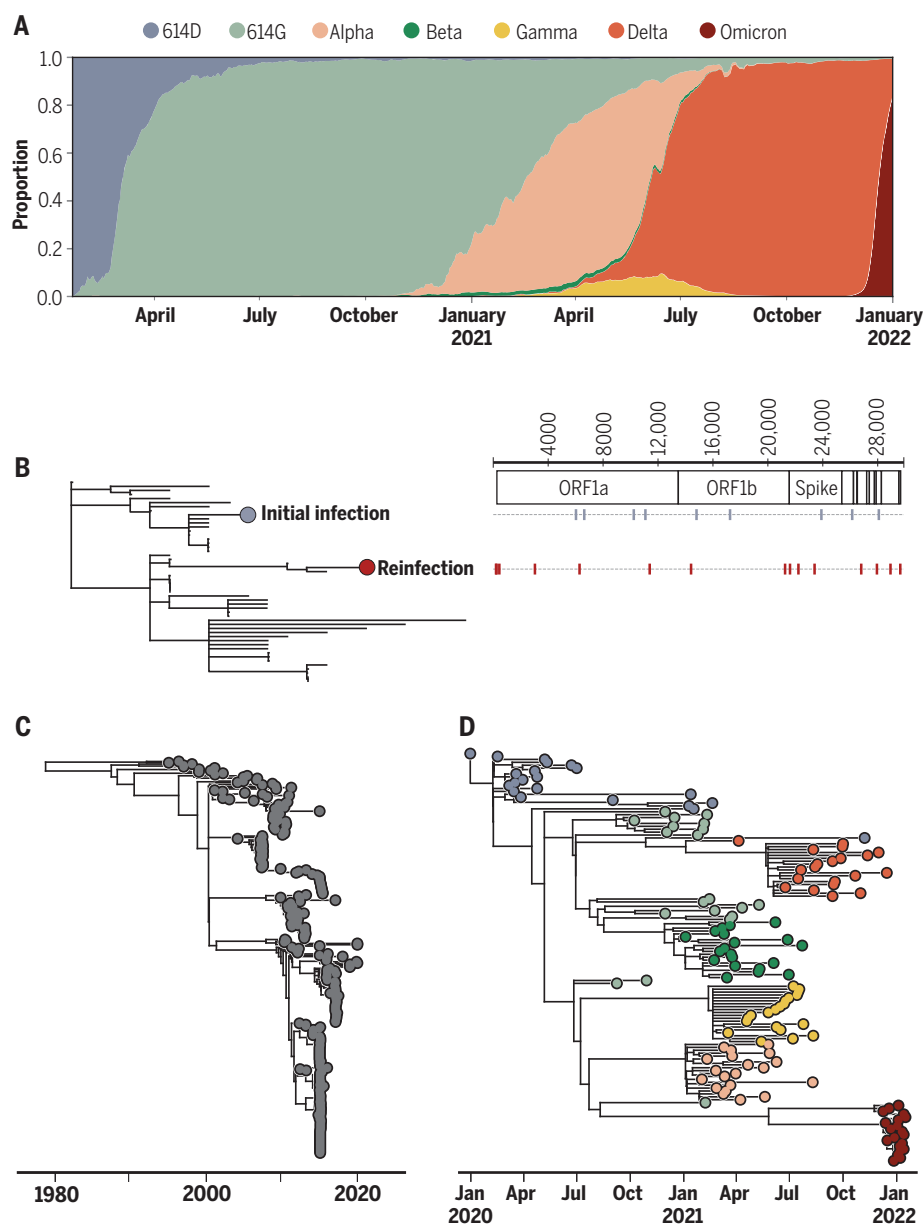
Fig. 3. Coronavirus sequence data have informed epidemiological understanding of SARS-CoV-2.
(**A**) The frequencies of SARS-CoV-2 variants of concern over time. The first indication that SAR-COV-2 was adapting to humans was the population-level replacement of the 614D allele with the 614G allele in early 2020. Variant frequencies on the y axis are calculated on the basis of SARS-CoV-2 sequence data deposited in GISAID (Global Initiative for Sharing Avian Influenza Data). Only major, globally circulating variant lineages are shown. (**B**) An example of viral sequencing that has provided evidence of reinfection with SARS-CoV-2. Reproduced with permission from (*48*). Samples from a patient in the context of circulating viral genetic variation indicate that the observed secondary infection is not a reactivation of a latent infection but is instead a reinfection. Schematic on the right shows the substitutions present in the primary and secondary infection viral samples. ORF, open reading frame. (**C**) A phylogeny of seasonal human coronavirus OC43 (lineage A). Phylogenetic analysis points toward antigenic evolution in this viral population. Reproduced with permission from (*50*). (**D**) A phylogeny inferred from SARS-CoV-2 sequence data showing the evolutionary relationships between the variant lineages included in (A).
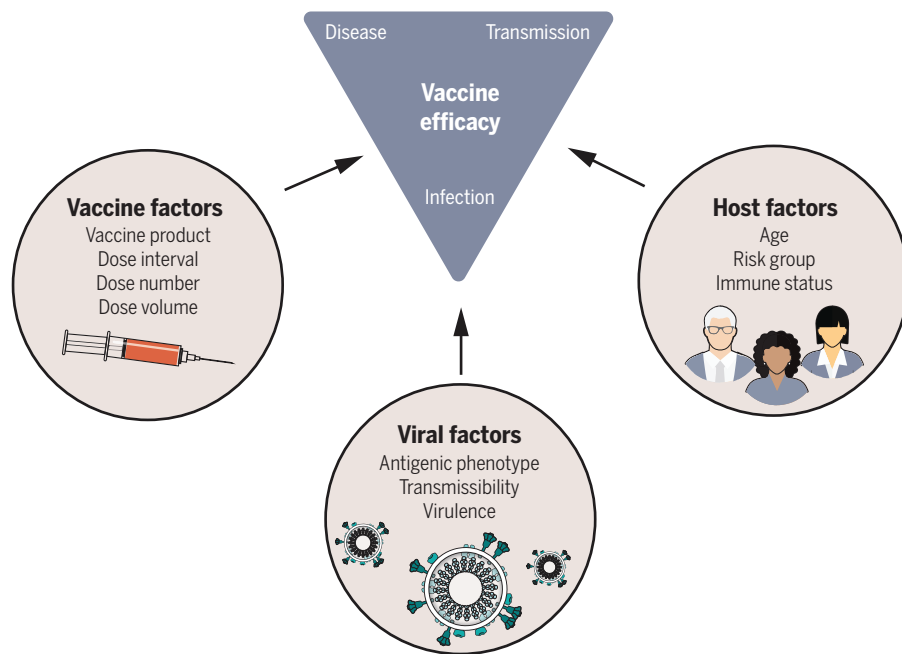
Fig. 4. Vaccine, host, and viral factors that affect vaccine efficacy. Each of the factors shown were investigated as potential moderators of vaccine efficacy in clinical trials and postintroduction observational studies. Scenario modeling projected the population-level effects of vaccination by incorporating data-driven assumptions on how vaccination affects susceptibility to infection, disease, and severe outcomes like hospitalization and death and how these effects differed by host factors, such as age. A key determinant of population impact of vaccination is the effect on transmission, which is a combination of infection-blocking and transmission-reducing effects of vaccines. As data emerged, mainly from observational studies, assumptions about the infectiousness of vaccine breakthrough cases were incorporated into models. With the evolution of Delta and then Omicron, variant-specific vaccine efficacy and the potential for viral immune escape became critical quantities to understand. With this array of vaccine, host, and viral data on the effects of vaccination on vaccine recipients, epidemiological models could project both the direct and indirect effects of vaccination on the population-level spread of SARS-CoV-2.

lineage significantly altered the viral phenotype, including enhanced binding to angiotensin-converting enzyme 2 (ACE2) (66), with consequences for transmissibility (67). Quantification of Alpha's selective advantage based on viral sequence data indicated that it had 1.5 to 2 times the reproduction number of other lineages circulating at that time (68).

Soon after the emergence of the Alpha lineage, a different viral lineage, B.1.617.2 (which, along with its descendant lineages, make up the Delta variant of concern), was fueling a large wave of SARS-CoV-2 infections in India (69). This lineage rapidly swept through India and became dominant worldwide. In vitro analyses indicated that convalescent human serum bound less efficiently to Delta than to ancestral lineages and that Delta had higher replication efficiency (70). Consistent with these findings, a modeling study found that Delta's growth advantage at the population level stemmed from a combination of Delta's higher transmissibility and its reduced sensitivity to host immune responses generated through prior infection (69). Fortunately, postintroduc-

tion vaccine effectiveness studies found that protection against symptomatic disease was largely maintained against both the Alpha and Delta variants (58, 71, 72). Subsequent studies clearly established that vaccine effectiveness against infection and symptomatic disease wane over time (73) such that additional vaccine doses (boosters) can be beneficial, particularly to older-aged individuals (74).

In November 2021, researchers in South Africa noticed a rising number of COVID-19 cases with S gene target failure (SGTF), indicative of an expanding viral subpopulation harboring the Δ69-70 spike gene deletion. Shortly thereafter, circulation of a new variant was confirmed through whole-genome sequencing. Based on its mutational profile and its extremely rapid spread in South Africa, the WHO designated this variant as one of concern, naming it Omicron (B.1.1.529). In December 2021, this variant spread rapidly throughout the world, resulting in more confirmed infections than ever before (Fig. 1) and rapidly replacing the Delta variant (Fig. 3A). Omicron's mutational profile relative to the

Wuhan-Hu-1 reference genotype consists of more than 30 mutations in the spike glycoprotein (75), many of which are predicted to enable immune escape. Early studies confirmed the ability of Omicron to evade antibodies from previous infection and vaccination (76) and indicated that it can use an alternative cell entry pathway, which improves its ability to infect cells in the upper respiratory tract (77). Animal models further indicated that infection with Omicron may be less clinically severe than infection with previous variants (78). Experimental predictions have been borne out in epidemiological studies, in which Omicron has been shown to be better at reinfecting previously infected people (79) and less clinically severe compared with Delta (80). Nevertheless, the rapid expansion of Omicron has resulted in the largest spike in confirmed SARS-CoV-2 infections thus far in many countries, and spikes in deaths attributed to the virus have followed. Vaccine effectiveness against severe disease and death, however, has remained high with the new variant, with Omicron-caused deaths concentrated among unvaccinated individuals.

The emergence of Omicron, and Alpha before it, raises important evolutionary questions regarding the sources of new, genetically divergent variants. At their time of emergence, both variants harbored more mutations than would have been expected given the rate at which SARS-CoV-2 evolves in the human population. Omicron's distant evolutionary relationship to other SARS-CoV-2 variants is notable (Fig. 3D). One possibility for the evolutionary origin of these variants is in an alternative host: SARS-CoV-2 has been observed in a number of domestic and wild animal species (as a result of reverse zoonotic events), and these new variants may have evolved in one of these alternative hosts and spilled back into the human population. Alternatively, these variants may have evolved in immunocompromised or immunosuppressed individuals with chronic SARS-CoV-2 infections, given documented examples of the extent to which adaptive evolution can happen in these types of infection (81). Both putative sources of new variants remain unobserved in human population surveillance sequencing data, which highlights the need for more comprehensive monitoring strategies to identify variants of concern earlier and to identify and stop potential variants of concern before they are widespread.

Looking ahead

SARS-CoV-2 has been difficult to control since the start of the pandemic, and the recent spread of Omicron has only amplified this challenge. It is inevitable that SARS-CoV-2 will become an endemic pathogen in the human population. However, its impact on our health and daily lives will change considerably over time (82). Governments and public health

officials must transition to a long-term strategy of dealing with SARS-CoV-2. Vaccines and therapeutics, including outpatient treatment to prevent progression of disease, are highly valuable tools. Yet the challenge ahead remains formidable, with transmission levels high even in areas with widespread vaccination and prior waves of infection, and many open questions remain about the future epidemiological impact of this virus.

The most evolutionarily successful variants of SARS-CoV-2 so far have evolved greater transmissibility relative to more ancestral variants. Although higher transmissibility will result in more infections, increasing population-level immunity from infections and vaccinations might make these infections less severe in terms of health impact. However, the evolutionary pressures facing the virus, and thus its patterns of adaptive evolution, will change. Variants, such as Omicron, that gain much of their transmission advantage from evading immunity may become the norm, as is the case with seasonal influenza.

The evolution of SARS-CoV-2 virulence in terms of how harmful or deadly it is has been a topic of debate since the beginning of the pandemic (83). Evolutionary theory has pointed out that we should not expect evolution toward lower virulence (84), and the last two variants of concern have demonstrated that there is not a clear, consistent trend in SARS-CoV-2 virulence evolution: Although Delta is thought to be slightly more virulent than previous variants, Omicron is less so. Although the virulence of SARS-CoV-2 may still evolve over time (in a direction that is not easily predicted), we expect the infection fatality ratio to decline for other reasons, including rising population immunity. As noted above, the available SARS-CoV-2 mRNA vaccines confer protection against severe disease, even where breakthrough infections occur. Vaccine boosters will further help in protecting against severe disease. Monoclonal antibodies and antivirals to be taken shortly after symptom onset will similarly reduce the risk of severe disease and death. Over time, population immunity will result in a younger age at first infection, which will reduce average infection severity and likely lead to milder infections later in life. Parallels can be drawn to the seasonal coronaviruses, where individuals are usually infected very early on in life and where secondary infections are milder than primary infections because of preexisting immunity (85).

Stepping back

The most pressing scientific questions have changed over the past 2 years, reflecting the complexity of this pandemic (Fig. 1). Early questions focused on placing the novel pathogen on a map with respect to its pathogenicity and transmissibility. Modeling assumptions relied heavily upon experiences with SARS-CoV-1 and MERS-CoV and on a limited number of SARS-CoV-2 datasets. Over time, we replaced broad assumptions with improved understanding. The rapid spread of the virus and the explosive waves of infection that ensued necessitated swift reactions, and questions shifted from short-term to longer-term considerations. We made a major leap forward with the deployment of safe and effective vaccines, but new variants and waning immunity dampened our hopes for greater control of transmission. Moving forward, key questions remain about the propensity of this virus to further adapt to our human population and the health impact of endemic transmission. Although our work as scientists studying this pathogen remains far from over, we recognize all that has been achieved to support the unifying mission of a safer and healthier world.

REFERENCES AND NOTES

1. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6146–6151 (2004).
2. R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, 2010).
3. S. W. Park et al., *J. R. Soc. Interface* **17**, 20200144 (2020).
4. R. Li et al., *Science* **368**, 489–493 (2020).
5. H. Nishiura et al., *Int. J. Infect. Dis.* **94**, 154–155 (2020).
6. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, *Euro Surveill.* **25**, 2000180 (2020).
7. X. He et al., *Nat. Med.* **26**, 672–675 (2020).
8. M. A. Johansson et al., *JAMA Netw. Open* **4**, e2035057 (2021).
9. T. Ganyani et al., *Euro Surveill.* **25**, 2000257 (2020).
10. X. Qiu et al., *Clin. Microbiol. Infect.* **27**, 511–519 (2021).
11. L. Van Tan, *Nat. Immunol.* **22**, 261 (2021).
12. M. G. Baker, N. Wilson, A. Anglemyer, *N. Engl. J. Med.* **383**, e56 (2020).
13. J. T. Wu, K. Leung, G. M. Leung, *Lancet* **395**, 689–697 (2020).
14. P. G. T. Walker et al., Imperial College COVID-19 Response Team, "Report 12: The global impact of COVID-19 and strategies for mitigation and suppression" (Imperial College London, 2020) .
15. D. Adam, *Nature* **580**, 316–318 (2020).
16. S. Cobey, *Science* **368**, 713–714 (2020).
17. I. Holmdahl, C. Buckee, *N. Engl. J. Med.* **383**, 303–305 (2020).
18. S. Flaxman et al., *Nature* **584**, 257–261 (2020).
19. D. Miller et al., *Nat. Commun.* **11**, 5518 (2020).
20. R. J. Hatchett, C. E. Mecher, M. Lipsitch, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7582–7587 (2007).
21. S. Pei, S. Kandula, J. Shaman, *Sci. Adv.* **6**, eabd6370 (2020).
22. J. T. Chen, N. Krieger, *J. Public Health Manag. Pract.* **27**, S43–S56 (2021).
23. C. Franco-Paredes et al., *PLOS Negl. Trop. Dis.* **14**, e0008409 (2020).
24. N. Bhala, G. Curry, A. R. Martineau, C. Agyemang, R. Bhopal, *Lancet* **395**, 1673–1676 (2020).
25. N. van Doremalen et al., *N. Engl. J. Med.* **382**, 1564–1567 (2020).
26. E. A. Meyerowitz, A. Richterman, R. T. Gandhi, P. E. Sax, *Ann. Intern. Med.* **174**, 69–79 (2021).
27. H. Nishiura et al., medRxiv 2020.02.28.20029272 [Preprint] (2020).
28. Y. Li et al., *Lancet Infect. Dis.* **21**, 193–202 (2021).
29. J. M. Brauner et al., *Science* **371**, eabd9338 (2021).
30. A. Endo, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, S. Abbott, A. J. Kucharski, S. Funk, *Wellcome Open Res.* **5**, 67 (2020).
31. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, *Nature* **438**, 355–359 (2005).
32. A. Goyal, D. B. Reeves, E. F. Cardozo-Ojeda, J. T. Schiffer, B. T. Mayer, *eLife* **10**, e63537 (2021).
33. T. C. Jones et al., *Science* **373**, eabi5273 (2021).
34. Q. Yang et al., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104547118 (2021).
35. J. Howard et al., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014564118 (2021).
36. A. Aleta et al., *Nat. Hum. Behav.* **4**, 964–971 (2020).
37. L. Ferretti et al., *Science* **368**, eabb6936 (2020).
38. D. B. Larremore et al., *Sci. Adv.* **7**, eabd5393 (2021).
39. M. J. Mina, R. Parker, D. B. Larremore, *N. Engl. J. Med.* **383**, e120 (2020).
40. D. H. Morris, F. W. Rossine, J. B. Plotkin, S. A. Levin, *Commun. Phys.* **4**, 78 (2021).
41. T. A. Perkins, G. España, *Bull. Math. Biol.* **82**, 118 (2020).
42. D. R. Olson, L. Simonsen, P. J. Edelson, S. S. Morse, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11059–11063 (2005).
43. B. Korber et al., *Cell* **182**, 812–827.e19 (2020).
44. B. Zhou et al., *Nature* **592**, 122–127 (2021).
45. E. Volz et al., *Cell* **184**, 64–75.e11 (2021).
46. P. E. Fine, *Epidemiol. Rev.* **15**, 265–302 (1993).
47. T. Britton, F. Ball, P. Trapman, *Science* **369**, 846–849 (2020).
48. K. K.-W. To et al., *Clin. Infect. Dis.* **73**, e2946–e2951 (2021).
49. A. W. D. Edridge et al., *Nat. Med.* **26**, 1691–1693 (2020).
50. K. E. Kistler, T. Bedford, *eLife* **10**, e64509 (2021).
51. R. T. Eguia et al., *PLOS Pathog.* **17**, e1009453 (2021).
52. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, *Science* **368**, 860–868 (2020).
53. C. M. Saad-Roy et al., *Science* **370**, 811–818 (2020).
54. C. M. Saad-Roy et al., *Science* **372**, 363–370 (2021).
55. S. Cobey, D. B. Larremore, Y. H. Grad, M. Lipsitch, *Nat. Rev. Immunol.* **21**, 330–335 (2021).
56. K. M. Bubar et al., *Science* **371**, 916–921 (2021).
57. S. Y. Tartof et al., *Lancet* **398**, 1407–1416 (2021).
58. P. Tang et al., *Nat. Med.* **27**, 2136–2143 (2021).
59. G. Regev-Yochay et al., *Lancet Reg. Health Eur.* **7**, 100150 (2021).
60. R. Ke et al., medRxiv 2021.08.30.21262701 [Preprint] (2021).
61. M. C. Shamier et al., medRxiv 2021.08.20.21262158 [Preprint] (2021).
62. P. Y. Chia et al., *Clin. Microbiol. Infect.* 10.1016/j.cmi.2021.11.010 (2021).
63. K. K. Riemersma et al., medRxiv 2021.07.31.21261387 [Preprint] (2021).
64. S. M. Kissler et al., *N. Engl. J. Med.* **385**, 2489–2491 (2021).
65. O. Prunas et al., *Science* **375**, 1151–1154 (2022).
66. T. N. Starr et al., *Cell* **182**, 1295–1310.e20 (2021).
67. N. G. Davies et al., *Science* **372**, eabg3055 (2021).
68. E. Volz et al., *Nature* **593**, 266–269 (2021).
69. M. S. Dhar et al., *Science* **374**, 995–999 (2021).
70. P. Mlcochova et al., *Nature* **599**, 114–119 (2021).
71. S. Nasreen et al., medRxiv 2021.06.28.21259420 [Preprint] (2021).
72. J. Lopez Bernal et al., *N. Engl. J. Med.* **385**, 585–594 (2021).
73. H. Chemaitelly et al., *N. Engl. J. Med.* **385**, e83 (2021).
74. Y. M. Bar-On et al., *N. Engl. J. Med.* **385**, 1393–1400 (2021).
75. R. Viana et al., *Nature* 10.1038/s41586-022-04411-y (2022).
76. L. Liu et al., *Nature* 10.1038/s41586-021-04388-0 (2021).
77. T. P. Peacock et al., bioRxiv 2021.12.31.474653 [Preprint] (2022).
78. E. G. Bentley et al., bioRxiv 2021.12.26.474085 [Preprint] (2021).
79. J. R. C. Pulliam et al., medRxiv 2021.11.11.21266068 [Preprint] (2021).
80. J. A. Lewnard et al., medRxiv 2022.01.11.22269045 [Preprint] (2022).
81. T. G. Maponga et al., SSRN 4014499 [Preprint] (2022).
82. R. Antia, M. E. Halloran, *Immunity* **54**, 2172–2176 (2021).
83. S. Alizon, M. T. Sofonea, *J. Evol. Biol.* **34**, 1867–1877 (2021).
84. T. Day, S. Gandon, S. Lion, S. P. Otto, *Curr. Biol.* **30**, R849–R857 (2020).
85. J. S. Lavine, O. N. Bjornstad, R. Antia, *Science* **371**, 741–745 (2021).
86. Q. Li et al., *N. Engl. J. Med.* **382**, 1199–1207 (2020).
87. J. Wallinga, M. Lipsitch, *Proc. R. Soc. B* **274**, 599–604 (2007).

REVIEW

The immunology and immunopathology of COVID-19

Miriam Merad[1]*, Catherine A. Blish[2]*, Federica Sallusto[3,4]*, Akiko Iwasaki[5,6]*

Considerable research effort has been made worldwide to decipher the immune response triggered upon severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections, identify the drivers of severe and fatal COVID-19, and understand what leads to the prolongation of symptoms after disease resolution. We review the results of almost 2 years of COVID-19 immunology research and discuss definitive findings and remaining questions regarding our understanding of COVID-19 pathophysiology. We discuss emerging understanding of differences in immune responses seen in those with and without Long Covid syndrome, also known as post-acute sequelae of SARS-CoV-2. We hope that the knowledge gained from this COVID-19 research will be applied in studies of inflammatory processes involved in critical and chronic illnesses, which remain a major unmet need.

In December 2019, the Chinese Center for Disease Control and Prevention reported a cluster of pneumonia cases with unknown cause. The first complete sequence of the novel beta coronavirus genome was then identified in January 2020 (1). Initially called 2019-nCoV (for 2019 novel coronavirus), the virus was renamed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the disease caused by the virus as coronavirus disease 2019 (COVID-19). COVID-19 manifests with symptoms ranging from fully asymptomatic to severe disease and death (Table 1). The most common manifestations are fever, cough, and shortness of breath; additional common symptoms include fatigue, myalgias, nausea, vomiting, diarrhea, headache, weakness, rhinorrhea, anosmia, and ageusia (2, 3). A wide range of complications can lead to severe illness and death, including pneumonia, acute respiratory distress syndrome, liver injury, cardiac injury, thrombosis including stroke, renal disease, neurologic disease, and sepsis (2–7).

As the virus evolves and new variants emerge, there have been concerns that such variants could increase pathogenesis by escaping from immunity generated through previous infection or vaccination or by inducing more severe disease. Some variants of concern (VOC), such as 1.351 (Beta), P.1 (Gamma), and the recently described B.1.1.529 (Omicron), have mutations that render them less susceptible to vaccine-mediated and infection-acquired immunity. It is less apparent whether some variants induce more severe disease upon primary infection than others, though strains such as B.1.1.7 (Alpha) and B.1.617.2 (Delta) are known

to spread more efficiently, making it difficult to distinguish increased infection rates from increased severity. As new variants emerge, it will be important to direct continued research efforts into identifying how such variants escape from either innate or adaptive immune responses. Here, we review progress in our understanding of the immunology and immunopathology of COVID-19.

Innate immunity to SARS-CoV-2

Coronaviruses are enveloped viruses with a positive-sense RNA-strand genome of ~30,000 bases. SARS-CoV-2 is a member of the *Betacoronavirus* genus, which also includes severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). The innate immune system uses various pattern recognition receptors (PRRs) to detect the presence of viruses. RNA viruses such as coronaviruses are detected through two separate strategies (8). In one strategy, specialized immune cells such as plasmacytoid dendritic cells (pDCs) detect the incoming viral genomic RNA in the endosome through Toll-like receptor 7 (TLR7). Other cell types express endosomal TLR3 (variety of cells) and TLR8 (myeloid cells) that can also recognize endocytosed double-stranded RNA (dsRNA) or single-stranded RNA (ssRNA), respectively. Alveolar macrophages survey the lumen of the respiratory tract and provide the first line of defense. The other strategy involves viral recognition within infected cells. During viral replication dsRNA intermediates can be recognized by cytosolic RNA sensors such as RIG-I and MDA5, or RIG-I–like receptors (RLRs). Upon engagement of TLRs and RLRs, signaling downstream activates the IRF3/IRF7-dependent transcription of type I and type III interferons (IFNs) as well as nuclear factor κB (NF-κB)–dependent proinflammatory cytokines and chemokines.

SARS-CoV-2 is adept at evading innate recognition, signaling, IFN induction, and IFN-stimulated genes (ISGs) through the expression of a number of viral proteins that block these pathways (9). Consequently, lower levels of IFN-I or IFN-III are detected in the lungs or peripheral

blood of SARS-CoV-2–infected individuals compared with other respiratory viruses (10–12). IFN-III (IFN-λ1 and IFN-λ3) and ISGs are expressed in the upper airways of patients with reduced disease risk or severity, whereas expression of IFN-λ2 and type I IFNs—but not ISGs—can be elevated in the upper airways of those who develop severe COVID-19 (13). Patients with genetic mutations or autoantibodies that interfere with IFN pathways suffer from life-threatening COVID-19 disease as discussed below. During the late phase of disease, prolonged IFN secretion correlates with worse disease outcomes (14, 15), possibly through the induction of chemokines that recruit inflammatory cellular infiltrates (16). In addition, COVID-19 is accompanied by a significant decline in the number of immune sentinel cells including pDCs and conventional DCs (cDCs) in the blood (14) and lungs (17).

Innate immune responses induced by PRR signaling activate effector cells to mediate viral clearance. NK cells, for example, contribute to detecting and eliminating virally infected cells. However, in patients with severe COVID-19, NK cells are depleted from the blood and become dysfunctional as a result of transforming growth factor–β (TGF-β), which impairs their antiviral function (18). Alveolar macrophages, which play an important sentinel role in the lungs by sensing and triggering potent antiviral immunity, are severely depleted in the lungs of patients with severe disease (19, 20). In contrast to the impaired early antiviral defense mediated by IFN-I and IFN-III, proinflammatory cytokines and chemokines are significantly elevated (16). In addition, T cells are profoundly depleted from the blood circulation of patients with severe disease compared with patients with moderate disease, which suggests that induction of adaptive effector T cells able to clear virally infected cells is likely impaired in severe patients (14, 21–24). Pathological immune responses that contribute to severe COVID-19 are discussed below.

Adaptive immunity to SARS-CoV-2

Given the key role of adaptive immunity in protection from viral infection and disease, the study of humoral antibody (Ab) and cell-mediated immunity has been aggressively pursued by the scientific community with the aim of defining correlates of protection, developing immune-based therapies, and optimizing vaccine design and administration (25–27).

Abs can neutralize SARS-CoV-2 by blocking binding of the viral spike (S) protein to the ACE2 receptor and can promote effector function by binding to the complement and Fc receptors (28). Structure-guided serological studies demonstrated that the bulk of neutralizing Abs bind to distinct epitopes in the receptor binding domain (RBD) of the S protein and primarily to sites 1a and 1b, which represent the receptor-binding motif (RBM). Additionally, a minor fraction of neutralizing Abs bind to the

[1]Precision Immunology Institute, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [2]Department of Medicine and Program in Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Institute of Microbiology, ETH Zürich, 8093 Zürich, Switzerland. [4]Institute for Research in Biomedicine, Università della Svizzera Italiana, 6500 Bellinzona, Switzerland. [5]Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA. [6]Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT, USA
*Corresponding author. Email: miriam.merad@mssm.edu (M.M.); cblish@stanford.edu (C.A.B.); federica.sallusto@irb.usi.ch (F.S.); akiko.iwasaki@yale.edu (A.I.)

N-terminal domain (NTD) (29–31). Furthermore, some RBM-specific Abs can mimic ACE2 and directly trigger and consequently disable the fusion protein, a property that was initially discovered in SARS-CoV. Neutralizing epitopes outside RBD and NTD are generally subdominant but may be more conserved as in the case of the stem-helix region (32). The quantity and quality of the Ab response is related to the amount of antigen and the duration of the germinal center reaction where the response matures (33). Ab responses of increasing magnitudes are detected in infected, preimmune, and vaccinated individuals, most likely reflecting differences in antigenic load and exposure (29, 34). In immunosuppressed patients and those undergoing dialysis, Ab response is generally poor and has been shown to contribute to the chronic persistence of the virus favoring the selection of variants (35).

CD4+ and CD8+ T cells elicited by SARS-CoV-2 infection are directed against a range of antigens including structural and nonstructural proteins and are significantly associated with milder disease (36). Ab-mediated depletion of CD8+ T cells in convalescent macaques partially abrogates protection against rechallenge with SARS-CoV-2, suggesting a role for CD8+ T cells in the face of waning Ab responses (37). In view of its relevance for the induction of neutralizing Abs, the CD4+ T cell response to the S protein has been studied in great detail in convalescent individuals and vaccinees with prediction algorithms, peptide or protein stimulation, and isolation of T cell clones (38–40). These studies revealed multiple T cell epitopes in the RBD—including an immunodominant site, which is highly conserved among Sarbecoviruses and VOCs. As discussed below, one of the limitations imposed by severe COVID-19 is the profound T cell lymphopenia that affects the host's ability to mount a robust immune response.

Abs and memory T cells elicited by common cold coronaviruses are recalled by SARS-CoV-2 infection or vaccination, but their role in protection or pathology remains to be established (41, 42). Further, it was suggested that preexisting memory CD4+ T cells specific to an early transcribed replication transcription complex found in abortive infection may confer cross-protection against SARS-CoV-2 (43). In addition, cross-reactive Abs elicited by seasonal coronaviruses may influence the course of the disease as well as de novo immune responses to SARS-CoV-2 (44, 45).

Seminal studies on RSV have paved the way for the use of stabilized prefusion proteins as vaccines to facilitate the production of neutralizing Abs (46). Similarly, a prefusion stabilized S protein in the Moderna and Pfizer/BioNTech mRNA vaccines and in the Ad26-based Johnson & Johnson vaccine was used to optimize the production of neutralizing Abs while limiting the production of non-neutralizing Abs specific for the postfusion conformation, which are elicited by natural infection (47). Emerging SARS-CoV-2 VOCs can, however, evade the neutralizing Ab response elicited by infection or by current vaccines based on the Wuhan strain (Wuhan-Hu-1) (48–50). Mutations found in VOCs primarily cluster in the RBM, resulting in increased binding to ACE2 and escape from neutralizing Abs. Furthermore, mutations and deletions in NTD change the domain structure and may account for differences in neutralizing titers observed against different VOCs (31). The Omicron variant in particular carries 15 mutations in the RBD, which substantially erode the neutralizing activity induced by previous infection or vaccination.

The interrogation of memory B cells has provided a plethora of monoclonal Abs that have been instrumental in dissecting the mechanisms of protection and informing vaccine design. Notably, several potent neutralizing mAbs have been rapidly developed and found to be effective as therapeutics when administered to patients who may be at risk during early infection (27). One limitation of this approach can be found in the variability of the S protein, particularly of the RBM, which has reduced or abolished the efficacy of several approved mAbs—even when used in combination—against the highly mutated Omicron variant (51). The current strategy is to focus on Abs that are less susceptible to variants because they mimic ACE2 or recognize conserved epitopes such as S309/sotrovimab, which binds to RBD site IV outside of the RBM and retains virus-neutralizing activity and effector functions against VOCs, including Omicron (51).

Immunopathology of severe COVID-19

SARS-CoV-2 shares 73% homology with SARS-CoV, and similar to SARS-CoV the pathogenesis of SARS-CoV-2–induced pneumonia likely occurs in two phases (52): First, the viral phase is characterized by viral replication resulting in direct virus-mediated tissue damage. The extent of this damage determines the pathogenesis of the secondary phase, which is characterized by the recruitment of effector immune cells causing a local and systemic inflammatory response that can persist even after viral clearance. The development of pulmonary disease is associated with excess vascular permeability leading to microthrombi deposition and vascular permeability (53), in

Table 1. Common manifestations and potential complications of COVID-19. N/A, not applicable.

| Organ system | Acute and severe Covid | | Long Covid |
	Common symptoms	Potential complications	Common symptoms
Respiratory	Cough Dyspnea, Rhinorrhea (runny nose)	Pneumonia Acute respiratory distress syndrome (ARDS) Respiratory failure	Dyspnea, cough, interstitial lung disease
Systemic	Fever Fatigue	Sepsis, kidney injury Liver injury	Fever, fatigue post-exertional malaise
Nervous System	Headache Confusion Anosmia (loss of smell) Ageusia (loss of taste)	Stroke Seizure Guillain-Barre syndrome	Memory deficit, loss of concentration, sleep difficulty, mood disorder, paresthesia, myelopathy, neuropathy, anosmia, ageusia, blurry vision, tinnitus, headache
Cardiovascular	Chest pain Shortness of breath	Myocardial infarction Myocarditis Cardiogenic shock Stress-induced cardiomyopathy Clotting abnormalities	Chest pain, palpitations, hypertension, angina, myocarditis, pericarditis, cardiac arrhythmias, postural orthostatic syndrome
Gastrointestinal	Nausea Vomiting Diarrhea Anorexia (loss of appetite)	Acute pancreatitis Acute appendicitis Intestinal obstruction Bowel ischemia Hemoperitoneum Abdominal compartment syndrome	Diarrhea, vomiting, nausea, loss of appetite, hepatitis
Skin	Rash	N/A	Hair loss
Musculoskeletal	Myalgias (muscle aches)	N/A	Myalgia, arthralgia, weakness
Endocrine	Hyperglycemia	Diabetes	Hot/cold sensation, thyroid abnormalities, diabetes, hormonal abnormalities
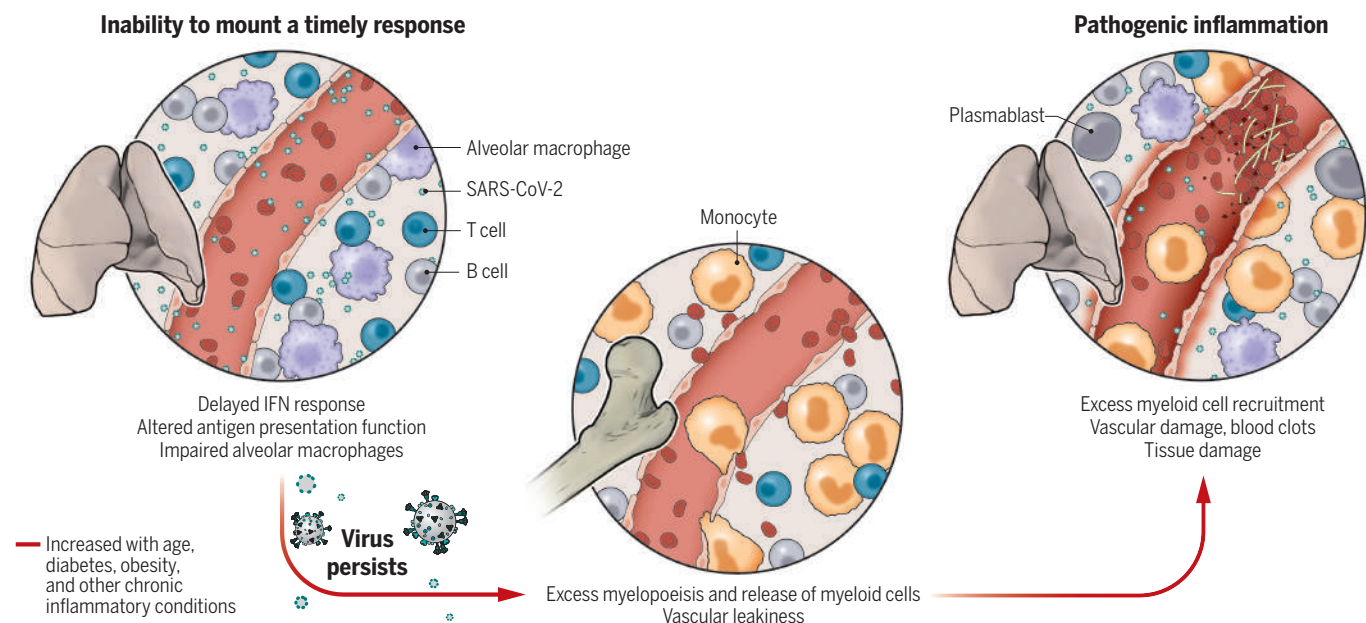
Fig. 1. COVID-19 pathogenesis. Inability to mount a timely and effective antiviral response because of delayed IFN response, altered antigen presentation function, or altered tissue resident macrophage pool, common in older individuals, promotes viral persistence and prolonged tissue damage that trigger the prolonged blood release and recruitment of damaging inflammatory myeloid cells to the infected site. Enhanced myelopoiesis and vascular damage common in older individuals and in patients with chronic inflammatory diseases further contribute to the enhanced release of inflammatory myeloid cells from the bone marrow to the blood circulation and their recruitment to the site of infection leading to profound tissue damage, vascular lesions, and blood clots common in patients with severe disease.

addition to a range of systemic symptoms including (i) olfactory dysfunction, which occurs in approximately half of people diagnosed with COVID-19 but has not been described for SARS-CoV (54); (ii) gastrointestinal (GI) symptoms (55); and (iii) cardiac, hepatobiliary, and renal dysfunction (53). The pathogenesis of extrapulmonary manifestations is likely multifactorial as a result of direct viral injury to the tissue cells, vessels, or neurons, as well as cytokine release, auto-Ab–induced tissue damage, vascular damage, or gut dysbiosis in the case of GI symptoms. Although several risk factors for severe COVID-19 have been identified, the pathophysiological underpinnings that contribute to disease severity are not fully understood. Below, we discuss two broad hypotheses that have emerged to explain severe COVID-19 pathophysiology and examine how the identified risk factors contribute to these pathophysiological drivers (Fig. 1).

Inability to mount a timely antiviral response

The inability to mount a timely and effective immune response against SARS-CoV-2 is multifactorial and is discussed as follows:

Viral load: The exact contribution of viral loads in disease course has been difficult to establish likely because of discrepancies in measurements between studies and sampling biases as well as viral dynamics, and it is still unclear how much this contributes to disease outcome.

Defects in type I IFN response: The timely production of type I IFN by host cells is critical for limiting viral replication and promoting antiviral immunity. Loss of function variants in loci that control TLR3- and IRF7-dependent type I IFN immunity have been identified in a small number of severe patients (56). Notably, these patients were adults who had never been hospitalized for severe viral illness before COVID-19. AutoAbs against IFN-α and IFN-ω (57–59) have been identified in patients with severe disease and have been shown to contribute to delayed viral clearance (59). Neutralizing IFN-α and IFN-ω autoAbs have recently been found to increase with age (60) suggesting that type I IFN autoAbs can precede the disease and serve as a biomarker of disease severity.

Imbalanced adaptive immunity: The reduction of DC numbers and profound T cell lymphopenia driven by T cell sequestration in tissues or T cell apoptosis as a result of proinflammatory cytokines is common in severe patients (22, 23) and may contribute to defective T cell–mediated viral clearance. Defects in type 1 immune responses (23) and excess type 2 immunity have both been shown to correlate with severe COVID-19 (14), suggesting that a maladapted adaptive immune response to the virus may also lead to delayed viral clearance and disease progression. Substantial plasmablast expansion reaching up to 30% of circulating B cells—sometimes associated with extrafollicular responses—has also been reported in severe patients (61). The massive plasmablast expansion may reflect polyreactivity given the low levels of somatic mutations in Ab clones observed in patients and may have

less potent viral control while contributing to tissue damage as suggested below.

Inability to control SARS-CoV-2–driven inflammatory responses

Autopsies of deceased COVID-19 patients have revealed very little active viral infection and substantial accumulation of activated immune cells, suggesting that organ failure is unlikely to result from extensive viral-induced tissue damage but is instead caused by an overactivated immune system or vascular damage. Below, we discuss the potential causes of these excessive inflammatory responses.

Altered myeloid responses: Excess circulating immature monocytes, neutrophils, and myeloid progenitors—referred to as emergency myelopoiesis—are almost pathognomonic of severe disease and are triggered during the first phase of infection, likely because of delayed viral clearance especially in conditions with preexisting altered myelopoiesis (62, 63). Circulating myeloid cells produce excessive amounts of inflammatory molecules that promote vascular permeability and organ damage (Fig. 1). By contrast, lung tissue–resident macrophages known to play a key role in tissue homeostasis and repair are often depleted in severe patients (19, 20). The exact nature of their depletion is unclear, however, and may be a result of direct viral injury or cell death induced by excess inflammation.

Pathogenic antibodies: AutoAbs have been identified in severe COVID-19 including a high proportion of Abs targeting nuclear antigen

(58, 59), phospholipids (64), T cell antigens, B cell antigens, chemokines, and cytokines (57–59). Anti-phospholipid Abs cause clots in mouse models (64) whereas immunoglobulin G (IgG) from patients with anti-CD38 or anti-CD3ε autoAbs exhibit increased Ab-dependent cellular phagocytosis by macrophages, which may contribute to the profound lymphopenia observed in some patients (59). AutoAbs have also been detected in the cerebrospinal fluid of patients with neurological symptoms (65).

In addition, afucosylated anti–SARS-CoV-2 IgG1 with enhanced affinity for the activating FcγRIIIa accumulates mainly in male patients with severe COVID-19 (66). These Abs fuel the production of inflammatory cytokines and NK cell degranulation, which may contribute to tissue damage; furthermore, they are absent in asymptomatic patients and seropositive children (66). The mechanisms that lead to the production of afucosylated Abs and whether these Abs can be induced upon vaccination still remains to be understood.

Vascular damage: Hypercoagulation, endothelial damage, and arterial and venous embolism are very common in severe COVID-19 (55). Although the exact mechanisms of hypercoagulability remain unclear, it may be a result of direct viral damage to the vasculature or severe inflammatory responses, which can alter the vascular endothelium and induce the activation of platelets, monocytes, and macrophages which in turn promote the release of tissue factor, von Willebrand factor, and factor VIII, leading to the production of thrombin and fibrin clot formation (53).

Predisposing risk factors

Age is by far the strongest risk factor for critical pneumonia with the risk of life-threatening disease sharply increasing from 65 years of age onward. Surprisingly, infants and young toddlers commonly at risk for influenza infections are usually protected from severe disease (67). Male patients are at much higher risk of developing severe COVID-19, which may be a result of enhanced innate immunity and impaired T cell activation compared with females (68) and/or higher levels of afucosylated anti–SARS-CoV-2 Abs (66). Diabetes, hypertension, obesity, and chronic kidney diseases are also important risk factors for severe COVID-19. These conditions can include chronic inflammatory lesions, which may enhance and perpetuate the inflammatory cycle. Metabolic syndrome—often associated with insulin resistance in diabetic and obese patients—also leads to vascular damage and impaired tissue repair. It is likely that the rapidity and quality of immune responses during the first phase of infection is particularly critical in these patients as delayed viral clearance is more likely to lead to overt recruitment and activation of immune cells, vascular damage, and impaired tissue repair.

Immune-directed therapy for COVID-19

Consistent with the major role of the immune response in driving pathogenesis, a wide range of immunomodulatory agents have been tested for the treatment of COVID-19, and several have demonstrated efficacy. Evidence-based treatment guidelines are regularly updated by the National Institutes of Health and contain references to all the primary literature guiding these recommendations (69). Such guidelines will continue to evolve as new drugs are developed and new clinical trials enable new indications. The recent approval of the direct-acting antivirals ritonavir-boosted nirmatrelvir (Paxlovid), remdesivir, and molnupiravir for nonhospitalized patients with mild-to-moderate COVID-19 who are at high risk of disease progression is an excellent example of such a recent modification. Patients who are not hospitalized but are at high risk of disease progression can benefit from anti–SARS-CoV-2 mAb products (either bamlanivimab plus etesevimab, or sasirivimab plus imdevimab or sotrovimab) (69). For hospitalized patients who require supplemental oxygen, remdesivir (a direct-acting antiviral) and dexamethasone (a broad-spectrum corticosteroid) are recommended (69). The success of dexamethasone in improving outcomes of patients with advanced disease highlights the critical role of inflammatory responses in mediating pathogenesis and marks a considerable departure from the treatment of influenza, in which such steroid treatments are harmful. Indeed, several more targeted agents have also proven beneficial, particularly agents counteracting the IL-6 pathway including tociluzimab and sarilumab (69). Inhibitors of the JAK–STAT signaling pathway such as baricitinib and tofactinib have also shown promise (69). Drugs that target the IL-1 pathway such as anakinra and canakinumab have also been tested, though there is a lack of sufficiently strong evidence of benefits to recommend their use (69). Critically, IL-6- and JAK–STAT-blocking agents must be used within a brief window in recently hospitalized patients with rapidly declining courses. Presumably, this is when interruption of the "cytokine storm" can change the outcome of the disease. When such anti-inflammatory drugs are used in less ill populations or too late, there is a trend toward harm, making it imperative to identify those patients with inflammation that warrants intervention. This opens the door for future diagnostics to better match a specific patient with a specific drug.

Immunology of Long Covid

It is now clear that COVID-19 can lead to long-term disease—often referred to as Long Covid syndrome or post-acute sequelae of SARS-CoV-2 infection (PASC)—in a significant proportion of survivors. Although there is no universal consensus in the definition of PASC, the American Centers for Disease Control and Prevention defines it as a wide range of new, returning, or ongoing health problems that people experience 4 or more weeks after first being infected with SARS-CoV-2. By contrast, the World Health Organization defines it as a condition that occurs in people with confirmed or probable SARS-CoV-2 infection, usually 3 months from the onset of COVID-19 with symptoms and that last for at least 2 months and cannot be explained by an alternative diagnosis (Fig. 2). A systematic review of 57 peer-reviewed studies with 250,351 survivors of COVID-19 who met their inclusion criteria for PASC showed that the median age of patients was 54.4 years, 56% were male, and 79% were hospitalized during acute COVID-19 (70). At 6 months, 54% of survivors suffered at least one PASC symptom. However, nonhospitalized COVID-19 survivors who developed PASC were primarily middle-aged women. In a survey of 445 nonhospitalized Danish COVID-19 patients, persistent symptoms—most commonly fatigue and difficulty with memory and concentration—were reported by 36% of symptomatic participants with a follow up of >4 weeks. Risk factors for persistent symptoms included female sex (44% and 24% for women and men respectively, odds ratio 2.7) and Body Mass Index (odds ratio 1.1) (71).

The immunobiology of PASC is currently under investigation. Leading hypotheses include (i) persistent virus or viral antigens and RNA in tissues that drive chronic inflammation, (ii) the triggering of autoimmunity after acute viral infection (72), (iii) dysbiosis of microbiome or virome, and (iv) unrepaired tissue damage (Fig. 2). SARS-CoV-2 viral proteins and/or RNA have been detected throughout respiratory, cardiac, renal, and reproductive systems, as well as in the brain, muscles, eyes (73), GI tract (73–75), and lymph nodes (73, 75) months after infection. Aberrant innate immune stimulation is associated with PASC during the early and late phases of disease. Studies have shown that some inflammatory cytokines—including IL-6, TNF-α, and IL-1β—are elevated in PASC patients (76, 77). IFN-β and IFN-λ1 remain elevated 8 months after infection in PASC patients compared with recovered controls. Furthermore, combinations of IFN-β, PTX3, IFN-γ, IFN-λ2/3, and IL-6 were associated with PASC with 78.5 to 81.6% accuracy (78). A deep, multiomic, longitudinal investigation of 309 COVID-19 patients (71% hospitalized) from initial diagnosis to 2 to 3 months after identified four PASC-anticipating risk factors at the time of initial COVID-19 diagnosis: type 2 diabetes, SARS-CoV-2 RNAemia, Epstein–Barr virus viremia, and autoAbs (79). AutoAbs were anticorrelated with anti–SARS-CoV-2 Abs. These data collectively suggest that the persistence of viral components may result in chronically elevated IFNs and cytokines. Moreover, certain autoAbs may provide permissive conditions for such viral persistence. However, there is no causal link between viral proteins and RNA and these elevated cytokines.

Adaptive immune responses to SARS-CoV-2 have been examined in PASC versus nonPASC

convalescent individuals. Studies of individuals with confirmed COVID-19 with or without PASC found no difference in anti-S Ab levels during the 8 months of study and no differences in the initial PCR Ct value for viral RNA (80, 81). Other studies found that anti-N, -M, and -S CD8 T effector memory (T_{EM}) cells and TEM cells that reexpress CD45RA (T_{EMRA}) in PASC patients were reduced compared with those of convalescent individuals, whereas anti-N T follicular helper cells (T_{FH}) and anti-N IgG levels were elevated in blood circulation (82). These studies suggest a possible persistence of viral antigens driving immune stimulation.

As discussed above, autoAbs have been detected in acute COVID-19 and evidence is emerging that autoAbs at the time of diagnosis were correlated with PASC (79). A study of 31 patients suffering from different PASC symptoms—including neurological and cardiovascular symptoms—revealed that in all 31 patients, 2 to 7 different GPCR-functional autoAbs (fAABs) were present, acting as receptor agonists. Some of those GPCR-fAABs activated their target receptors, causing a positive chronotropic effect in neonatal rat cardiomyocytes, whereas others caused a negative chronotropic effect (83). AutoAbs to GPCR have been associated with numerous diseases of the cardiovascular, pulmonary, and the central nervous system in addition to autoimmune conditions (84) and may play a pathological role in mediating PASC symptoms.

Emergent results suggest that COVID-19 vaccines may have an effect on PASC symptoms in a subset of individuals. A prospective case control study using self-reported data found that two doses of vaccines before SARS-CoV-2 infection (n = 906) reduced the risk of prolonged symptoms after 28 days compared with unvaccinated matched controls (n = 906) (85). Analysis of existing data on a 1296-patient cohort with PASC emulated a 1:1 matched (vaccinated: unvaccinated) trial to determine the effect of vaccination in PASC patients, and found that vaccination ~38 days postinfection significantly reduced the risk of prolonged symptoms 120 days postinfection (86). Even after infection, vaccination within the first 12 weeks was associated with a reduced risk of developing Long Covid on the basis of a retrospective analysis of the medical history of 240,648 SARS-CoV-2–infected persons (87). How exactly the vaccine prevents or treats PASC is currently unclear. It is possible that the anti-S Abs and T cells elicited by the vaccines promote clearance of residual antigens or viral particles, eliminating the cause of chronic inflammation. It is also possible that vaccine-induced cytokines act on autoreactive lymphocytes

and shut down the production of pathogenic cytokines or reprogram pathogenic lymphocytes. Evaluation of immune responses in PASC patients to vaccines is needed to reveal the underlying mechanism of protection.

Concluding remarks and future directions

The COVID-19 pandemic has wrought massive disruption and resulted in the loss of countless lives; however, there have been silver linings. The particularly rapid development of highly efficacious vaccines is foremost among these and has established a playbook for the response to future pandemics. One comforting prospect is the degree to which advances in our understanding and treatment of COVID-19 have been aided by an unprecedented degree of scientific cooperation. Free sharing of data has allowed us to

rapidly glean critical insights into the role of the immune system in contributing to both protection and pathogenesis in COVID-19. Such insights will undoubtedly help us prepare for the next pandemic, just as decades of previous immunological research led to our current COVID-19 vaccines. However, many challenges remain, and our progress in ending this pandemic is threatened by inequitable distribution of vaccines and the rise of variants that are less susceptible to vaccination and prior-infection–mediated immunity. As infections continue to occur, there remains a need for new therapeutics and hence a need for a better understanding of the pathophysiology of COVID-19. In addition to treating acute infections, there is a dire need to better understand and develop treatments for individuals with PASC. Another threat is the amount of misinformation and erroneous theories about the pandemic, vaccines, and therapeutic efforts that have been circulating in social media, some unfortunately introduced by scientists. More than ever, interdisciplinary and integrative approaches to scientific collaboration and fighting misinformation are necessary to tackle these and other challenges that lie ahead.
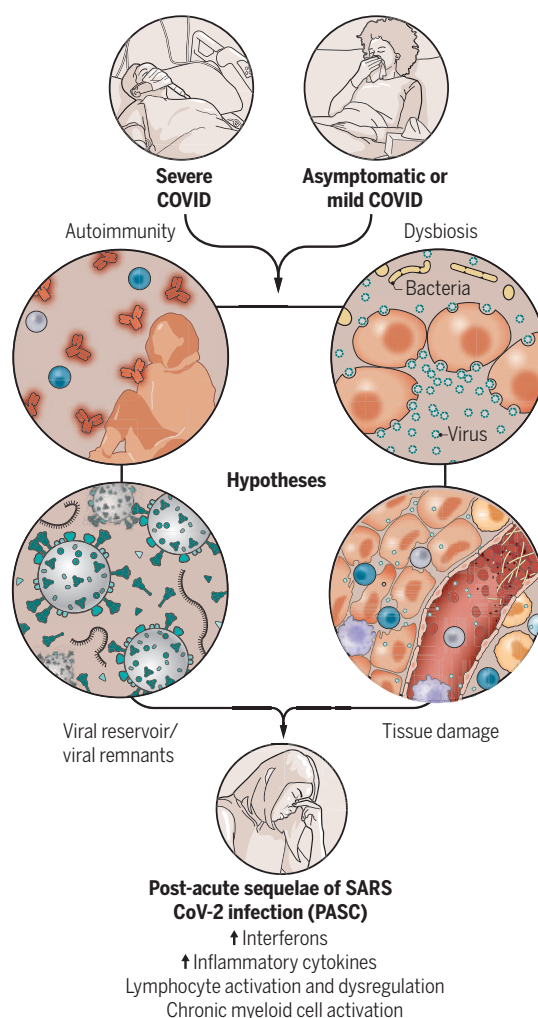

Fig. 2. Immunology of PASC. A fraction of COVID-19 patients with either severe or mild COVID-19 develop a variety of new, recurring, or ongoing symptoms and clinical findings 4 or more weeks after infection. Analyses of immune responses in people with PASC reveal key inflammatory cytokines and cellular activation phenotypes that are significantly elevated over nonPASC convalescent controls. Further studies are needed to identify the drivers of PASC pathophysiology.

REFERENCES AND NOTES

1. W. Tan et al., A Novel Coronavirus Genome Identified in a Cluster of Pneumonia Cases - Wuhan, China 2019-2020. China CDC Weekly **2**, 61–62 (2020).
2. W. J. Wiersinga, A. Rhodes, A. C. Cheng, S. J. Peacock, H. C. Prescott, JAMA **324**, 782–793 (2020).
3. L. Mao et al., JAMA Neurol. **77**, 683–690 (2020).
4. Y. T. Chen et al., Crit. Care **24**, 346 (2020).
5. R. Mao et al., Lancet Gastroenterol. Hepatol. **5**, 667–678 (2020).
6. B. Long, W. J. Brady, A. Koyfman, M. Gottlieb, Am. J. Emerg. Med. **38**, 1504–1507 (2020).
7. K. Qi et al., Medicine **100**, e25230 (2021).
8. E. A. Madden, M. S. Diamond, Curr. Opin. Virol. **52**, 30–38 (2021).
9. Y. Kasuga, B. Zhu, K.-J. Jang, J.-S. Yoo, Exp. Mol. Med. **53**, 723–736 (2021).
10. D. Blanco-Melo et al., Cell **181**, 1036–1045.e9 (2020).
11. I.-E. Galani et al., Nat. Immunol. **22**, 32–40 (2021).
12. J. Hadjadj et al., Science **369**, 718–724 (2020).
13. B. Sposito et al., Cell **184**, 4953–4968.e16 (2021).
14. C. Lucas et al., Nature **584**, 463–469 (2020).
15. J. S. Lee et al., Sci. Immunol. **5**, eabd1554 (2020).
16. B. Israelow et al., J. Exp. Med. **217**, e20201241 (2020).
17. I. Sánchez-Cerrillo et al., medRxiv 2020.2005.2013.20100925 [Preprint] (2020); doi: 10.1101/2020.05.13.20100925.
18. M. Witkowski et al., Nature **600**, 295–301 (2021).
19. M. Liao et al., Nat. Med. **26**, 842–844 (2020).
20. T. M. Delorey et al., Nature **595**, 107–113 (2021).
21. A. J. Wilk et al., Nat. Med. **26**, 1070–1076 (2020).
22. R. Zhou et al., Immunity **53**, 864–877.e5 (2020).
23. D. Mathew et al., Science **369**, eabc8511 (2020).
24. A. J. Wilk et al., J. Exp. Med. **218**, e20210582 (2021).
25. K. A. Earle et al., Vaccine **39**, 4423–4428 (2021).
26. D. S. Khoury et al., Nat. Med. **27**, 1205–1211 (2021).
27. D. Corti, L. A. Purcell, G. Snell, D. Veesler, Cell **184**, 3086–3108 (2021).
28. T. Zohar, G. Alter, Nat. Rev. Immunol. **20**, 392–394 (2020).
29. L. Piccoli et al., Cell **183**, 1024–1042.e21 (2020).
30. T. N. Starr et al., Cell **182**, 1295–1310.e20 (2020).
31. M. McCallum et al., Science **373**, 648–654 (2021).
32. D. Pinto et al., Science **373**, 1109–1116 (2021).
33. Z. Wang et al., Nature **595**, 426–431 (2021).
34. C. Lucas et al., Nature **600**, 523–529 (2021).
35. S. A. Kemp et al., Nature **592**, 277–282 (2021).

36. C. Rydyznski Moderbacher et al., Cell 183, 996–1012.e19 (2020).
37. K. McMahan et al., Nature 590, 630–634 (2021).
38. A. Grifoni et al., Cell 181, 1489–1501.e15 (2020).
39. N. Le Bert et al., Nature 584, 457–462 (2020).
40. J. S. Low et al., Science 372, 1336–1341 (2021).
41. A. Sette, S. Crotty, Nat. Rev. Immunol. 20, 457–458 (2020).
42. J. J. Guthmiller, P. C. Wilson, Science 370, 1272–1273 (2020).
43. L. Swadling et al., Nature 601, 110–117 (2022).
44. K. W. Ng et al., Science 370, 1339–1343 (2020).
45. H. L. Dugan et al., Immunity 54, 1290–1303.e7 (2021).
46. B. S. Graham, K. Modjarrad, J. S. McLellan, Curr. Opin. Immunol. 35, 30–38 (2015).
47. D. M. Altmann, R. J. Boyton, Science 375, xxx–xxx (2022).
48. D. A. Collier et al., Nature 593, 136–141 (2021).
49. D. Zhou et al., Cell 184, 2348–2361.e6 (2021).
50. P. Wang et al., Nature 593, 130–135 (2021).
51. E. Cameroni et al., Nature (2021)
52. R. Channappanavar et al., Cell Host Microbe 19, 181–193 (2016).
53. L. Perico et al., Nat. Rev. Nephrol. 17, 46–64 (2021).
54. M. S. Xydakis et al., Lancet Neurol. 20, 753–761 (2021).
55. A. E. Livanos et al., Gastroenterology 160, 2435–2450.e34 (2021).
56. Q. Zhang et al., Science 370, eabd4570 (2020).
57. P. Bastard et al., Science 370, eabd4585 (2020).
58. S. E. Chang et al., Nat. Commun. 12, 5417 (2021).
59. E. Y. Wang et al., Nature 595, 283–288 (2021).
60. P. Bastard et al., Sci. Immunol. 6, eabl4340 (2021).
61. N. Kaneko et al., Cell 183, 143–157.e13 (2020).
62. J. Schulte-Schrepping et al., Cell 182, 1419–1440.e23 (2020).
63. A. Silvin et al., Cell 182, 1401–1418.e18 (2020).
64. Y. Zuo et al., Sci. Transl. Med. 12, eabd3876 (2020).
65. C. Franke et al., Brain Behav. Immun. 93, 415–419 (2021).
66. S. Chakraborty et al., Nat. Immunol. 22, 67–73 (2021).
67. E. J. Williamson et al., Nature 584, 430–436 (2020).
68. T. Takahashi et al., Nature 588, 315–320 (2020).
69. NIH, Coronavirus Disease 2019 (COVID-19) Treatment Guidelines (2022) www.covid19treatmentguidelines.nih.gov).
70. D. Groff et al., JAMA Netw. Open 4, e2128568–e2128568 (2021).
71. S. Bliddal et al., Sci. Rep. 11, 13153 (2021).
72. D. Massey et al., medRxiv 2021.2007.2021.21260391 [Preprint] (2021); doi: 10.1101/2021.07.21.21260391.
73. C. Daniel et al., version 1, Research Square [Preprint] (2021); doi: 10.1101/10.21203/rs.3.rs-1139035/v1.
74. C. Gaebler et al., Nature 591, 639–644 (2021).
75. C. C. L. Cheung et al., Gut 71, 226–229 (2021).
76. M. J. Peluso et al., medRxiv 2021.2007.2009.21260287 [Preprint] (2021); doi: 10.1101/2021.07.09.21260287.
77. C. Schultheiß et al., medRxiv 2021.11.16.21266391 [Preprint] 2021); doi: 10.1101/2021.11.16.21266391.
78. C. Phetsouphanh et al., Nat. Immunol. 23, 210–216 (2022).
79. Y. Su et al., Cell (2022).
80. C. Pereira et al., J. Infect. Dis. 223, 1671–1676 (2021).
81. H. Fang et al., Viruses 13, 916 (2021).
82. L. Visvabharathy et al., medRxiv 2021.2008.2008.21261763 [Preprint] (2021).
83. G. Wallukat et al., J. Transl. Autoimmun. 4, 100100 (2021).
84. M. A. Skiba, A. C. Kruse, Trends Pharmacol. Sci. 42, 135–150 (2021).
85. M. Antonelli et al., Lancet Infect. Dis. (2021).
86. V.-T. Tran, E. Perrodeau, J. Saldanha, I. Pane, P. Ravaud, SSRN [Preprint] (2021); http://doi.org/10.2139/ssrn.3932953.
87. M. A. Simon, R. D. Luginbuhl, R. Parker, medRxiv 2021.2011.2017.21263608 [Preprint] (2021); doi: 10.1101/2021.11.17.21263608.

REVIEW

COVID-19 vaccination: The road ahead

Daniel M. Altmann[1]* and Rosemary J. Boyton[2,3]

A diverse array of successful, first-generation SARS-CoV-2 vaccines have played a huge role in efforts to bring the COVID-19 pandemic under control, even though inequitable distribution still leaves many vulnerable. Additional challenges loom for the next phase. These include optimizing the immunological rationale for boosting—how often and with what—and the best approaches for building a future-proofed, durable immune repertoire to protect against oncoming viral variants, including in children. The landscape of vaccine producers and technologies is likely to become even more heterogeneous. There is a need now for appraisal of future approaches: While some favor frequent boosting with the first-generation, ancestral spike vaccines, others propose frequent readjustment using current variant sequences, polyvalent vaccines, or pan-coronavirus strategies.

The rapid production and clinical development of efficacious vaccines to mitigate the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has been a testament to decades of research and sequential advances in immunology, vaccinology, and adjuvant biology. Clinically tested in the midst of the battle against COVID-19, the returns have been greater than anticipated, providing new momentum for vaccinology with respect to many other infectious diseases.

A number of outstanding and comprehensive recent review articles have compared individual vaccine details, clinical trial data, immune parameters, and safety profiles, such that there would be little added value here to reiterating those points (1–4). Less than a year elapsed between identification of a novel pathogen sequence and wide-scale vaccine rollout. Hundreds of vaccines were developed, more than 100 of these were taken into clinical trials, and some 24 are currently authorized and in use (5). The hit rate has been high, with only 10 vaccine candidates abandoned because of poor efficacy in clinical trials. However, this scientifically successful vaccinology has been counterbalanced by inequitable global distribution (6, 7). More than 9 billion doses have been administered—more than enough for one dose to each living person of vaccinable age—yet vaccine distribution is heavily skewed. Half of the planet is currently unvaccinated, and only 4% of populations in low-income countries have received a dose (6). As the COVID-19 pandemic continues to wreak havoc, events (and mortality) would have been considerably worse without the mitigation effects from successful vaccine programs. It has been challenging to estimate the number of lives saved by COVID-19 vaccines, not least when global fatalities are themselves a source of uncertainty, but estimates are currently in the region of more than 700,000 lives saved.

Many of the high-level vaccine decisions of the past 2 years were uncontentious: the centrality of effective vaccines in the antiviral strategy, the key role in protection of targeting spike neutralization (based on prior knowledge of SARS-CoV and MERS), and the application of potent, exciting, scalable vaccine technologies-in-waiting for vaccine delivery—RNA, DNA, and recombinant protein with novel adjuvants and adenovirus platforms (8). However, for technologies never before deployed at this scale, vaccine safety and durability have imposed a steep learning curve.

Vaccine safety

Coming into the COVID-19 global vaccination program, vaccine safety and vaccine hesitancy posed considerable challenges: Variable levels of vaccine hesitancy across countries had the potential to undermine protective coverage, with hesitancy generally linked to safety concerns and/or suspicion of governmental authority. With billions of vaccine doses given across a wide age range in more than 180 countries, the safety record has been good. Vaccination uptake in most countries exceeded initial projections. Noteworthy serious or lethal adverse event reports included anaphylaxis, myocarditis, and occasionally lethal thrombotic events with thrombocytopenia, with some variability across the different vaccine platforms. Vaccine-induced immune thrombocytopenia and thrombosis (VITT) was specifically associated with the adenovirus-vectored vaccines (9). "Yellow Card" adverse event reporting in the UK resulted in 260 VITT cases among 31 million vaccine doses, which were skewed toward younger recipients (10). Mortality among the initial cases was 22%. The mechanism was defined as induction of anti-platelet factor–4 (PF4) autoantibodies, targeted to an epitope in the heparin binding site and induced in response to neoantigens generated by interaction of PF4 with adenovirus components (11). Further characterization of this pathway should now enable vaccine modifications to

[1]Department of Immunology and Inflammation, Faculty of Medicine, Hammersmith Hospital Campus, Imperial College London, London, UK. [2]Department of Infectious Disease, Faculty of Medicine, Hammersmith Hospital Campus, Imperial College London, London, UK. [3]Lung Division, Royal Brompton & Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust, London, UK.

*Corresponding author. Email: d.altmann@imperial.ac.uk

abrogate this effect. A small number of cases of myocarditis and pericarditis were reported in certain recipients of mRNA vaccines. Again, the challenge has been to use standardized national and international adverse event reporting systems to move to a population-level appraisal of risk. A case series of reports from England estimated additional myocarditis cases at 1 to 2 per million for the first dose of AZ/ChadOx1 and Pfizer/BNT162b2, and 6 per million for Moderna/mRNA-1273 (12). These risks should be considered against the risk of 40 myocarditis events per million during the month after SARS-CoV-2 infection.

Waning and boosting

Entering the third year of the pandemic in 2022, the terms of engagement for SARS-CoV-2 vaccinology have progressed beyond the initial emergency development of first-generation vaccines as mitigation for a respiratory virus that often causes lethal pneumonia. Our aim in reflecting on the state of play is to consider some of the much-debated questions and options moving forward.

One key concern is the seemingly rapid waning of effective immune protection against infection (but not against severe or lethal disease) (7, 13, 14). Do we face a future of frequent boosters, and to what extent does this carry us into uncharted territory of incremental focusing of immune repertoire and memory? Will the strategy be to continue with ancestral spike sequence, frequently recalibrate to the most troubling mutations [i.e., variants of concern (VOCs)] at any given time, and develop polyvalent vaccines? Or will it be to prioritize the development of vaccines with broader cross-protection, and would these depend on spike or on the viral immunome more widely (1, 15) (Fig. 1)? Many of these questions loop back to the most central uncertainties: What is envisaged as the aspiration for our future relationship with SARS-CoV-2, and what do we therefore need our vaccine strategies to achieve (16)?

Countries with high vaccine coverage are often considered to have broken the link from COVID-19 infection to hospitalization and mortality, offering a future where we settle for "living with the virus." But this assumption encompasses an ethical conundrum, even in countries with relatively high vaccination rates. At a time when virtually all current infections are attributable to the Omicron variant, if current conditions are extrapolated forward, the UK may be on course for a rumbling, ongoing turnover of 15 million infections and more than 50,000 excess deaths per annum, and the US for ~26 million annual infections and 400,000 excess deaths. These statistics would not have previously been tolerated in modern medical practice and would force a reappraisal of wider healthcare provisions. However, elimination of SARS-CoV-2 looks increasingly untenable.

An alternative scenario is that viral evolutionary mutations carry us into a relationship of endemicity with a more benign, common cold–like pathogen. Evidence for such a prediction is sometimes inferred from evolutionary analysis of the four key common cold human coronaviruses, OC43, 229E, NL63, and HKU1. However, the time scale for this might range from years to decades or even centuries (17, 18).

Immunological considerations of current SARS-CoV-2 vaccines

For most of the vaccines in wide use, the starting premise had been that the Achilles' heel of SARS-CoV-2 viral transmission is the interaction between the receptor-binding domain (RBD) of the spike protein and human angiotensin-converting enzyme 2 (ACE2), and that an immunogen eliciting high levels of neutralizing antibodies (nAbs) against epitopes at this interface would be effective. This led to a wide range of approaches for expression of spike, stabilized or not, for immune presentation. The mRNA (Pfizer, Moderna) and adenovirus (AstraZeneca, Gamaleya, Johnson & Johnson) vaccine platforms that have predominated clinical development have undergone experimental trials in diverse geographic and disease settings over decades. However, we arrived at clinical trials and COVID-19 vaccine implementation strategies with relatively sparse datasets about key features including immune priming and comparative, long-term, immune memory. It soon became clear that, generally, the mRNA vaccines are outstanding at inducing high nAb responses and the adenovirus vaccines somewhat less so, although they may induce higher T cell responses (19). Neutralizing antibody titer has been considered a key correlate of protection (COP) with respect to the initial trials on efficacy against symptomatic disease (20, 21), and these higher nAb responses may underpin the greater efficacy of the mRNA vaccines in these trials. In real-life settings, especially considering the more pertinent endpoint of severe disease and death, the adenovirus and mRNA vaccines have, depending on the VOC in question, been similarly and highly effective.

Recent studies on COP in the ChAdOx1 nCov19 and Moderna/mRNA-1273 arrive at similar conclusions with respect to the levels of binding Abs and nAbs associated with protection from symptomatic infection (22–24). There is increasing attention to wider consideration of the different facets of vaccine-induced protective immunity and its durability (22–24). This is emerging from studies of systems vaccinology, systems serology (including the contributions of non-neutralizing Abs), and detailed immune phenotyping, specificity, and immune repertoire analysis of B and T cell memory responses, including from human challenge studies (Fig. 2). It is a truism of textbook immunology and the linchpin of vaccinology that initial serum antibody output wanes in the absence of ongoing antigenic stimulation, but that B and T cell memory will have been primed for a speedy, protective response upon repeat encounter. Seen in this light, analysis of waning of serum nAbs would helpfully be broadened to consider the contributions of other immune compartments, including mucosal antibodies and systemic and lung-resident memory T cells (Fig. 2) (22).

With most European and North American vaccination programs starting in late 2020 or early 2021, it took until the latter part of 2021 to accrue data on vaccine durability and breakthrough infections. Baseline assumptions at the start of the pandemic were that B cell immunity to live infection was likely to show poor durability by analogy to recurrent seasonal susceptibility to the common cold HCoVs (25). In line with this, studies early in the pandemic highlighted the ability of the virus (in severe infection) to subvert the establishment of germinal centers, which constitute "mission control" for establishing immune memory with affinity maturation (26). By contrast, vaccine-induced immunity was expected to be both more uniform and more durable, because the vaccine expression platforms lack the coronavirus evolutionary adaptations for subverting immune priming. However, predictions about long-term immune memory after administration of either mRNA- or adenovirus-based vaccines were difficult in light of the thin prior

CD8 antigens
S, M, N, E
Nsp (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15)
ORF (3a, 6, 7a, 8, 10)

CD4 antigens
S, M, N, E
Nsp (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15)
ORF (3a, 6, 7a, 8, 10)

Seroreactive antigens
S, M, N
Nsp (1, 5)
ORF (3a, 3d, 7a, 7b, 8, 9b)

SARS-CoV-2 genome

5′ — ORF1a — ORF1b — Spike (S) — 3a — E — M — 6 — 7a8 — 7b — 9b — N — 14? — AAAAA... 3′
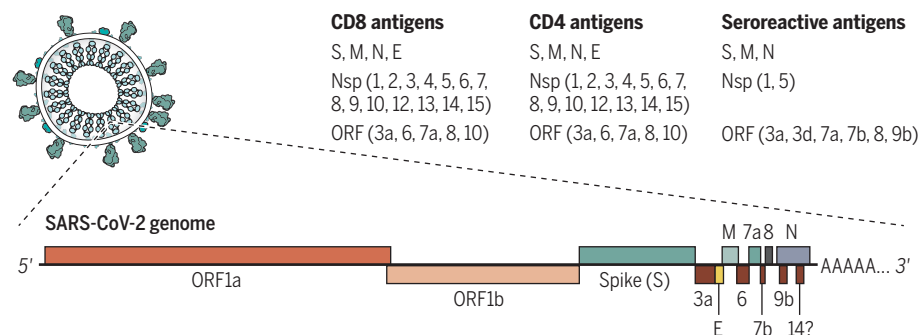
Fig. 1. The SARS-CoV-2 immunome at the level of B cells and CD4+ and CD8+ T cells. Whereas current vaccines largely focus on the spike protein as the immunogen, the immune system recognizes epitopes across the SARS-CoV-2 viral proteome, including seroreactive antigens (67) and antigens recognized by CD4 and CD8 T cells (68).

datasets. Given the relatively ephemeral expression of spike antigen at the injection site for uptake and processing by dendritic cells, would this yield a sufficient antigen depot to trigger enduring, fully fledged, immune memory? Elegant studies (with mRNA vaccines) that used fine needle aspirates to analyze vaccinated patients' draining lymph node cells show that vaccination indeed induces the expected germinal center response (*27*). Modeling of the decay kinetics of Ab responsiveness, whether after infection or after two-dose vaccination, predicted a response half-life that would leave most people protected beyond a year (*21, 28*).

A year out, one can start to appraise the real-life data from the largest-ever experiment in human immunological memory. There are three groups to compare: infected, vaccinated, and both vaccinated and infected (so-called hybrid immunity) (*29, 30*). SARS-CoV-2 immune response phenotypes to infection show the expected Gaussian distribution: Low responders are thought to be more susceptible to reinfection, and high responders are considered to have greater protective headroom before waning back into susceptibility (*31*). Ab levels after mRNA vaccination are generally higher than after viral infection, whereas levels after adenoviral vaccines are roughly in line with viral infection. Serum Ab levels wane rapidly in the months after vaccination, and there is a massive increment to the quality and quantity of response in those infected and then vaccinated (*29, 30*). Also, abundant B cell and T cell memory responses are indeed present, along with evidence of B cell somatic hypermutation and affinity maturation—the hallmarks of successfully primed adaptive immune memory (*32*). In infected or vaccinated groups, antibody waning has been more rapid than had been envisaged, as is apparent in Ab binding assays or functional neutralization assays from around 4 to 5 months (*33–35*). The litmus test of whether concern over Ab waning is overstated, given that memory T and B cells are primed and ready, comes from data on breakthrough cases, hospitalizations, and deaths. The answer is somewhat gray rather than black or white: From 5 months or so after the second vaccine dose, breakthrough infections become considerably more common, in line with waning nAb titers (*36*). On the other hand, there is enduring protection from severe infection, hospitalization, and death, if not from transmission
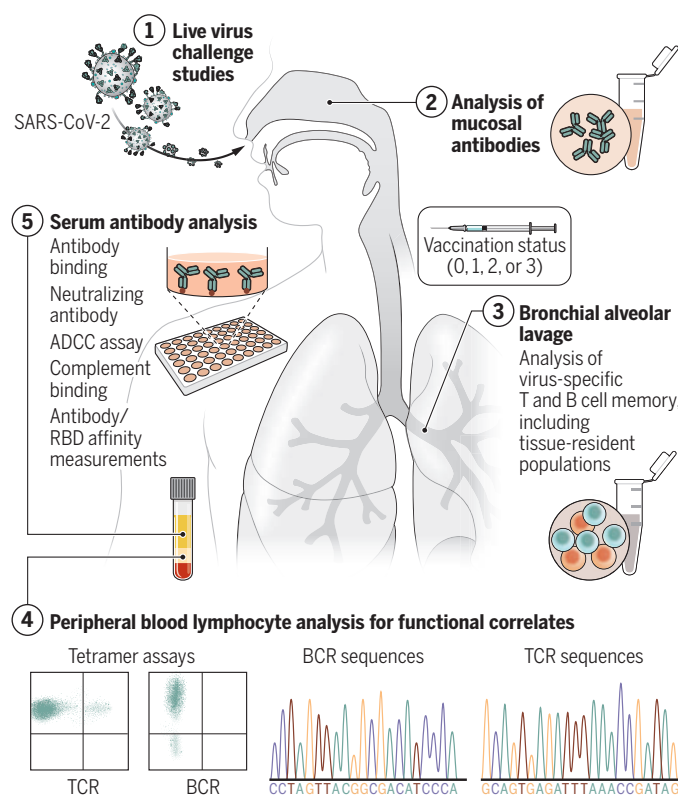
CREDIT: KELLIE HOLOSKI/*SCIENCE*

Fig. 2. Strategies to broaden understanding of systemic and mucosal correlates of protection. To date, most immune analysis derives from monitoring systemic immunity in peripheral blood. Moving forward, it will be important to understand the kinetics and durability of mucosal and lung-resident immunity, with a focus on specific B and T cell immune repertoires. This could include live human challenge models. See (*69, 70*).

and mild disease, probably through activation of memory T and B cell protection. Nonetheless, the disease load from Delta and then Omicron infections and reinfections has been so onerous that many countries have instigated a third vaccine booster dose (*37*). In the US, heterologous boosters are approved at 2 months after the Johnson & Johnson vaccine and 6 months after the mRNA vaccines. In January 2022, Israel became the first country to start a wide-scale fourth-dose booster program. The rapid increase in understanding of immune waning and susceptibility emphasizes the fragility of the datasets that underpin vaccine certification: Knowing that a person was vaccinated decreases the statistical likelihood of being hospitalized (and death), but no more than that, and double- or triple-vaccinated individuals can become infected (sometimes asymptomatically) and transmit SARS-CoV-2 infection.

Among the many vaccination protocol issues needing to be resolved in real time by national advisory bodies is the optimization of sequential booster programs. This encompasses the need to understand the safety and immunology of heterologous, "mix and match" vaccine combinations (*38*). Most countries have to date favored mRNA vaccines for booster programs. Considering the cost and supply-chain issues of

these vaccines, the approach would likely limit wider access to boosters. Use of AZ/ChAdOx1 as a third-dose booster 38 weeks after the second dose is associated with strong immunity, and there is no evidence that repeated use of adenoviral constructs becomes self-limiting through induction of anti-vector Ab (*39*). A recent study compared homologous Coronavac boosting to heterologous adenovirus or mRNA boosters given 6 months from the second dose of Coronavac. Boosting after Coronavac is arguably the "typical" future global challenge, given the billions of doses that have been used (*40*). The adenoviral-vectored vaccines yielded nAb titers that were higher than the previously calculated 90% vaccine efficacy correlate by a factor of 2.4, whereas BNT162b2 yielded a geometric mean that was higher than the 90% correlate by a factor of 4.8 (*41*). That is, nAb levels in each of the heterologous boosted groups, by adenovirus or mRNA, would be associated with high protection against symptomatic infection with a circulating VOC.

Boosting or sharing

It has been overwhelmingly demonstrated that Delta or Omicron breakthrough infection can be considerably mitigated by a third booster shot, as shown, for example, by the rapid success of the booster rollout in Israel (*37*). However, the case has been argued through the World Health Organization and elsewhere that stockpiling of vaccine doses for additional boosts when much of the world is in danger, not yet having accessed a first dose, is both unnecessary and unethical (*7*). It is hard to obtain accurate estimates of vaccine wastage, but estimates range from 20 to 50% of all stocks, including discarded residual vaccine in vials, storage and transport issues, or failure to use vaccines within expiration dates (*42*). Equitable vaccine coverage is inextricably linked to future virus control strategies and the ongoing global percolation and emergence of new VOCs. Policy-makers need to be persuaded that equitable vaccine sharing is not an act of altruism, but rather is essential to limit the emergence of further VOCs, thus shortening the timeline of the global pandemic. The US government recently increased its COVAX pledge of vaccines to 1 billion doses, prompting the question of whether the vaccination shortfall is best addressed through initiatives such as COVAX, through improved regional manufacture of regionally produced and licensed vaccines, or through some combination of these.

The initial vaccine rollouts were dominated by rapid clinical trials and the success of the

SCIENCE science.org

11 MARCH 2022 • VOL 375 ISSUE 6585 **1129**

Pfizer/BioNTech, Moderna, and AstraZeneca vaccines. Now, however, vaccine development, manufacture, and clinical testing take place across multiple countries and continents, as shown by the list of vaccines that were recently authorized or are in phase 3 trials (5). Two major vaccines from India, the inactivated vaccine Covaxin and the DNA vaccine ZyCoV-D, are making a major impact regionally (43, 44). The recombinant spike CORBEVAX vaccine (Baylor College of Medicine) has emergency use authorization in India and is being produced there, with 300 million doses sold to the Indian government (45). As the routes to development of safe and effective COVID-19 vaccines become more established, licensure may become more straightforward. Also, placebo-controlled phase 3 trials will become harder to conduct, empha-

sizing the need for agreed COP that can be reliably used to benchmark vaccine efficacy, as has been done to authorize other vaccines (22). The vaccines that most rapidly rose to prominence were a small subset of all those that had passed through phase 2 and 3 trials; there are many other promising approaches, including self-amplifying RNA, DNA vaccines, adjuvanted recombinant proteins, and approaches based on pan-coronavirus neutralization (8, 46–48) (Table 1). The National Institute of Allergy and Infectious Diseases recently made three large awards for pan-coronavirus vaccine strategies. One successful approach to pan-coronavirus vaccination has been through immunization with 24–amino acid SARS-CoV-2 RBD nanoparticles conjugated to a ferritin scaffold (48). This gave high nAb titers against VOC

as well as for bat coronaviruses. The feasibility of cross-clade vaccination is further supported by studies showing broad neutralization by monoclonal antibodies generated from SARS-CoV-1 infection survivors later given a SARS-CoV-2 mRNA vaccine (47). This suggests that empirical sequential immunization using vaccines based on distantly related S or RBD immunogens, in cross-clade prime/boost protocols, could generate pan-sarbecovirus protection (47). In light of waning vaccine immunity, there is also renewed appeal of looking to the immunology of known, durable-immunity vaccine platforms, exemplified by spike expression in the context of the live-attenuated yellow fever vaccine platform YF17D (49). Considerable efforts are also going into the potential for added value of mucosal vaccines in rapid, local protection against transmission (50).

Table 1. Future approaches to protective immunity against SARS-CoV-2. A spectrum of the key approaches envisaged for the future stages of SARS-CoV-2 vaccination. The opinions range from reliance on immune memory from live infection, to ongoing use of first generation spike vaccines, through to clinical trials of future-proofing using pan-coronavirus strategies.

Future protective immunity approach	Pros	Cons
Stop boosting and rely on existing memory	Cheap, simple; all rely on existing primed immunity	Immunity likely to wane; high probability of long-duration COVID-19 waves
Homologous first-generation spike vaccine boosters frequently (as needed)	Simple, safe, established, with supportive evidence of short-term protection against VOC	May be expensive (depending on which platforms); may be increasingly suboptimal against immune-selected variants; suboptimal immune boosting relative to heterologous boosters
Heterologous first-generation spike vaccine boosters frequently (as needed)	Simple, safe, established, with supportive evidence of short-term protection against VOC; strong supportive evidence for enhanced immunity, including the ability to rescue responses in those who may initially have received weaker vaccines	May be expensive (depending on which platforms); may be increasingly suboptimal against oncoming immune-selected variants
As above, but using spike boosters based on wider rollout of **second-generation** production platforms, e.g., DNA vaccines, self-amplifying RNA, recombinant protein with adjuvant	Likely to be effective, globally scalable, cheap, and therefore advantageous for global coverage	Little data thus far in large-scale heterologous boosting protocols
First-generation platforms modified for **specific** VOC spike inserts	Likely to be highly effective against a given VOC; simple, safe, established	VOC waves tend to arise considerably faster than new vaccines can be modified and tested; unpredictability of protective phenotype due to prior imprinting; wave-specific and lacks future-proofing
First-generation platforms modified for **polyvalent** VOC spike inserts	Likely to be highly effective against multiple VOCs; potential for global relevance; simple, safe, established	Unpredictability of protective phenotype due to prior imprinting; lacks future-proofing
Boost with a wider viral immunome: **Polyvalent adjuvanted proteins**	Accessible technologies: Enhanced potential to avoid immune evasion mutants	Lack of strong evidence to date for additional protection through non-spike immunogens; lack of data in heterologous protocols to date
Boost with a wider viral immunome: **Whole inactive virus**	Accessible with much prior data and production line infrastructure and safety; immunogen-agnostic potential to prime with much of the viral immunome; adaptability to modification for oncoming VOCs	In some cases, whole inactivated virus has been less immunogenic; would not give faithful expression of the full proteome as seen in infection
Boost with **live attenuated virus** carrying polyvalent spike for lifelong durability, e.g., YF17 platform	Good prior track record of YF17D platform	Relatively untried; greater safety concerns
Sequential immunization with spike from SARS-CoV clades for pan-coronavirus coverage	Potential for completely future-proofed pan-coronavirus protection	Would need further development; potential for unpredicted effects of immune imprinting
Immunization with **adjuvanted RBD nanoparticles** for pan-coronavirus coverage	Potential for completely future-proofed pan-coronavirus protection	Would need further development; potential for unpredicted effects of immune imprinting

Future vaccine strategies in the face of emerging variants of concern

There is currently a lack of consensus on precisely where SARS-CoV-2 vaccines are going next and what the coming decade might look like. Policy-makers often state that COVID-19 will be controlled "like flu" with "seasonal jabs." But what does this mean, what level of disease control do we envisage, and crucially, what will be in that jab? Even among immunologists, virologists, and vaccinologists, these topics are subjects of hot debate encompassing the major uncharted research questions. Much of the research effort across the past 2 years has involved the necessity of "building the airplane as we fly." However, faced with complex choices about the future, we need to take the opportunity to slow down and carefully plan the future vaccine strategy. These complex choices relate to a lack of shared agreement and/or understanding as to what we require vaccines to achieve—that is, what are the potential endpoints of an efficacious vaccine (16, 22)? At the start of vaccine trials, endpoints encompassed the ability to significantly reduce new PCR-confirmed infection, to reduce hospital admissions and mortality, or both. Setting the bar for effective protection at different places along the spectrum—prevention of transmission, of symptomatic disease, of requirement for hospital or intensive therapy unit admission, and of fatal outcome—produces different answers. The current situation in many Western European countries as well as in the US is one of good uptake of potent vaccines, yet there is sufficient susceptibility in the population for a high daily caseload. This translates into high hospital admissions and mortality, although with case numbers blunted relative to the infection waves prior to vaccine rollout. However, tolerance of a high level of percolating transmission in the population, even translating into proportionally lower hospitalizations and mortality than in the early waves (this heavily skewed to the unvaccinated), has implications for healthcare provision, for the generation of new VOCs, and for building a substantial, lasting legacy of long COVID, even in the mildly infected. During the Delta and Omicron infection waves, an increased proportion of cases were observed in children. Several countries implemented vaccine programs for teenagers, and a smaller number did so for children in the 5- to 12-year age group. Moving forward, we will need considerable effort to optimize the specific regimen for childhood vaccinations, including decisions about how a new vaccine is best accommodated within the existing framework for childhood vaccine programs.

The vaccine strategy until now has been dominated by targeting spike of the ancestral Wuhan Hu-1 sequence, a single antigen from the large viral immunome (Fig. 1). There may be a price for this very focused approach in driving immune escape (15). Then there is the big question of ability to respond rapidly to forthcoming VOCs.

Possible strategies to consider include a seasonally adjusted VOC spike within existing vaccine platforms (51), pooling sets of VOC sequences for multivalent vaccines, using ancestral and/or VOC inactivated whole virus or antigens to present a wider immunome, or prioritizing rational, pan-neutralizing approaches, epitope-based or otherwise (46–48). As yet, there has been little breathing space for comparative evaluation of these very diverse options with their diverse ramifications.

Which antigens and immune effector pathways?

In terms of the cell biology of RBD/ACE2 binding, it is unsurprising that whether in analysis of human data or in animal models, spike nAb titer is a key COP (21–24). However, given the breadth and complexity of adaptive immunity to this virus, it is important to consider whether other aspects of systemic and mucosal protective immunity should be considered in future approaches and COP analysis (22). The SARS-CoV-2 immunome has been characterized at both the Ab and T cell level (Fig. 1) (52, 53). There are immunoreactive Ab and T cell epitopes in each of the expressed SARS-CoV-2 proteins, probably to the extent of many thousands in total. The problem is that a considerable difference exists between "immunoreactive" and "immunoprotective" recognition. CD4 and CD8 T cell repertoires against peptides from the SARS-CoV-2 proteome are extensive, yet of 29 proteins expressed by SARS-CoV-2, we have unequivocal evidence only for immunity to spike as immunoprotective. However, it is inconceivable that antiviral T cells are not important for broad protection, as supported by studies in animal models (20, 54, 55). A study of highly exposed healthcare workers who show evidence of abortive infection through an initial innate IFI27 signal in the absence of seroconversion finds this to be correlated with a preexisting T cell repertoire against the early-transcribed replication-transcription complex (56).

Immune evasion properties conferred on VOCs by spike mutations have been achieved through disruption of key Ab epitopes, leaving T cell recognition relatively intact (57, 58). This is taken to indicate that unperturbed T cell epitope responses may indeed afford cross-protection against VOCs. However, the predominance of Ab epitope escape mutants may alternatively indicate that during the course of infection, the T cell protective effect is too nuanced to drive viral selection pressure—as seen, for example, from CD8 immune selection pressure on HIV sequences (59). Unlike the patent detection of nAb escape mutations, detection of T cell escape is confounded by the effects of human leukocyte antigen polymorphism and the fact that studies of any T cell selection pressure on SARS-CoV-2 genomes is being probed during short-term, acute infection.

The relative contributions of different immune effector functions have been considered across case reports of human infection in the face of immunodeficiencies affecting specific immune compartments as well as in mouse or macaque challenge (16, 55, 60). The animal models allow the application of Koch's postulates to immune functions that are necessary and sufficient to confer protection, either by depletion of specific cell types or by specific transfer to SARS-CoV-2–challenged recipients. However, even the nonhuman primate models do not always faithfully replicate differences such as the differential pathophysiology of VOCs. From such studies, anti-spike Ab is paramount, specifically in the context of functional nAb titer—that is, not all anti-spike Abs or even all RBD Abs are able to neutralize virus. However, there are additional, non-neutralizing, Fc-dependent effector functions, such as antibody-dependent cellular cytotoxicity, that contribute to protection (61). Furthermore, the subsidiary ligand/receptor interactions beyond spike/ACE2 involved in cellular entry by SARS-CoV-2 are incompletely understood. The animal studies show an additional protective role of CD8 T cells. For example, in the AAV-hACE2 transgenic mouse model of acute infection, depleting CD8 T cells leads to inhibition of viral clearance, with a factor of 7.4 increase in virus (55). If T cells, and especially CD8 T cells, contribute to the early recognition and clearance of virus, this reiterates a need to consider the candidacy of all structural and nonstructural SARS-CoV-2 proteins as vaccine immunogens. Nucleocapsid is highly immunogenic with respect both to the Ab and T cell repertoire and is under development as a vaccine candidate. In animal models, multi-antigen vaccines incorporating nucleocapsid facilitate virus clearance from the upper respiratory tract, although an additive effect of nucleocapsid specific immunity has yet to be defined (1, 62).

While emphasizing the likelihood of residual protection from the diverse, unperturbed and largely intact T cell and memory B cell repertoire (35, 58, 63), we need improved granularity in T cell memory studies, allowing establishment of causal relationships between specific T cell repertoires and protective phenotypes. Differential kinetics and abundance of viral gene expression and display after cellular entry suggest the need for efforts to better elucidate which are the key virus antigen targets of functionally protective T cells. The logic arguing that the building frequency at each antigen exposure of memory T and B cells and the associated antibody affinity maturation must have considerably blunted the pathological impact of the Delta and Omicron waves is inescapable, but confidence in this model must not deter us from designing the studies to prove and maximize the point.

As each of the VOCs has emerged since the Alpha variant in early 2021, the race has been on to characterize the extent to which mutations within targeted epitopes erode vaccine-elicited

protective headroom. Omicron appeared at a time when 6 months of global dominance by Delta had led many to believe that the peak variant may have been reached. The likelihood that the saltational mutations in Omicron had arisen in an immunocompromised individual offers a more daunting vision for the appearance of further VOCs. The vaccines, especially the mRNA vaccines, initially induce such high nAb levels in most people that even the reduced neutralization, most notably against Delta and Omicron VOCs, would be expected to leave a safety margin (*31*). However, with relatively rapid waning, many individuals have serum nAb IC_{50} (half-maximal inhibitory concentration for virus neutralization) values that fall below the notional COP. Attention thus focuses on boosting against breakthrough infections and, more generally, mitigation against forthcoming VOCs. Of all the crystal-ball gazing questions, one of the most important is that of how far along we are in the ongoing arms race against phenotypically significant SARS-CoV-2 new mutations generating VOCs. The period of global spread from the end of 2020 through 2021 generated VOC Alpha through Delta and Omicron, the last two of which were game-changers in terms of a successful vaccine escape route. The appearance of additional mutant sequences is a function of the number of viral copies, that is, the number infected at any given time (and especially, the number of immune-compromised, chronically infected individuals). While additional mutations are being sequenced all the time, Delta had remained in place as the dominant strain in most regions, then displaced because of the additional changes and superior transmission and immune evasion of Omicron. Are there any indications that vaccine programs may be allowing us to outrun SARS-CoV-2 in the arms race? Although structural modeling allows the hypothesis that additional mutations would confer even more devastating phenotypes than Delta and Omicron, and these may indeed emerge, we have previously argued the case that the arms race eventually enters an endgame, set by the point at which advantageous and immune-evasive (or transmission) mutations become offset by mutations that impede viral fitness (*31*). Depending on the will to control ongoing global caseload, that endgame should be tractable.

Nonetheless, the immunology of how best to generate potent protection against current and future VOC spike sequences is not trivial. Starting from a tabula rasa with respect to SARS-CoV-2 immune repertoires, we now have a world of people with diverse histories reflected in the specific B cell and T cell receptor repertoires of their memory populations: presence or absence of acute infection by the ancestral Wuhan Hu-1 virus, and/or by VOC Alpha to Omicron, then variably overlaid with one to four subunit or inactivated virus vaccine doses.

In terms of immune imprinting ("original antigenic sin"), the data show that different repertoires emerge, with associated implications for variable quality and quantity of neutralization of current or future VOC (*35*, *64*–*66*). For example, our comparative analysis of differential VOC neutralization patterns in vaccinees shows the development of imprinted differences between those who had a prior infection with either the ancestral or Alpha virus (*35*). Faced with these diverse scenarios, the question is whether to keep developing boosters carrying prototypic Wuhan Hu-1 spike sequence or focus on being reactive to regionally predominant VOCs. The iteration of this that pools VOC sequences into multivalent vaccines has appeal, although the immune imprinting data argue the potential for unforeseen, differential response patterns dependent on prior history and subsequent SARS-CoV-2 exposure. There is a danger that, even with "plug and play" platforms and rapid pipelines, this entails a future of playing catchup against oncoming VOCs for diminishing and unpredictable returns in protective immunity. Implicit in any VOC mitigation strategy is the notion that our vaccine approaches will need to be specifically linked to strong, networked, international programs of genomic horizon-scanning early warning for oncoming VOCs.

While the heavy lifting for entry into a postpandemic world seems destined to continue with the push to global, vaccine-induced protection complemented by behavioral change, there will also be growing contributions to disease management from advances in point-of-care testing and antiviral drugs such as Merck's Molnupiravir and Pfizer's Paxlovid.

Concluding remarks

In the face of the global cataclysm of the pandemic, the many strands of the vaccine effort have gone outstandingly well. Effective vaccines were rapidly conceived, manufactured at scale, trialed, licensed, and rolled out, with excellent impact. Even some of the more pessimistic projections of vaccine hesitancy were negated by strong uptake in many countries. We now need strong data to address the next unknowns. This encompasses good early warning for emerging VOCs; optimizing future vaccine strategy in terms of design, number of doses, dosing interval, and approaches to achieve safe, durable, vaccine immunity in both children and adults; and characterization of the optimal strategies to generate truly variant cross-protective immunity, irrespective of prior infection history.

REFERENCES AND NOTES

1. L. Dai, G. F. Gao, *Nat. Rev. Immunol.* **21**, 73–82 (2021).
2. J. S. Tregoning, K. E. Flight, S. L. Higham, Z. Wang, B. F. Pierce, *Nat. Rev. Immunol.* **21**, 626–636 (2021).
3. D. S. Khoury *et al.*, *Nat. Rev. Immunol.* **20**, 727–738 (2020).
4. M. Sadarangani, A. Marchant, T. R. Kollmann, *Nat. Rev. Immunol.* **21**, 475–484 (2021).
5. New York Times Coronavirus Vaccine Tracker; www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html.
6. J. H. Kim, F. Marks, J. D. Clemens, *Nat. Med.* **27**, 205–211 (2021).
7. P. R. Krause *et al.*, *Lancet* **398**, 1377–1380 (2021).
8. M. S. Gebre *et al.*, *Cell* **184**, 1589–1603 (2021).
9. S. Pavord *et al.*, *N. Engl. J. Med.* **385**, 1680–1689 (2021).
10. J. Hippisley-Cox *et al.*, *BMJ* **374**, n1931 (2021).
11. A. Huynh, J. G. Kelton, D. M. Arnold, M. Daka, I. Nazy, *Nature* **596**, 565–569 (2021).
12. M. Patone *et al.*, *Nat. Med.* 10.1038/s41591-021-01630-0 (2021).
13. J. S. Scott, A. Richterman, M. Cevik, *BMJ* **374**, n2320 (2021).
14. E. Dolgin, *Nature* **597**, 606–607 (2021).
15. G. A. Poland, I. G. Ovsyannikova, R. B. Kennedy, *Vaccine* **39**, 4239–4241 (2021).
16. S. H. Hodgson *et al.*, *Lancet Infect. Dis.* **21**, e26–e35 (2021).
17. S. Su *et al.*, *Trends Microbiol.* **24**, 490–502 (2016).
18. A. King, *New Sci.* **249**, 12–13 (2021).
19. I. McDonald, S. M. Murray, C. J. Reynolds, D. M. Altmann, R. J. Boyton, *NPJ Vaccines* **6**, 74 (2021).
20. J. Yu *et al.*, *Science* **369**, 806–811 (2020).
21. D. S. Khoury *et al.*, *Nat. Med.* **27**, 1205–1211 (2021).
22. P. J. M. Openshaw, *Science* **375**, 22–23 (2022).
23. P. B. Gilbert *et al.*, *Science* **375**, 43–50 (2022).
24. S. Feng *et al.*, *Nat. Med.* **27**, 2032–2040 (2021).
25. A. W. D. Edridge *et al.*, *Nat. Med.* **26**, 1691–1693 (2020).
26. N. Kaneko *et al.*, *Cell* **183**, 143–157.e13 (2020).
27. K. Lederer *et al.*, medRxiv 21263686 [preprint] (2021).
28. K. W. Cohen *et al.*, *Cell Rep. Med.* **2**, 100354 (2021).
29. C. J. Reynolds *et al.*, *Science* **372**, 1418–1423 (2021).
30. C. Manisty *et al.*, *Lancet* **397**, 1057–1058 (2021).
31. D. M. Altmann, R. J. Boyton, R. Beale, *Science* **371**, 1103–1104 (2021).
32. A. Sokal *et al.*, *Immunity* **54**, 2893–2907 (2021).
33. M. Shrotri *et al.*, *Lancet* **398**, 385–387 (2021).
34. A. Y. Collier *et al.*, *N. Engl. J. Med.* **385**, 2010–2012 (2021).
35. C. J. Reynolds *et al.*, *Science* **375**, 183–192 (2021).
36. E. G. Levin *et al.*, *N. Engl. J. Med.* **385**, e84 (2021).
37. N. Barda *et al.*, *Lancet* **398**, 2093–2100 (2021).
38. X. Liu *et al.*, *Lancet* **398**, 856–869 (2021).
39. A. Flaxman *et al.*, *Lancet* **398**, 981–990 (2021).
40. S. Mallapaty, *Nature* **598**, 398–399 (2021).
41. S. A. Costa Clemens *et al.*, Preprints with The Lancet, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3989848.
42. A. Timsit, How many vaccines go to waste? *Quartz* (2021); https://qz.com/2013918/some-countries-are-wasting-more-covid-19-vaccines-than-others/.
43. A. K. Singh *et al.*, *Vaccine* **39**, 6492–6509 (2021).
44. T. Momin *et al.*, *EClinicalMedicine* **38**, 101020 (2021).
45. "Texas Children's Hospital and Baylor College of Medicine COVID-19 Vaccine Technology Secures Emergency Use Authorization in India" (2021); www.texaschildrens.org/texas-children%E2%80%99s-hospital-and-baylor-college-medicine-covid-19-vaccine-technology-secures-emergency.
46. D. R. Burton, L. M. Walker, *Cell Host Microbe* **27**, 695–698 (2020).
47. C. W. Tan *et al.*, *N. Engl. J. Med.* **385**, 1401–1406 (2021).
48. K. O. Saunders *et al.*, *Nature* **594**, 553–559 (2021).
49. L. Sanchez-Felipe *et al.*, *Nature* **590**, 320–325 (2021).
50. D. Lapuente *et al.*, *Nat. Commun.* **12**, 6871 (2021).
51. A. Choi *et al.*, *Nat. Med.* **27**, 2025–2031 (2021).
52. D. M. Altmann, R. J. Boyton, *Sci. Immunol.* **5**, eabd6160 (2020).
53. A. Sette, S. Crotty, *Cell* **184**, 861–880 (2021).
54. A. Bertoletti, A. T. Tan, N. Le Bert. *Oxford Open Immunol.* **2**, iqab006 (2021).
55. B. Israelow *et al.*, *Sci. Immunol.* **6**, eabl4509 (2021).
56. L. Swadling *et al.*, *Nature* **601**, 110–117 (2022).
57. W. T. Harvey *et al.*, *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
58. C. Riou *et al.*, *Sci. Transl. Med.* **14**, eabj6824 (2021).
59. H. E. Roberts *et al.*, *PLOS Genet.* **11**, e1004914 (2015).
60. K. S. Corbett *et al.*, *Science* **373**, eabj0299 (2021).
61. T. Zohar *et al.*, *Cell* **183**, 1508–1519.e12 (2020).
62. W. Jiang *et al.*, *Cell Rep.* **37**, 110112 (2021).
63. R. R. Goel *et al.*, *Science* abm0829 (2021).
64. T. Aydillo *et al.*, *Nat. Commun.* **12**, 3781 (2021).
65. A. K. Wheatley *et al.*, *Trends Immunol.* **42**, 956–959 (2021).
66. N. Faulkner *et al.*, *eLife* **10**, e69317 (2021).
67. A. Hachim, N. Kavian, S. A. Valkenburg, *Curr. Opin. Virol.* **50**, 139–146 (2021).
68. A. Tarke *et al.*, *Cell Rep. Med.* **2**, 100204 (2021).
69. G. Rapeport *et al.*, *N. Engl. J. Med.* **385**, 961–964 (2021).
70. B. Killingley *et al.*, Research Square [preprint]; www.researchsquare.com/article/rs-1121993/v1.

Stopping pandemics before they start: Lessons learned from SARS-CoV-2

Aled M. Edwards[1]*, Ralph S. Baric[2], Erica Ollmann Saphire[3,4], Jeffrey B. Ulmer[5,6]

The vaccine and drug discovery responses to COVID-19 have worked far better than could have been imagined. Yet by the end of 2021, more than 5 million people had died, and the pandemic continues to evolve and rage globally. This Review will describe how each of the vaccines, antibody therapies, and antiviral drugs that have been approved to date were built on decades of investment in technology and basic science. We will caution that the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus has so far proven a straightforward test of our pandemic preparedness, and we will recommend steps we should undertake now to prepare for, to minimize the effects of, and ideally to prevent future pandemics. Other Reviews in this series describe the interactions of SARS-CoV-2 with the immune system and those therapies that target the host response to infection.

Within 2 years of the discovery of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the concerted effort of funders, industry, academia, government agencies, regulatory bodies, and foundations has yielded a range of remarkably effective vaccines and a number of antiviral therapies. This "rapid" success is based on decades of public and private sector investments in science and technology, including in (i) vaccine rapid response technologies that enabled multiple vaccine candidates to be produced within months of the sequencing of SARS-CoV-2 (*1*); (ii) prototype strain coronavirus research that identified that certain conformations of the Spike (S) glycoprotein, the main protein on the surface of the virus, were likely to confer maximal immunity (*2*); (iii) vaccine studies that had demonstrated good mRNA and viral vector vaccine performance against other, more lethal prototype emerging coronaviruses—Middle East respiratory syndrome coronavirus (MERS-CoV) and SARS-CoV (2003)—in small-animal models of human disease (*3*); and (iv) drug research on emerging coronaviruses (*4*).

Although we should rightly celebrate the technological achievements, we must also acknowledge that SARS-CoV-2 until now may not have represented the strongest test of our preparedness. Many other RNA viruses, such as HIV (*5*) and hepatitis C virus (HCV) (*6*), mutate rapidly, and for HCV and HIV, decades of effort has failed to yield even one effective vaccine. By contrast, the genome of SARS-CoV-2 is relatively stable (*7*), and vaccination with the SARS-CoV-2 S protein, regardless of the technology platform, induces an immune response that significantly reduces hospitalization and deaths, so far even to viral variants (including Omicron) (*8*). Similarly, antibody drugs and small-molecule drugs for COVID-19 were developed rapidly and are proving effective for preventing serious disease. Yet despite these scientific successes, by the end of 2021, more than 5 million people have already died, the virus continues to transmit efficiently across the world, and many regions lack access to these life-saving treatments. Moreover, the SARS-CoV-2 *spike* gene is evolving substantially, which portends cyclical epidemic waves reminiscent of influenza and noroviruses. We need to be far better prepared for the next pandemic.

Box 1. Lessons learned from COVID-19.

- Successful development of vaccines and therapies was built on the foundation of decades of basic and applied research on coronaviruses.
- Concerted efforts by public-private partnerships are essential to an effective rapid response.
- New vaccine technologies and antibody antiviral therapies can be effective in humans and exemplified a "just-in-time" capability to infectious disease outbreaks.
- Repurposed antiviral drugs have shown utility against SARS-CoV-2.
- In the long term, we need vaccines that provide broader and more durable protection.
- We must ensure equitable global access of vaccines and therapies.

The development of SARS-CoV-2 vaccines and therapies has been well documented in the popular press and in scientific reviews. In this Review, we will provide an overview of the progress made so far, provide some ideas about key knowledge and technology gaps, and suggest how the lessons learned can better prepare us for the next pandemic. Treatments designed to counter the out-of-control and sometimes lethal host inflammatory response will not be discussed in this Review but are touched on in a companion Review on the immunology and immunopathology of COVID-19 (*9*).

Coronaviruses and pandemics

The early 21st century has now experienced two epidemics caused by new coronaviruses, SARS-CoV-1 (2003) and MERS-CoV (2012 to present), and now a pandemic caused by SARS-CoV-2 (2019 to present). Although there are some differences in terms of the severe respiratory, end-stage lung diseases and mortality caused by these three viruses, for the purposes of vaccination and therapeutics, there are important commonalities that guide preventative and therapeutic strategies. First, the three coronaviruses share molecular features, from their genes to their overall shapes—meaning that studies with prototype emerging strains are highly portable to the others and future emergent strains. Second, although infections from these viruses all elicit cold symptoms, hospitalization and deaths arise from the combination of viral infection, comorbidities, and an uncontrolled immune response in some patients, which can lead to acute respiratory distress syndrome (ARDS) (*10*, *11*). The aim of vaccination and treatments is to avoid the progression of the disease to this more serious phase. Last, SARS-CoV and MERS-CoV can also induce long-term lung disease; roughly 20 to 30% of patients infected with SARS-CoV, MERS-CoV, and potentially SARS-CoV-2 developed some permanent lung damage (*12*–*14*). If this is indeed the case for SARS-CoV-2, then vaccination is expected to reduce the risk not only of acute disease but also of the potential long-term effects, including pulmonary, neurological, and other symptoms that can extend for months to years after the acute infection (*15*).

Preventing infections with vaccines

Vaccination against SARS-CoV-2 has proven very protective so far, although booster immunizations are clearly required for maintenance of effectiveness over time. Moreover, current vaccines are less efficacious against protecting from infection from recently emerging viral variants such as Omicron, even after a third dose (*16*). For context, however, at the outset of the pandemic the World Health Organization (WHO) and regulatory agencies

[1]Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7, Canada. [2]Rapidly Emerging Antiviral Drug Development Initiative (READDI), Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [3]Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA. [4]Department of Medicine, University of California, San Diego, La Jolla, CA 92037, USA. [5]TechImmune, Newport Beach, CA 92660, USA. [6]Immorna Biotherapeutics, Durham, NC 27703, USA.
*Corresponding author. Email: aled.edwards@utoronto.ca

set a minimum target of 50% efficacy as the threshold for success (*17*). Now, for those fortunate to have access, we are in the unexpected and enviable position of having multiple vaccines that are well over 90% effective in preventing symptomatic disease for most variants and that also reduce hospitalization and deaths for all variants so far (*8*). This success is both deserved and serendipitous: deserved, in that progress can be traced directly to long-term support for vaccine technologies and research in coronavirus biology and immunology; but also serendipitous, in that coronavirus S happens to display many regions that are excellent targets of protective immunity.

By 2019, coronavirus vaccination strategies were well conceptualized through work on prototype human and animal coronaviruses, and the S protein had already been established as the lead candidate immunogen for emerging *sarbecoviruses*, the coronavirus subgenus that includes SARS-CoV-1 and SARS-CoV-2. Structure-guided vaccinology (*18*) had been used to establish that one particular conformation of the S protein, the "prefusion" conformation, was an excellent target for vaccines and could be stabilized through engineering specific mutations in the protein (*2, 19*). mRNA vaccines and recombinant protein vaccines based on prefusion forms of the MERS S protein (MERS S-2P) were shown to protect rodents

from lethal infection with that coronavirus (*3*). When SARS-CoV-2 emerged, the main question was whether these technologies and preclinical advances would translate into successful real-world application in humans.

The seriousness and urgency of the outbreak provided the answer; within a year, more than 150 clinical-stage vaccine programs using a range of approaches were launched (https://covid19.trackvaccines.org), and various approaches proved to be effective. These included established vaccine technologies and vaccine technologies that had not previously achieved licensure for use in humans. Established vaccine technologies being evaluated include live-attenuated viruses, which have been rendered safe through mutation and typically elicit broad and potent humoral (antibody) and cellular immune responses; inactivated viruses, which use a virus inactivated either by heat, ultraviolet light, or through chemical treatment and provide a safer approach with a simplified development path; and recombinant subunit protein vaccines, which focus the immune response against only the key viral proteins of interest. The emerging vaccine technologies in play include those based on nucleic acids (such as plasmid DNA, mRNA, and viral vectors), which mimic a virus infection without the need for live virus, and on self-assembling nanoparticles, which take advantage of the inherent immunogenicity of particulate antigens.

The breadth of vaccine technologies approved for use against SARS-CoV-2 is listed in Table 1.

One key takeaway from the pandemic vaccination efforts is that in part because the S protein is highly immunogenic, most technologies have conferred protective immunity. As such, the outbreak has provided an unparalleled opportunity to evaluate the comparative strengths and limitations of vaccine technologies against multiple metrics, including potency, speed, stability, and ease of manufacturing (Table 1). From a vaccine innovation perspective, the pandemic also provided the means to bring the first nucleic acid–based vaccines to regulatory approval after more than 30 years of development of these technologies (*20*). Particularly impressive have been the real-world data that confirm the efficacy and safety of the mRNA-based vaccines and the speed with which established technologies were harnessed to manufacture the vaccines. The mRNA-1273 (Moderna) vaccine, for example, entered a human phase I clinical trial about 2 months after the sequence of the *spike* gene was made public (*3*). The impact of this work will extend beyond COVID-19; the success of the mRNA vaccines has already emboldened the community to invest in the wider application of this technology to various other infectious disease targets, as well as for applications to oncology, metabolic diseases, gene therapy, and gene editing.

However, although the effectiveness of the vaccines has been impressive, they are far from perfect. Immunity wanes over time, and antigenic variation, such as seen in Omicron, compromises vaccine-induced immunity. This not only could increase the risk of breakthrough infection but also opens the slight possibility that subsequent infection by SARS-CoV-2 may induce vaccine-associated enhanced respiratory disease, as has been noted in animal studies for some vaccines (*21, 22*). An effective short-term solution is to increase the level of immunity with booster immunizations (*23*). Another limitation with current vaccines is the narrow focus of immunity toward the earliest versions of S protein. Although current vaccines are robust against closely matched viral strains, variants with more extensive mutations in the S protein escape many aspects of the vaccine-induced immune response (*24–26*), and periodic boosting or updating of the vaccines are required to control the severity of Omicron infections (*8, 16*). We must develop "broad-spectrum" coronavirus vaccines that will be active against any new, antigenically distinct SARS-CoV-2 variants and novel zoonotic coronaviruses that make the jump from animals to people. To provide broader and durable protection, new vaccine efforts should also target both arms of the adaptive immune system (B and T cell responses) toward regions of viral proteins that are conserved both among

Table 1. SARS CoV-2 vaccine technologies. According to WHO, as of 14 January 2022 these vaccines have been "approved, authorized, licensed, granted emergency use authorization, or made available for use outside of clinical trials via any pathway by a regulatory agency, a national authority, or another entity" in at least one country (https://covid19.trackvaccines.org/agency/who).

Technology	Number approved	Strengths	Limitations
Inactivated virus	8	Established development path	Limited CD8+ T cell immunity
		Extensive human experience	Potency
			Cold storage
Subunit protein	5	Established development path	Limited CD8+ T cell immunity
		Extensive human experience	Cold storage
		Potency	
mRNA	3	Streamlined development	Cold storage
		Broad-based immunity	Limited long-term human experience
		Potency	
		Manufacturing scale up	
Viral vector	6	Broad-based immunity	Cold storage
		Potency	Limited long-term human experience
Plasmid DNA	1	Streamlined development	Potency
		Broad-based immunity	Limited long-term human experience
		Thermostability	
		Manufacturing scale up	
Nanoparticles	1	Potency	Limited long-term human experience
			Cold storage

SARS-CoV-2 variants as well as in related coronaviruses (27, 28). Nucleic acid–based vaccines are particularly well suited to eliciting this type of broad-based immunity.

In addition to the technical innovations described above, four aspects of vaccination implementation should also be improved. First, we need to increase the number of doses that can be manufactured to produce billions of doses rapidly and meet global demand. From a practical perspective, this can be achieved by reducing either the quantity of the active component in the vaccine (for example, mRNA) or the number of doses required for sustained immunity, such as by amplifying the immune response through inclusion of potent adjuvants, particularly for protein-based vaccines, or by using replicating systems such as viral vectors or self-amplifying mRNA. Second, we need to solve several logistical challenges, such as creating and maintaining the infrastructure needed to produce large quantities of vaccine quickly and overcoming cold-chain requirements for those vaccines that are not stable at ambient temperatures. Third, to provide more equitable access, we must create a distributed manufacturing ecosystem that includes sites in developing nations as well as implement a technology transfer system that can rapidly transfer both know-how and manufacturing rights to sites in the distributed network. Fourth, we must better understand the impact of social media, which has proven to be a formidable weapon for the distribution of mis- and disinformation that foments increased vaccine hesitancy and has eroded public health intervention strategies. Enhancement of early education programs on the science, history, and public health benefits of vaccination would be a good place to start.

Preventing disease progression with antiviral therapies

The development of COVID-19 therapeutics has been guided by our increasing understanding of the timing of SARS-CoV-2 infection and COVID-19 disease. Once infected, and depending on the variant and the immune status of the individual, SARS-CoV-2 replication occurs in the upper respiratory tract for 1 to 4 days before the onset of symptoms and peaks before or during the early phases of symptomatic disease. Virus is usually difficult to isolate after about day 7 after symptomatic disease (29), but the viral genomes can be detected for 2 to 3 additional weeks. Early-stage acute infection often is mild and resolves in most people; however, in some cases infection initiates a more serious infection of the lower respiratory tract and induces an immunopathogenic and reparative phase that emerges around 7 to 10 days after symptomatic disease. This immunopathologic stage is characterized by extensive dysregulation of the immune sys-

tem. In the worst cases, this can lead to lung damage, blood clotting problems, and ARDS, which is a severe end-stage lung disease with ~30% mortality rates (30, 31). In some patients whose symptoms have resolved, some immune dysregulation persists (32), potentially leading to chronic disease. The aim of antiviral therapies is to halt progression of COVID-19 before patients advance to more serious acute and/or chronic disease. To this

end, both antibody (delivered with injection or intravenously) and drug approaches (delivered as a pill or intravenously) have been aggressively pursued.

Antiviral drugs

It has been known for more than 50 years that viral infections can be treated by using drugs or combinations of drugs that interfere with viral replication. On occasion, these treatment
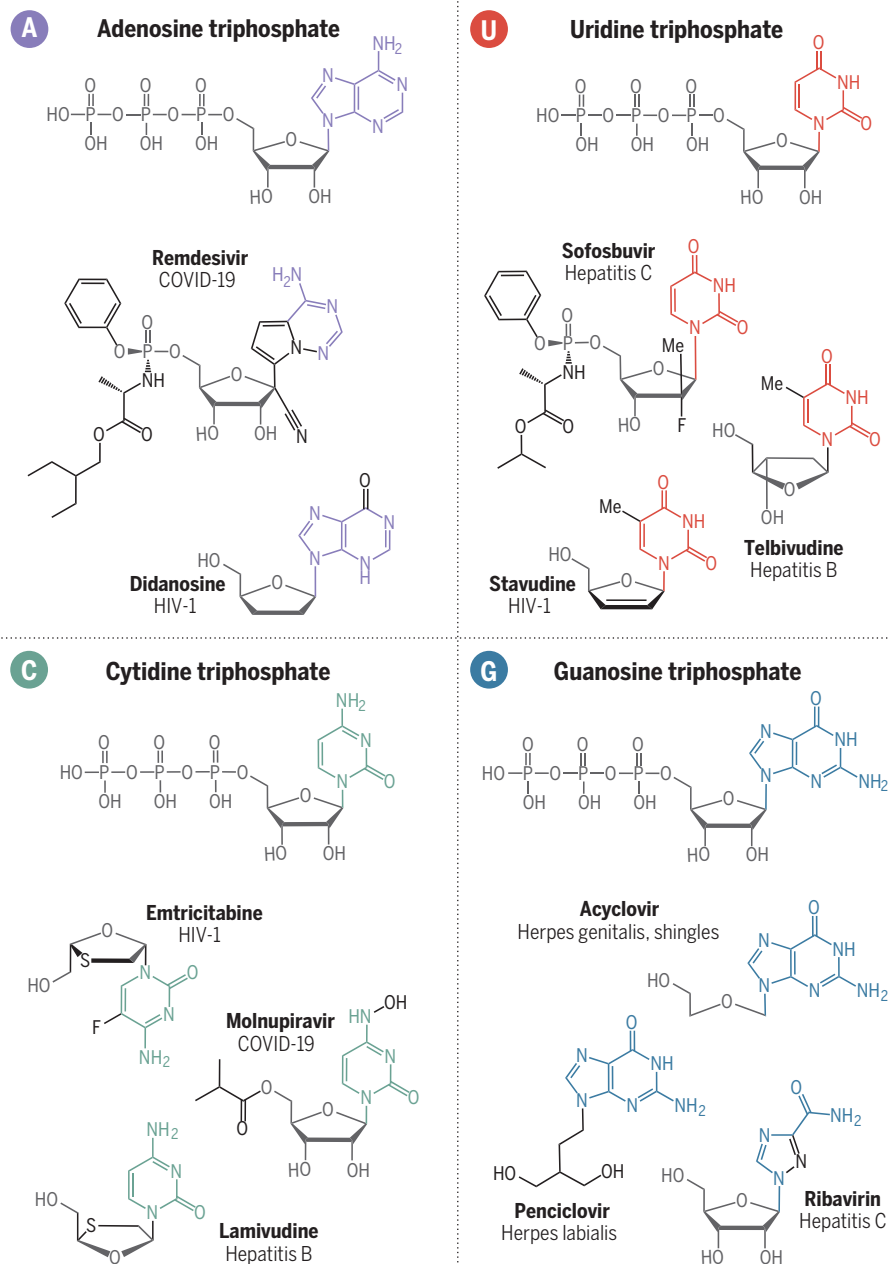

Fig. 1. Where to search for new antiviral drugs. Manmade chemicals that resemble the building blocks of RNA and DNA have proven to be, and promise to be, the richest source of new antiviral drugs. The structures of the building blocks of RNA (A, C, G, and U) are shown in the middle, with their ribose and phosphate parts colored in dark gray and their bases represented in different colors. The structures of an assortment of approved antiviral drugs are displayed, colored to highlight their similarity to the building blocks they are designed to mimic.

approaches have transformed the course of a disease. For example, even though HIV and HCV have resisted all attempts at developing a vaccine, infections that were often fatal can be managed (HIV) or even cured (HCV) by use of antiviral pills.
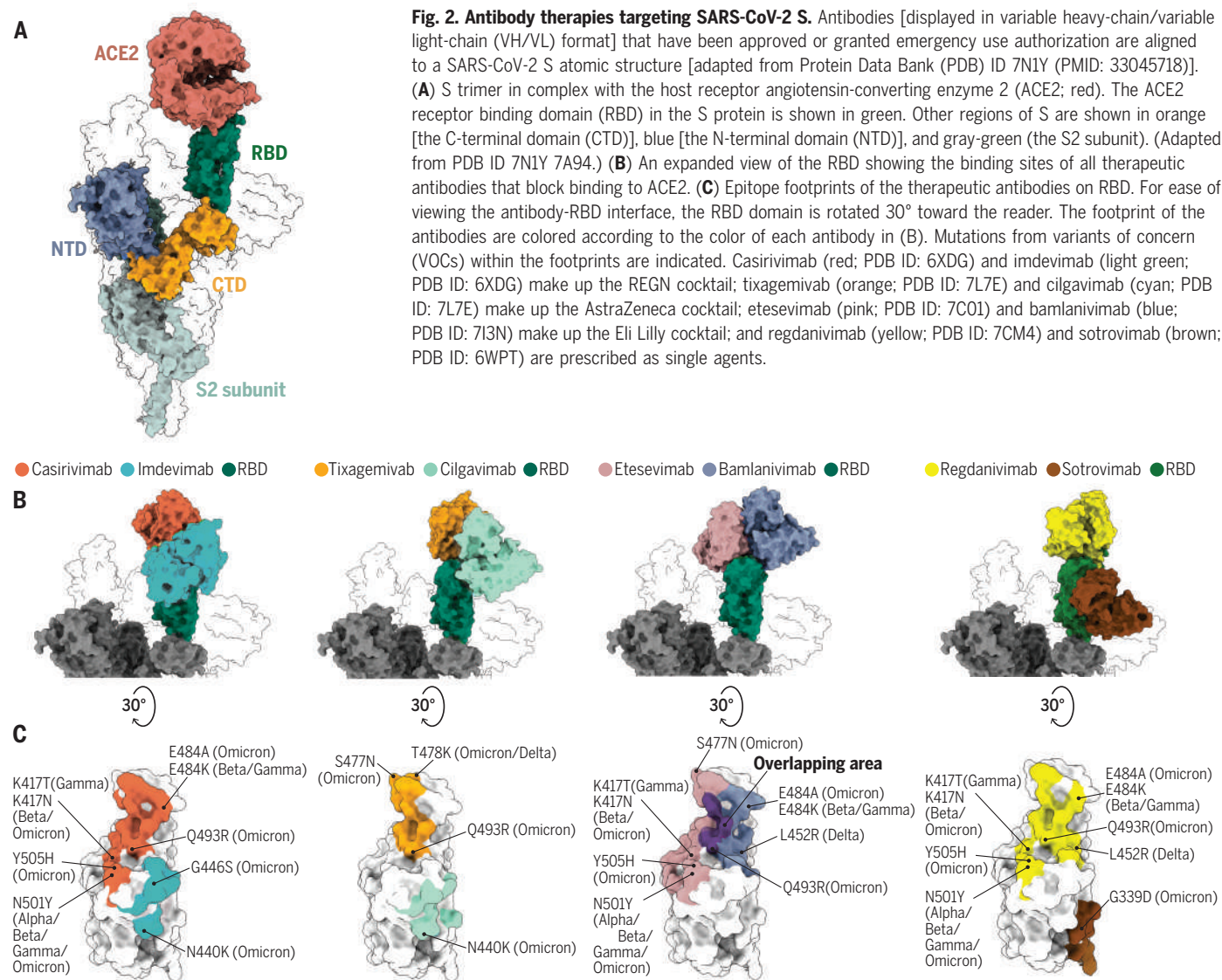
It commonly takes 3 to 5 years or longer to develop a new drug, so when the pandemic exploded and an immediate solution was needed, massive and uncoordinated "drug repurposing" efforts (33) were launched in the hopes of finding COVID-19 drugs among drugs already approved for other diseases. As described in the accompanying Review (9), these drug repurposing efforts showed that several immune modulating drugs improved COVID-19 outcomes. Drug repurposing screens also identified several drugs with antiviral activity. Among these, the only ones that have been approved for use derived from antiviral programs that target other RNA viruses, including other coronaviruses. These repurposing efforts were fruitful because they relied on the structural and functional connectivity among viral drug targets. All RNA viruses, including coronaviruses, encode a common set of enzymes (drug targets) that are essential for replication; these include enzymes that replicate and/or transcribe the genome (RNA polymerases) and enzymes that cleave viral proteins into their active parts (proteases). In practice, this means that the similarity in the chemical reactions that the enzymes perform is mirrored in the similarity of their active sites and thus in the structures of the drugs that act on them. Accordingly, the best place to start fishing for drugs for new RNA viruses is in a pool stocked with compounds that resemble drugs active against other RNA viruses.

All three approved drugs that inhibit SARS-CoV-2 replication originated from drug discovery programs against another RNA virus. Molnupiravir, which acts through the SARS-CoV-2 polymerase (34), was characterized first as an inhibitor of bovine viral diarrhea virus and HCV in 2003 (35). Remdesivir, which also acts through the SARS-CoV-2 polymerase, was originally developed for Ebola (36) starting in 2015. Both remdesivir (37) and molnupiravir (38) had also been shown to block the in vitro and in vivo replication and pathogenesis of other emerging coronaviruses, before the COVID19 pandemic. Nirmatrelvir (39), which inhibits the SARS-CoV-2 M^{pro} protease, is an orally available derivative of a shelved SARS-CoV-1 protease drug developed almost 20 years ago (40). More polymerase and protease drugs from legacy drug programs are in the COVID-19 pipeline.

Molnupiravir, remdesivir, and nirmatrelvir interfere with different stages of the viral life cycle, but there are important commonalities.

Fig. 2. Antibody therapies targeting SARS-CoV-2 S. Antibodies [displayed in variable heavy-chain/variable light-chain (VH/VL) format] that have been approved or granted emergency use authorization are aligned to a SARS-CoV-2 S atomic structure [adapted from Protein Data Bank (PDB) ID 7N1Y (PMID: 33045718)]. (**A**) S trimer in complex with the host receptor angiotensin-converting enzyme 2 (ACE2; red). The ACE2 receptor binding domain (RBD) in the S protein is shown in green. Other regions of S are shown in orange [the C-terminal domain (CTD)], blue [the N-terminal domain (NTD)], and gray-green (the S2 subunit). (Adapted from PDB ID 7N1Y 7A94.) (**B**) An expanded view of the RBD showing the binding sites of all therapeutic antibodies that block binding to ACE2. (**C**) Epitope footprints of the therapeutic antibodies on RBD. For ease of viewing the antibody-RBD interface, the RBD domain is rotated 30° toward the reader. The footprint of the antibodies are colored according to the color of each antibody in (B). Mutations from variants of concern (VOCs) within the footprints are indicated. Casirivimab (red; PDB ID: 6XDG) and imdevimab (light green; PDB ID: 6XDG) make up the REGN cocktail; tixagemivab (orange; PDB ID: 7L7E) and cilgavimab (cyan; PDB ID: 7L7E) make up the AstraZeneca cocktail; etesevimab (pink; PDB ID: 7C01) and bamlanivimab (blue; PDB ID: 7I3N) make up the Eli Lilly cocktail; and regdanivimab (yellow; PDB ID: 7CM4) and sotrovimab (brown; PDB ID: 6WPT) are prescribed as single agents.

First, if given early in the infection, within days of first symptoms, all seem to be effective in preventing progression to serious disease, as seen so far in controlled clinical trials. Second, each of the drugs must be modified to work effectively. Remdesivir and molnupiravir are produced as pro-drugs; each is chemically decorated to make the drug stable in blood as well as to promote efficient uptake into cells, and each is converted inside cells into its active form. Nirmatrelvir does not need to be chemically modified to be functional, but it is not suited to be prescribed alone. Pharmacological studies showed that it was too rapidly inactivated by one of the body's chemical defense enzymes (CYP3A4). The solution was to extend the life span of nirmatrelvir in blood by combining it with ritonavir (*41*), a safe drug that blocks the activity of CYP3A4 and has been used in many other protease-containing drug combinations. It is the combination of nirmatrelvir and ritonavir that is marketed as Paxlovid. Both Paxlovid and molnupiravir are available as oral medications.

The initial results from the clinical trials with these drugs are impressive, but the story is not over. First, although the most advanced clinical studies on these drugs have been reviewed by regulators, as of February 2022, much of the late-stage clinical trial data involving Paxlovid and molnupiravir have not appeared in peer-reviewed journals, and the safety and efficacy data should be made available for scrutiny by the scientific community. Second, there are logistical issues with these medications. Antiviral drugs are most effective if administered soon after infection, and this necessitates molecular diagnosis before treatment. Treatment with remdesivir is further complicated because it must be given intravenously, which places practical limitations on its use, although an oral analog of remdesivir has now been reported (*42*). Third, even though SARS-CoV-2 usually mutates slowly, given that there will be billions of SARS-CoV-2 infections, it is likely that resistance will eventually emerge to one or more of the drugs. The emergence of resistant strains can be delayed by giving the drugs in combination, as demonstrated for HIV and HCV therapies, but the coronavirus combination therapies will have to be first tested rigorously in clinical trials, and this will take time.

The real-world and massive application of drug repurposing also provides key lessons for the next pandemic. First, we learned that the most effective approach to finding drugs or good drug starting points rapidly for any new RNA virus is to start the search before the pandemic within existing drug programs for other RNA viruses—to fish where there are fish, and to test them against prototype viruses in each family. This proved effective for remdesivir and molnupiravir. Second, we

need to increase the size and quality of our antiviral compound libraries. This latter approach must include assembling larger libraries of compounds that mimic the building blocks of the genetic-code polymers (A, G, C, T, and U) because historically, these have proven to be the richest source of antiviral drugs (Fig. 1). Third, our experience with other RNA viruses shows that polymerase and protease drugs should be complemented by drugs to other viral proteins. Fourth, there is no guarantee that the existing antiviral drug libraries or drug programs will yield candidate drugs for all possible viruses. We therefore also need to diversify our antiviral drug pipelines and invest in antiviral programs that focus on other protein targets in the viruses, that target replicase/polymerases and proteases in different ways, and that target host proteins essential for viral replication. Fifth, we should initiate comprehensive drug discovery programs for RNA viruses that have pandemic potential

> "The COVID-19 outbreak demonstrated the strengths and gaps in our pandemic preparedness...and highlights approaches that should be implemented to prevent or lessen the impact of the next pandemic."

now while we have time, focusing on prototype pathogens. These approaches should prioritize oral drugs that act broadly against such virus families (broad-spectrum antivirals) to provide protection against novel viruses that emerge. Any drugs that show promise in cells and show the appropriate safety and pharmacology in animal studies should be advanced through human safety studies (phase I clinical trials) so that they are poised to be trialed in infected patients when the next pandemic emerges.

Antibody drugs

For more than a century, diseases caused by some toxins (*43*) and viruses (*44*) have been prevented by using pre- or postexposure injections of antibodies derived from convalescent animal or human plasma. In many cases, these treatments were not optimally effective because the concentration of neutralizing antibodies in the plasma was too low. In the early 1980s, the advent of techniques to isolate and manufacture individual (monoclonal) antibodies opened the exciting opportunity that we could rapidly develop, design, and purify potent antibodies that could be

given in sufficient quantities not only to prevent but also to treat viral diseases. The excitement was realized quickly; in 1998, the first monoclonal antibody therapy for an infectious disease was approved to prevent infection of neonates with respiratory syncytial virus (*45*). However, widespread adoption since then has been slower than hoped, limited by both scientific and practical issues. On a scientific level, for some viruses, it can be difficult to identify neutralizing antibodies of sufficient breadth and durability. From a practical perspective, monoclonal antibodies are more expensive to manufacture than small-molecule treatments, and historically the treatments have had to be administered in a hospital setting, either with injection or intravenously. However, the relative speed with which antibodies can now be identified and their safety profile compared with other drugs are both major benefits. The Ebola outbreaks and the SARS-CoV-2 pandemic changed the calculus, and there is increasing interest and effort in developing antibodies and antibody-like molecules as antiviral therapies and in improving technologies to better facilitate global access.

The Ebola outbreak in 2018–2019 demonstrated the utility of monoclonal antibodies as antiviral treatments. Both a fixed-dose combination of three monoclonal antibodies (atoltivimab, maftivimab, and odesivimab-ebgn) and a monotherapy (ansuvimab-zykl) directed to a single protein on the Ebola virus surface reduced deaths by almost half in patients already showing symptoms (*46*). Both products have been approved by US regulators as a treatment for Ebola virus disease. As expected, these studies indicate that early treatment with the antibodies would be more effective than is treatment later in disease course.

The Ebola experience emboldened antibody developers to invest in the development of potential therapeutic antibodies for COVID-19 and to test the limits of their technologies and their application to disease. Previous studies on coronaviruses pointed to the S protein as the most suitable target for monoclonal antibody therapy. When the pandemic emerged, the challenge was to identify monoclonal antibodies that bind tightly to the SARS-CoV-2 S protein and neutralize the virus potently. Antibodies with therapeutic potential were identified within weeks by using a range of technologies, and many were rapidly advanced to clinical studies. The results from those first trials were encouraging but not spectacular. In the earliest trials, therapies only marginally reduced hospitalizations and deaths. However, in subsequent trials in which the treatments were given earlier in infection, antibodies demonstrated substantial benefit, roughly equivalent to that seen with some of the antiviral pills.

Mobilizing a new antibody drug against a novel pathogen likely provides one of the

fastest routes to new drug approval: The common framework and safety profiles of antibodies mean a higher rate of success than that of other new treatments. The first monoclonal antibody therapy for COVID-19 (bamlanivimab) entered clinical trials only 5 months after SARS-CoV-2 was identified, and more antiviral monoclonal therapies are likely to be approved in 2022 than have ever been approved in history. Today, there are more than 100 separate efforts targeting the SARS-CoV-2 S protein, with several already proven to be effective in preventing hospitalization and death if given early in the infection.

Five antibody products that block binding of the S protein from early SARS-CoV-2 variants to its receptor on cells (Fig. 2) have thus far been approved or authorized for use in patients as treatments in early COVID-19. These include both monotherapies [regdanvimab (47) and sotrovimab (48)] and cocktails [bamlanivimab + etesevimab (49); casirivimab + imdevimab (50, 51); and tixagevimab + cilgavimab (52)]. Some are also approved or authorized for use to prevent infection in high-risk individuals, such as family members of COVID patients (50). Many of these treatments are predicted on the basis of in vitro studies to be rendered less effective by the multiple mutations in Omicron (53, 54) (Fig. 2), but sotrovimab, for example, which was developed from a survivor of SARS-CoV-1, is predicted to remain effective (55, 56). Emergence of variants of concern has underscored the previously anticipated need to identify antibodies against conserved regions less susceptible to mutagenic escape. Further, the emergence of variants also supports the need for multiple potential therapeutics at the ready, so that replacement drugs are available to respond to changing waves of viral sequences.

Antibodies broadly active against multiple variants and multiple related viruses are the focus of current development efforts. At least one international consortium is systematically mapping the binding of and measuring the activities of hundreds of different antibodies (57). Antibodies that recognize the receptor-binding motif of the S protein can provide the most potent neutralization in vitro but are among the most susceptible to change in viral variants. Those antibodies that continue to bind despite the emergence of variants will provide more durable treatment options. Some antireceptor binding motif antibodies exhibit broad neutralization, as do antibodies that target other, more conserved regions of the S protein (57–63). Further, cocktails of antibodies mitigate escape during treatment and increase chances of success against unexpected mutations (64, 65). Promising antibodies can also be reengineered to increase their activity and the length of time that they persist in the body (22). Although the field is still learning
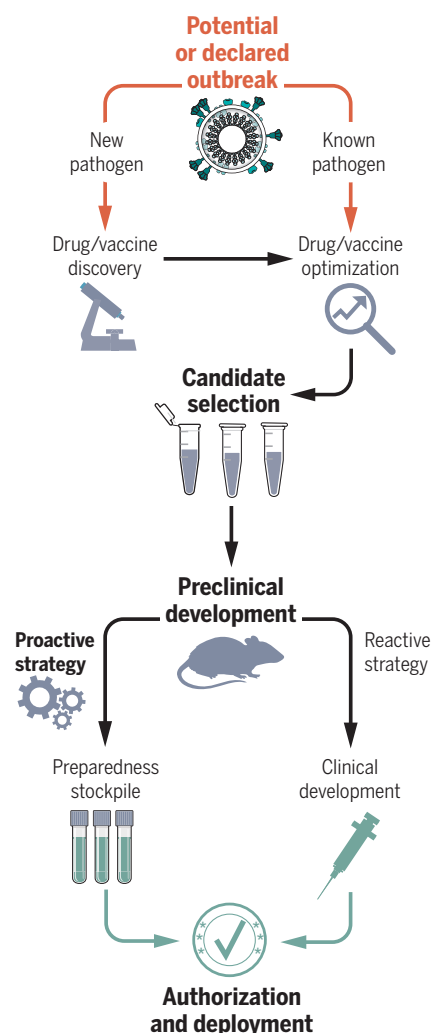
Fig. 3. Response to real or potential infectious disease outbreaks. A comprehensive approach to preparing for future infectious disease outbreaks must include both proactive and reactive approaches. Proactive discovery and development efforts toward vaccines and therapeutics against pathogens with pandemic potential should proceed now to a point at which drugs and vaccines clinically proven to be safe could be stockpiled. In the advent of an outbreak, these assets would be rapidly mobilized for clinical development, authorization, and deployment. In addition, to guard against emerging pathogens, we need to invest in rapid-response vaccine and drug technologies for the discovery, optimization, development, authorization, and deployment of effective countermeasures.

what modifications are optimal, there is great optimism as to its potential.

The main hurdles to widespread use of antibodies are the cost of manufacturing and the logistics of treatment. Both of these hurdles are the focus of substantial public and private investment. Efforts are focused on reducing the manufacturing scale and developing intra-

muscular, inhaled, and other alternatives to intravenous delivery. There remain scientific hurdles to our understanding of how to measure and predict the protective activities of antiviral antibody therapeutics as well. At the heart lies the "black box" of how to elicit the best immune-related protection. Neutralization, the mechanical block of viral entry, can be recapitulated in cell culture, rodent, and primate models, but the ways in which the antibodies interact with the immune system to sound an alarm, clear pathogen and infected cells, and inspire other protective functions are more difficult to measure. These interactions are more complex, may be transient, and differ among model species. Understanding which antibodies are best, the features that make them best, and the differences between what types of antibodies are needed pre-exposure versus post-exposure is important for the development of future antibody drugs and is also key information that can inform vaccine design.

Preventing future pandemics

The COVID-19 outbreak demonstrated the strengths and gaps in our pandemic preparedness (Box 1) and highlights approaches that should be implemented to prevent or lessen the impact of the next pandemic (Fig. 3). Lessons learned include the following.

First, partnerships involving the public sector, the private sector, and nonprofits are central elements of an effective pandemic response, but we must make real effort, before work starts, to agree to terms that ensure timely and equitable global access to technologies, know-how, and products.

Second, an array of new vaccine technologies—such as mRNA, plasmid DNA, viral vectors, and nanoparticles, which had been supported by 30 years of research but hitherto not achieved licensure for human use—can be safe and effective. In addition, the speed by which these technologies were harnessed also shows that a true rapid-response vaccine platform is feasible, as well as a "just-in-time" approach to address new SARS-CoV-2 variants. The availability of a range of effective vaccine technologies also positions us well to generate candidate vaccines in advance of the next pandemic. We must prospectively develop potential vaccines for all viruses with pandemic potential.

Third, our therapeutic antiviral drug discovery pipelines are well positioned to respond to viral threats. Within 2 years, they delivered effective antibody-based treatments and highly effective drugs. Two of these drugs have been demonstrated to be highly potent against many, genetically distinct coronaviruses. However, more than 5 million people have died from COVID-19 in this period. To minimize loss of life in future pandemics, we must also prospectively produce broad-spectrum antiviral therapies for

viruses with pandemic potential. Ideally, these therapies should be in the form of a pill, to increase their ability to be distributed to high-need, low-resource settings. The treatments must also be accessible and not restricted by the private sector or geopolitical forces. We advocate for these activities to take place in the public domain (66, 67) and be coordinated by nonprofits (such as www.readdi.org, https://dndi.org, https://vimiopen.org, and https://carb-x.org) that have clear global access mandates.

Fourth, there is a difference between having a vaccine in hand and making it widely available. As detailed in the previous section, the logistics of vaccine deployment, including manufacturing and access, need to be improved. The Coalition for Epidemic Preparedness Innovations (CEPI; https://cepi.net) is a global partnership that has been launched to prevent future epidemics, but more efforts and international support are needed to maximize its impact. These technical preparations must be complemented by the need for studies of human behavior, to improve compliance and develop strategies to counter misinformation.

Last, more applied and basic research is needed. SARS-CoV-2 was vulnerable to countermeasures developed by using our current scientific pandemic preparedness technologies, but performance has been stressed by the emergence of new, antigenically distinct variants, such as Omicron. A wide variety of technologies succeeded, to varying degrees, but it is also uncertain whether the next pandemic virus will be vulnerable to these approaches or even that the next pandemic will be caused by a virus. In the absence of foresight as to what will come, we need to invest in applied research on new technologies to facilitate the discovery and development of new drugs and vaccines. We must also invest widely and substantially in basic research on microbes of pandemic potential, on viral evolution and antigenic drift, on the pathophysiology of infectious diseases, on human immunology, and on their intersection.

The costs of these initiatives will be substantial, yet trivial in comparison with the impact of COVID-19 on human health (both physical and mental) and world economies, which in 2020 alone was estimated at more than $2 trillion (68).

REFERENCES AND NOTES

1. N. Pardi, M. J. Hogan, D. Weissman, *Curr. Opin. Immunol.* **65**, 14–20 (2020).
2. J. Pallesen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7348–E7357 (2017).
3. K. S. Corbett *et al.*, *Nature* **586**, 567–571 (2020).
4. T. P. Sheahan *et al.*, *Sci. Transl. Med.* **9**, eaal3653 (2017).
5. J. H. Lee, S. Crotty, *Semin. Immunol.* **51**, 101470 (2021).
6. J. D. Duncan, R. A. Urbanowicz, A. W. Tarr, J. K. Ball, *Vaccines* **8**, 90 (2020).
7. G. Tonkin-Hill *et al.*, *eLife* **10**, e66857 (2021).
8. M. G. Thompson *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **71**, 139–145 (2022).
9. M. Merad, C. A. Blish, F. Sallusto, A. Iwasaki, *Science* **375**, XXXX (2022).
10. H. A. Rothan, S. N. Byrareddy, *J. Autoimmun.* **109**, 102433 (2020).
11. X. Li, X. Ma, *Crit. Care* **24**, 198 (2020).
12. P. Zhang *et al.*, *Bone Res.* **8**, 8 (2020).
13. O. O'Sullivan, *Clin. Med. (Lond.)* **21**, e68–e70 (2021).
14. F. Wang, R. M. Kream, G. B. Stefano, *Med. Sci. Monit.* **26**, e928996 (2020).
15. H. Ward *et al.*, *Lancet* **398**, 2057–2059 (2021).
16. W. F. Garcia-Beltran *et al.*, *Cell* **185**, 457–466.e4 (2022).
17. WHO, WHO Target Product Profiles for COVID-19 Vaccines (2020); www.who.int/publications/m/item/who-target-product-profiles-for-covid-19-vaccines.
18. B. S. Graham, M. S. A. Gilman, J. S. McLellan, *Annu. Rev. Med.* **70**, 91–104 (2019).
19. C. L. Hsieh *et al.*, *Science* **369**, 1501–1505 (2020).
20. J. B. Ulmer, M. A. Liu, *Molec. Front. J.* **5**, 1–20 (2021).
21. M. Bolles *et al.*, *J. Virol.* **85**, 12201–12215 (2011).
22. A. T. DiPiazza *et al.*, *Immunity* **54**, 1869–1882.e6 (2021).
23. Y. M. Bar-On *et al.*, *N. Engl. J. Med.* **385**, 1393–1400 (2021).
24. P. Naaber *et al.*, *Lancet Reg. Health Eur.* **10**, 100208 (2021).
25. A. Choi *et al.*, *J. Virol.* **95**, e0131321 (2021).
26. V. Servellita *et al.*, *Nat. Microbiol.* **7**, 277–288 (2022).
27. Y. Goldberg *et al.*, *N. Engl. J. Med.* **385**, e85 (2021).
28. S. J. Thomas *et al.*, *N. Engl. J. Med.* **385**, 1761–1773 (2021).
29. W. A. Fischer 2nd *et al.*, *Sci. Transl. Med.* **14**, eabl7430 (2022).
30. A. B. Pavel *et al.*, *Front. Genet.* **12**, 706902 (2021).
31. R. Kumar *et al.*, *ERJ Open Res.* **6**, 00405-2020 (2020).
32. C. Phetsouphanh *et al.*, *Nat. Immunol.* **23**, 210–216 (2022).
33. A. Edwards, *J. Chem. Inf. Model.* **60**, 5727–5729 (2020).
34. C. J. Gordon, E. P. Tchesnokov, R. F. Schinazi, M. Götte, *J. Biol. Chem.* **297**, 100770 (2021).
35. L. J. Stuyver *et al.*, *Antimicrob. Agents Chemother.* **47**, 244–254 (2003).
36. T. K. Warren *et al.*, *Nature* **531**, 381–385 (2016).
37. C. J. Gordon, E. P. Tchesnokov, J. Y. Feng, D. P. Porter, M. Götte, *J. Biol. Chem.* **295**, 4773–4779 (2020).
38. M. L. Agostini *et al.*, *J. Virol.* **93**, e01348 (2019).
39. D. R. Owen *et al.*, *Science* **374**, 1586–1593 (2021).
40. B. Boras *et al.*, bioRxiv 293498 [Preprint] (2021).
41. M. W. Hull, J. S. Montaner, *Ann. Med.* **43**, 375–388 (2011).
42. A. Schäfer *et al.*, bioRxiv 460111 [Preprint] (2021).
43. G. B. Dunmire, *JAMA* **23**, 853–857 (1893).
44. J. Stokes Jr., J. R. Neefe, *J. Am. Med. Assoc.* **127**, 144–145 (1945).
45. *Pediatrics* **102**, 531–537 (1998).
46. S. Mulangu *et al.*, *N. Engl. J. Med.* **381**, 2293–2303 (2019).
47. Y. Y. Syed, *Drugs* **81**, 2133–2137 (2021).
48. A. Gupta *et al.*, *N. Engl. J. Med.* **385**, 1941–1950 (2021).
49. M. Dougan *et al.*, *N. Engl. J. Med.* **385**, 1382–1392 (2021).
50. M. P. O'Brien *et al.*, *N. Engl. J. Med.* **385**, 1184–1195 (2021).
51. E. D. Deeks, *Drugs* **81**, 2047–2055 (2021).
52. *JAMA* **327**, 384–385 (2022).
53. D. Mannar *et al.*, *Science* **375**, eabn7760 (2022).
54. K. G. Nabel *et al.*, *Science* **375**, eabl6251 (2022).
55. L. A. VanBlargan *et al.*, *Nat. Med.* (2022).
56. E. Cameroni *et al.*, *Nature* 10.1038/s41586-021-04386-2 (2021).
57. K. M. Hastie *et al.*, *Science* **374**, 472–478 (2021).
58. S. Changrob *et al.*, *mBio* **12**, e0297521 (2021).
59. C. A. Jette *et al.*, *Cell Rep.* **36**, 109760 (2021).
60. D. Pinto *et al.*, *Science* **373**, 1109–1116 (2021).
61. M. A. Tortorici *et al.*, *Nature* **597**, 103–108 (2021).
62. H. Liu *et al.*, *Immunity* **53**, 1272–1280.e5 (2020).
63. M. Yuan *et al.*, *Science* **369**, 1119–1123 (2020).
64. J. Dong *et al.*, *Nat. Microbiol.* **6**, 1233–1244 (2021).
65. D. M. Weinreich *et al.*, *N. Engl. J. Med.* **384**, 238–251 (2021).
66. F. von Delft *et al.*, *Nature* **594**, 330–332 (2021).
67. A. Morris *et al.*, *Chem. Commun.* **57**, 5909–5912 (2021).
68. Z. Kolahchi *et al.*, *Adv. Exp. Med. Biol.* **1318**, 825–837 (2021).

RESEARCH

VOLCANOLOGY
Water-controlled magma depth

Magma is stored deep underground until something triggers an eruption or it cools into a pluton. One traditional view of why the magma stays where it does is the assumption that it has the same density as the surrounding rock, keeping it from ascending upward. Rasmussen *et al.* found that the amount of water that arc magmas have in them determines their depth. Degassing water changes the viscosity, allowing it to stall out at a depth where it is still buoyant. This situation may provide some additional buoyancy to get the magma to the surface during an eruption. —BG   *Science*, abm5174, this issue p. 1169

Half Dome at Yosemite National Park, an ancient pluton formed of arc magma that solidified underground

CELL BIOLOGY
Monitoring cell death in vivo

Programmed cell death and the subsequent clearance (efferocytosis) of the cell corpses are essential for organismal development and homeostasis. However, uncleared corpses are rarely detected in vivo, even in tissues with high cell turnover. Raymond *et al.* engineered a genetically encoded dual reporter called CharON to track cell death and efferocytosis in real time within developing fruit fly embryos. CharON revealed previously unrecognized features of macrophage efferocytosis in vivo: rapid clearance with tremendous heterogeneity in corpse burden between macrophages plus a disconnect between ingestion and digestion in macrophages eating multiple corpses successively. These high-burdened macrophages also altered their clearance of dying cells after inflammatory insults. —SMH

Science, abl4430, this issue p. 1182

EVOLUTIONARY BIOLOGY
How organisms make minerals

For more than 500 million years, Earth's organisms have had the ability to make calcium carbonate structures through a process called biomineralization. Such capacity enables the construction of skeletons, shells, teeth, and beaks, although it evolved differently in different phyla. In a Review, Gilbert *et al.* propose an integrated mechanistic model for calcium carbonate biomineralization. Based on constraints derived from established evolutionary history, omics, and a meta-analysis of isotopic data, the model treats the biomineralization process as the result of a single, unifying set of processes. In the authors' view, the importance of specific processes varies among different organisms. This would explain the known divergent evolutionary pathways of different phyla. By providing this unifying perspective, the authors aim to provide a better understanding of both past extinctions and the capacity of life to adapt to climate change. —KVH

Sci. Adv. 10.1126/sciadv.abl9653 (2022).

Shells, bones, and other calcium carbonate structures are proposed to have arisen through convergent evolution in different lineages.

DENDRITIC CELLS
Encoding human dendritic cell fate

Recent advances in human dendritic cell (DC) biology have been accelerated by single-cell technologies, but DCs remain challenging to study because of their rarity and localization within tissues. Rosa *et al.* used single-cell transcriptomics to guide reprogramming of human fibroblasts into induced DCs and benchmarked their function against blood-derived DCs. Enforced expression of three transcription factors, PU.1, IRF8, and BATF3, induced a small

population resembling type 1 conventional DCs (cDC1s), which could be improved to more than 75% efficiency with a constitutive promoter and a trio of inflammatory cytokines. PU.1 chromatin binding early during reprogramming recruited IRF8 and BATF3 to facilitate cDC1-associated gene expression. Systems to generate functional, patient-specific DCs represent a promising tool for studying human DCs and developing novel immunotherapies. —CO

Sci. Immunol. **7**, eabg5539 (2022).

MAGNETISM
Magnetism hits the high notes

The generation and propagation of magnetic excitations such as magnons and spin waves in ferromagnetic thin films provides a platform for the development of spin-based device technology. Koerner *et al*. report measurements on the magnetization dynamics of a nickel–iron film excited coherently by microwave magnetic fields from a coplanar wave guide. Using nitrogen vacancy center–based magnetometry and a time-resolved magneto-optical Kerr effect, the authors show that low excitation frequencies and low bias fields in the range of only a few milli-tesla result in the generation of magnons emitted at higher frequency. Extending over 60 harmonics of the excitation frequency, such upconversion of magnetic excitation frequencies should prove useful for spintronics applications. —ISO

Science, abm6044, this issue p. 1165

NEUROSCIENCE
Boosting endogenous opioid signaling

Angiotensin-converting enzyme (ACE) is expressed in brain tissue, but the central function of ACE in the brain has proven enigmatic. Trieu *et al*. discovered that ACE has a noncanonical role in governing endogenous opioid signaling in the brain. ACE cleaves and degrades an unconventional enkephalin called Met-enkephalin-Arg-Phe (MERF). Unlike conventional enkephalin pentapeptides, MERF is selectively degraded by ACE and enhances μ-opioid receptor activation in the nucleus accumbens, perhaps explaining its antidepressant effects in patients taking ACE inhibitors. —PRS

Science, abl5130, this issue p. 1177

GLOBAL PRODUCTIVITY
Uncovering an uncoupling

The amount of CO_2 in the atmosphere and global biosphere productivity are closely related, but sometimes they do not quite covary. Yang *et al*. present a triple-isotope record of O_2 for the past 800,000 years, which reveals how biospheric productivity has varied over that time and provides a negative feedback that can help atmospheric CO_2 fluxes decouple from sea level, typically before deglaciations (see the Perspective by LeQuéré and Mayot). This finding supports the idea that low CO_2 depresses global photosynthesis and reduces further CO_2 drawdown. —HJS

Science, abj8826, this issue p. 1145; see also abo1262, p. 1091

CHROMOSOMES
Finding and replenishing chromosome ends

Telomeres cap the ends of eukaryotic chromosomes and are vital for maintaining genome stability. The enzyme telomerase helps to compensate for the telomere shortening that occurs with each cell division, thereby prolonging cellular life span. Sekne *et al*. report cryo–electron microscopy structures of human telomerase bound to telomeric DNA and telomerase recruitment factors. The structures reveal how the recruitment factors facilitate telomerase engagement with the telomeric DNA for efficient extension of telomeres. —DJ

Science, abn6840, this issue p. 1173

IN OTHER JOURNALS
Edited by **Caroline Ash** *and* **Jesse Smith**

ECOTOXICOLOGY
Lead in the details

There is increasing evidence that lead from ammunition has entered the food chain of many ecosystems, largely through the activities of scavengers, including the otherwise iconic American bald eagle. Understanding the effects of this chronic toxic load is challenging, however, especially in cases where data collection is not designed for this purpose. Hanley *et al*. examined count and necropsy data over 30 years from bald eagles in the northeastern United States to determine if and how lead has affected the population. Although the eagle population size has increased as the species recovers from the influence of DDT, differences were discerned in population structure between lead-exposed and lead-free bird populations. In lead-exposed groups, breeding individuals and hatchlings showed decreased survival, and population reorganization resulted in more nonbreeding females. Although these changes in population structure often go unnoticed, they can affect eagle resilience over time, and stricter control on the use of lead in ammunition is required to resolve this. —SNV   *J. Wildl. Manage.* **86**, e22177 (2022).

Scavenging bald eagles can be exposed to the toxic effects of lead from ammunition.

NEUROSCIENCE
Astrocytes source glutamine

Because de novo glutamate synthesis does not take place at nerve terminals, its uptake by astrocytes has been postulated to fuel a local mechanism to recycle glutamate to neurons. Cheung *et al*. assessed the direct roles of astroglial glutamine supply during physiological neuronal activity. The authors developed a

POLLUTION
Pharmaceuticals foul freshwater

Contamination of water resources by medicinal chemicals has been monitored patchily in some parts of the world, and unfortunately, the collective data are often incomplete or missing. Using a standardized analysis, Wilkinson *et al.* achieved a global view of the scale of pollution by 61 pharmaceutically active substances by sampling 258 rivers in 104 countries. The greatest levels of river pollution were seen in low- to middle-income countries, and were associated with some affluence, weak regulation, and poor wastewater management. Some chemicals have a geographical signature, such as artemisinin in Africa and ketoconazole in Asia, and many are anticipated to interact with other pollutants to exceed threshold effects. The most abundant chemicals found included carbamazepine, metformin, and caffeine (also used recreationally) in concentrations considered unsafe for aquatic organisms and potentially also for human consumers of those waters. —CA

Proc. Natl. Acad. Sci. U.S.A. **119**, e2113947119 (2022).

Pollution of rivers by pharmaceuticals is a concern in low- to middle-income countries, where regulation may not be fully developed.

---

fluorescent rhodamine–tagged glutamine probe to visualize glutamine directly in live cells from mouse brain tissue. This approach revealed an activity-dependent mobilization of glutamine from astroglial networks into adjacent presynaptic structures. The exchange process was mediated by connexin 43 (Cx43), an astroglial protein with both gap-junction and hemi-channel functions. This supply of glutamine is essential for sustaining physiological glutamatergic synaptic transmission and cognitive performance during recognition memory. —SMH

Nat. Commun. **13**, 753 (2022).

NEUROSCIENCE
Differentiating sense into action

How does the brain decide when to translate perception or cognition into action? Jagadisan and Gandhi studied how action potentials from external sensory inputs or internally generated commands are translated in a brain region called the superior colliculus into eye movements called saccades. Monkeys were trained to detect a visual stimulus

but also to be able to delay their saccades. Microstimulation experiments and computational modeling complemented electrophysiological recordings. Although both sensory and movement-related signals involved increased rates of neuronal activity, they differed in their temporal structure. The neural population vector became very stable just before a movement. By contrast, other responses in the same population were equally large but temporally unstable and therefore did not cause eye movements. —PRS

Curr. Biol. 10.1016/j.cub.2022.01.015 (2022).

METALLURGY
Temperature-invariant alloy

Alloys with properties that change little to not at all as a function of temperature are naturally appealing for certain applications. That certainly is the case for alloys that exhibit an Elivar effect, in which the elastic modulus is nearly constant over a given temperature range. He *et al.* found a cobalt–nickel–hafnium–titanium–zirconium high-entropy

alloy that has this sort of invariance between room temperature and 900 kelvin. This type of alloy could find uses in high-precision devices and space applications, in which materials experience wide temperature fluctuations. —BG

Nature **602**, 251 (2022).

OPTOELECTRONICS
Actively controlling metasurfaces

Over the past decade, the functionality of many bulk optical components has been reduced in scale to that of engineered layers of thin films, or metasurfaces, just several hundreds of nanometers thick. The first generation of metasurfaces were passive devices with function that was fixed after fabrication. Recent efforts have been applied to developing active metasurfaces for which the function can be switched or varied by external stimulation. Weiss *et al.* demonstrate an electrically tunable metasurface using a thin film of lithium niobate to modulate the reflectance of telecom wavelength light. Electrical control provides a route to develop on-chip integrated optical devices

that are lightweight and have a smaller footprint. —ISO

ACS Photonics **9**, 605 (2022).

BIOFUELS
Good intentions are not enough

The Renewable Fuel Standard (RFS) specifies how biofuels must be blended into the transportation fuel supply in the United States in an attempt to reduce anthropogenic carbon dioxide emissions. Does the program live up to that expectation? Lark *et al.* show that the production of corn-based ethanol in the United States has not met the policy's greenhouse gas emissions targets because the carbon intensity of that corn ethanol is no less than that of gasoline and likely nearly one-quarter higher. Furthermore, corn ethanol production has degraded water quality, reduced the area of land available for conservation, and had a negative impact on important ecosystem services. These findings call into question whether the use of corn ethanol actually can help to meet climate mitigation goals. —HJS

Proc. Natl. Acad. Sci. U.S.A. **119**, e2101084119 (2022).

Edited by **Michael Funk**

NEUROSCIENCE
Cortical microcircuits in humans and mice

Insights into the organization and dynamics of local neo-cortical circuits are central to our understanding of brain processing. Neither macro-scopic connectome analyses nor ultramicroscopic con-nectome studies provide such information. Using slice electrophysiology, Campagnola *et al.* collected a very large dataset to study the organiza-tion of cortical microcircuits in mice and humans. They analyzed the strength and probability of chemical and electrical connections; the latency and kinetic properties of synaptic responses; and cell classification features such as morphology, intrinsic physiol-ogy, transgenic markers, and cortical layer. The authors also modeled quantal release and short-term plasticity. Synaptic variability was a primary driver of cross-subclass differences. Compared with the connec-tions in mice, human excitatory synapses had lower variability. —PRS

Science, abj5861, this issue p. 1144

CELL BIOLOGY
Tracking proteins

Improved understanding of how proteins are organized within human cells should enhance our systems-level understand-ing of how cells function. Cho *et al.* used CRISPR technol-ogy to express more than 1000 different proteins at near endogenous amounts with labels that allowed both fluorescent imaging of their location and immunoprecipita-tion and mass spectrometry analysis of interacting protein partners (see the Perspective by Michnick and Levy). The large-scale data are made avail-able on an interactive website, with clustering and analysis performed by machine learning.

The studies emphasize the unusual properties of RNA-binding proteins and indicate that protein localization is very specific and may allow predic-tions of function. —LBR

Science, abi6983, this issue p. 1143; see also abo2360, p. 1093

MOLECULAR TRANSPORT
DNA walkers handle molecular cargos

Controlling the long-range movement of molecules is challenging. Cells use molecu-lar motors such as dynein and kinesin and cytoskeletal features such as microtubules to achieve active transport over long distances relative to cargo sizes. Taking inspiration from these natural systems, Ibusuki *et al.* outfitted the motor protein dynein with a DNA-binding module that enables it to grab onto and move along an engineered DNA track (see the Perspective by Gandavadi and Hariadi). This system is attrac-tive because the DNA track can adopt precisely designed structures, and the DNA-binding module creates specificity for different sequences. Using these features, the authors cre-ated sorters that could separate cargos between two tracks and integrators that could bring together two streams of cargos. The average speed of the engineered motors is about 220 nanometers per second, comparable to some molecular motors in the cell. —MAF

Science, abj5170, this issue p. 1159; see also abn9659, p. 1089

SURFACE CHEMISTRY
Imaging polyethylene synthesis

The long-accepted mechanism for olefin polymerization growth is that the polymer elongates at one end through olefin insertion into the metal−carbon bond that anchors the chain to the active catalyst. Using in situ scanning tunneling microscopy, Guo *et al.* visualized this pro-cess for polyethylene on an iron carbide surface formed from Fe(110) (see the Perspective by Wintterlin). Triangular iron sites at the boundary between two neighboring carbide domains could oligomerize ethylene mol-ecules at room temperature and allowed for real-time imaging of the growing chain. —PDS

Science, abi4407, this issue p. 1188; see also abo2194, p. 1092

INFLAMMATION
Taking the gas out of lung inflammation

In patients with acute lung injury (ALI)/acute respiratory distress syndrome (ARDS), sepsis induces immune cell infiltration in the lungs and the production of excessive amounts of pro-inflammatory cytokines. Noting that the circulating amounts of the protein Gas6, a ligand of the receptor tyrosine kinase Mer, are increased in patients with septic ALI/ARDS, Fukatsu *et al.* examined the effects of the Mer inhibitor UNC2250 in a mouse model of ALI. Blocking Gas6-Mer signaling reduced inflammation, immune cell infil-tration, and damage in the lungs and improved survival. —JFF

Sci. Signal. **15**, eabd2533 (2022).

CANCER
Stemming AML with UBE2N inhibition

Relapse in acute myeloid leukemia (AML) is often due to the inability to fully eradi-cate leukemic hematopoietic stem and progenitor cells (HSPCs). These cells co-opt innate immune signaling pathways for tumorigenesis and represent a potential target of therapeutic interven-tion. Barreyro *et al.* found that dysregulated innate immune pathway activation depends on ubiquitin-conjugating enzyme E2 N (UBE2N). Pharmacologic targeting of UBE2N in mice car-rying patient-derived xenografts suppressed AML and inhibited the expansion of leukemic HSPCs. UBE2N represents an actionable target for AML and other malignant hematopoietic cell disorders. —DLH

Sci. Transl. Med. **14**, eabb7695 (2022).

CORONAVIRUS
Protection, whether direct or not

Vaccination provides both direct protection of vaccinated indi-viduals and indirect protection of individuals living in vacci-nated communities. Two studies based on data from Israel investigated the efficacy and indirect protection of the Pfizer/BioNTech messenger RNA vaccine (see the Perspective by Dean and Halloran). Prunas *et al.* used statistical approaches to analyze transmission in households from June 2020 to July 2021. People who were vaccinated and subsequently infected were less infectious than unvaccinated persons. Moreover, less transmission occurred within households with vaccinated members than in those with unvaccinated individuals. However, the abil-ity of the vaccine to prevent transmission waned with time and with the advent of the Delta variant. Hayek *et al.* investigated whether older and vaccinated household members reduced the risk of infection to younger children who are as yet ineligible for vaccination. Regardless of household size, parental vacci-nation substantially reduced the risk of children up to 12 years old becoming infected. This indirect effect will protect chil-dren from risk of severe disease and reduce the propagation of transmission chains. —CA

Science, abl4292, abm3087, this issue p. 1151, 1155; see also abo2959, p. 1088

CELL BIOLOGY

OpenCell: Endogenous tagging for the cartography of human cellular organization

Nathan H. Cho†, Keith C. Cheveralls†, Andreas-David Brunner†, Kibeom Kim†, André C. Michaelis†, Preethi Raghavan†, Hirofumi Kobayashi, Laura Savy, Jason Y. Li, Hera Canaj, James Y. S. Kim, Edna M. Stewart, Christian Gnann, Frank McCarthy, Joana P. Cabrera, Rachel M. Brunetti, Bryant B. Chhun, Greg Dingle, Marco Y. Hein, Bo Huang, Shalin B. Mehta, Jonathan S. Weissman, Rafael Gómez-Sjöberg, Daniel N. Itzhak, Loïc A. Royer, Matthias Mann, Manuel D. Leonetti*

INTRODUCTION: Proteins are the product of gene expression and the molecular building blocks of cells. Examples include enzymes that orchestrate the cell's chemistry, filaments that shape the cell's structure, or the pharmacological targets of drugs. The genome sequence provides us with the complete set of proteins that give rise to the human cell. However, systematically characterizing how proteins organize within the cell to sustain its operation remains an important goal of modern cell biology. A comprehensive map of the human proteome's organization will serve as a reference to explore gene function in health and disease.

RATIONALE: Subcellular localization and physical interactions are key aspects tightly related to the function of any given protein. Proteins localize to different subcellular compartments, which enables a spatial separation of cellular functions. Proteins also physically interact with one another, forming molecular networks that connect proteins involved in the same processes. There-

fore, mapping the cell's molecular organization requires a comprehensive description of where different proteins localize and how they interact. Among other strategies, a powerful approach to map cellular architecture is to visualize individual proteins using fusions with fluorescent protein "tags." These tags allow us not only to image protein localization in live cells, but also to measure protein interactions by serving as handles for immunopurification–mass spectrometry (IP-MS). Recent advances in genome engineering facilitate tagging of endogenous human genes, so that the corresponding proteins can be characterized in their native cellular environment.

RESULTS: Using high-throughput CRISPR-mediated genome editing, we constructed a library of 1310 fluorescently tagged cell lines. By performing paired IP-MS and live-cell imaging using this library, we generated a large dataset that maps the cellular localization and physical interactions of the corresponding 1310 proteins. Applying a combination of unsupervised clustering and

machine learning for image analysis allowed us to objectively identify proteins that share spatial or interaction signatures. Our data provide insights into the function of individual proteins, but also enable us to derive some general principles of human cellular organization. In particular, we show that proteins that bind RNA form a separate subgroup defined by specific localization and interaction signatures. We also show that the precise spatial distribution of a given protein is very strongly correlated with its cellular function, such that fine-grained molecular insights can be derived from the analysis of imaging data. Our opensource dataset can be explored through an interactive web interface at opencell.czbiohub.org.

CONCLUSION: Our results show that endogenous tagging coupled with interactome and microscopy analysis provides new systems-level insights about the organization of the human proteome. The information contained within the subcellular distribution of each protein is highly specific and can be paired with advances in machine learning to extrapolate fine-grained functional information using microscopy alone. This opens exciting avenues for the characterization of understudied proteins, high-throughput screening, and modeling of complex cellular states during differentiation and disease. ∎

S READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abi6983
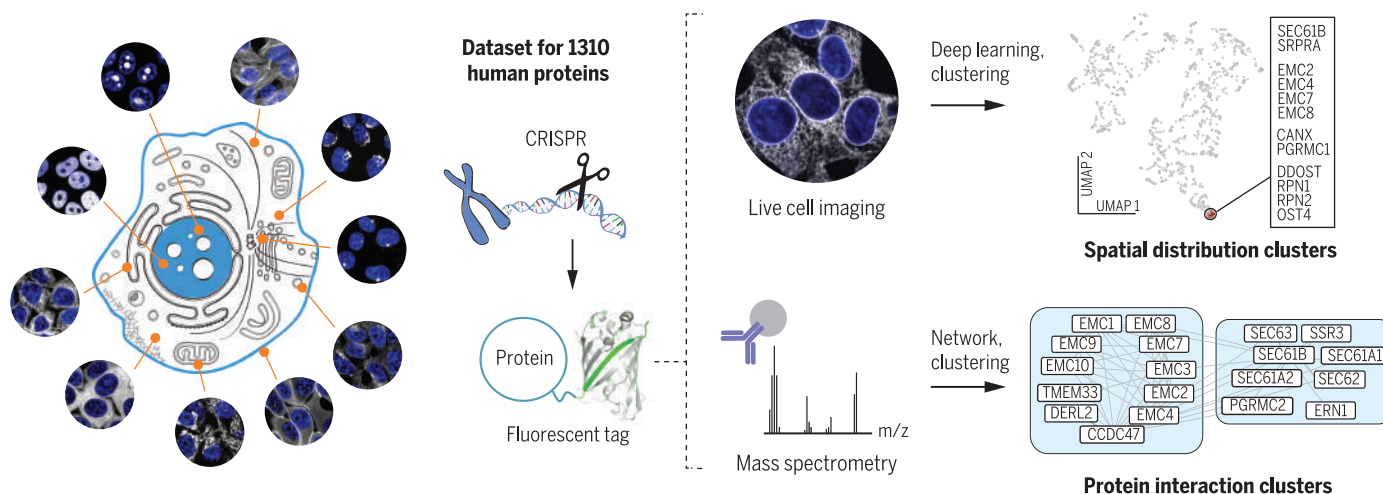
OpenCell: Combining endogenous tagging, live-cell imaging, and interaction proteomics to map the architecture of the human proteome. We created a library of engineered cell lines by using CRISPR to introduce fluorescent tags into 1310 individual human proteins. This allowed us to image the localization of each protein in live cells, as well as the interactions between a given target and other proteins within the cell. This large dataset enables a systems-level description of the organization of the human proteome.

CELL BIOLOGY

OpenCell: Endogenous tagging for the cartography of human cellular organization

Nathan H. Cho[1]†, Keith C. Cheveralls[1]†, Andreas-David Brunner[2]†, Kibeom Kim[1]†,
André C. Michaelis[2]†, Preethi Raghavan[1]†, Hirofumi Kobayashi[1], Laura Savy[1], Jason Y. Li[1],
Hera Canaj[1], James Y. S. Kim[1], Edna M. Stewart[1], Christian Gnann[1,3], Frank McCarthy[1],
Joana P. Cabrera[1], Rachel M. Brunetti[4], Bryant B. Chhun[1], Greg Dingle[5], Marco Y. Hein[1], Bo Huang[1,4,6],
Shalin B. Mehta[1], Jonathan S. Weissman[7,8], Rafael Gómez-Sjöberg[1], Daniel N. Itzhak[1], Loïc A. Royer[1],
Matthias Mann[2,9], Manuel D. Leonetti[1]*

Elucidating the wiring diagram of the human cell is a central goal of the postgenomic era. We combined genome engineering, confocal live-cell imaging, mass spectrometry, and data science to systematically map the localization and interactions of human proteins. Our approach provides a data-driven description of the molecular and spatial networks that organize the proteome. Unsupervised clustering of these networks delineates functional communities that facilitate biological discovery. We found that remarkably precise functional information can be derived from protein localization patterns, which often contain enough information to identify molecular interactions, and that RNA binding proteins form a specific subgroup defined by unique interaction and localization properties. Paired with a fully interactive website (opencell.czbiohub.org), our work constitutes a resource for the quantitative cartography of human cellular organization.

Sequencing the human genome has transformed cell biology by defining the protein parts list that forms the canvas of cellular operation (*1, 2*). This paves the way for elucidating how the ~20,000 proteins encoded in the genome organize in space and time to define the cell's functional architecture (*3, 4*). Where does each protein localize within the cell? Can we comprehensively map how proteins assemble into larger functional communities? A main challenge to answering these fundamental questions is that cellular architecture is organized along multiple scales. Therefore, several approaches need to be combined for its elucidation (*5*). In a series of pioneering studies, human protein-protein interactions have been mapped using ectopic expression strategies with yeast two-hybrid (*6*) or epitope tagging coupled to immunoprecipitation–mass spectrometry (IP-MS) (*7, 8*), whereas protein

localization has been charted using immunofluorescence in fixed samples (*9*). A complementary approach is to directly modify genes in a genome by appending sequences that illuminate specific aspects of the corresponding proteins' function [commonly referred to as "endogenous tagging" (*10*)]. For example, endogenously tagging a gene with a fluorescent reporter enables imaging of protein subcellular localization in live cells and supports functional characterization in a native cellular environment (*10, 11*). The use of endogenous tagging to study the organization of a eukaryotic cell is illustrated by seminal work in the budding yeast *Saccharomyces cerevisiae*. There, libraries of tagged strains have enabled the comprehensive mapping of protein localization and molecular interactions across the yeast proteome (*12–14*). These libraries were made possible by the relative simplicity of homologous recombination and genome engineering in yeast (*15*). In human cells, earlier work has leveraged alternative strategies including expression from bacterial artificial chromosomes (*16*) or central-dogma tagging (*17*) because of the difficulty of site-specific gene editing. CRISPR-mediated genome engineering now allows for homologous recombination–based endogenous tagging to be applied for the interrogation of the human cell (*10, 11, 18*).

Here, we combined experimental and analytical strategies to create OpenCell, a proteomic map of human cellular architecture. We generated a library of 1310 CRISPR-edited human embryonic kidney (HEK) 293T cell lines harboring fluorescent tags on individual proteins,

which we characterized by pairing confocal microscopy and mass spectrometry. Our dataset constitutes the most comprehensive live-cell image collection of human protein localization to date. In addition, integration of IP-MS using the fluorescent tags for affinity capture enables measurement of localization and interactions from the same samples. For a quantitative description of cellular architecture, we developed a data-driven framework to represent protein interactions and localization features, supported by a new machine learning algorithm for image encoding.

This approach allows us to delineate communities of functionally related proteins by unsupervised clustering and facilitates the generation of mechanistic hypotheses, including for proteins that had so far remained uncharacterized. We further demonstrate that the localization pattern of each protein is defined by unique and specific features that can be used for functional interpretation, to the point that spatial relationships often contain enough information to predict interactions at the molecular scale. Finally, our analysis enables an unsupervised description of the human proteome's organization; in particular, we found that RNA binding proteins exhibit unique functional signatures that shape the proteome's network.

Engineered cell library

Fluorescent protein (FP) fusions are versatile tools that can enable the measurement of protein localization (by microscopy) as well as protein-protein interactions (by acting as affinity handles for IP-MS) (*18, 19*) (fig. S1A). Here, we constructed a library of fluorescently tagged HEK293T cell lines by targeting human genes with the split-mNeonGreen2 system (*20*) (Fig. 1A). Split FPs greatly simplify CRISPR-based genome engineering by circumventing the need for molecular cloning (*18*) and allowed us to generate endogenous genomic fusions (Fig. 1B) that preserve native expression regulation. A full description of our pipeline is available in (*21*) and is summarized in Fig. 1, C to E. In brief, FP insertion sites (N or C terminus) were chosen on the basis of information from the literature or structural analysis (fig. S1B and table S1). For each tagged target, we isolated a polyclonal pool of CRISPR-edited cells, which was then characterized by live-cell three-dimensional (3D) confocal microscopy, IP-MS, and genotyping of tagged alleles by next-generation sequencing. Open-source software development and advances in instrumentation supported scalability (Fig. 1C). In particular, we developed crispycrunch, a CRISPR design software that enables guide RNA selection and homology donor sequence design (github.com/czbiohub/crispycrunch). We also fully automated the acquisition of microscopy data in Python for

[1]Chan Zuckerberg Biohub, San Francisco, CA, USA.
[2]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [3]Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH–Royal Institute of Technology, Stockholm, Sweden. [4]Department of Biochemistry and Biophysics, University of California, San Francisco, CA, USA. [5]Chan Zuckerberg Initiative, Redwood City, CA, USA. [6]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, USA. [7]Whitehead Institute, Koch Institute, Howard Hughes Medical Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, USA. [9]NNF Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
*Corresponding author. Email: manuel.leonetti@czbiohub.org
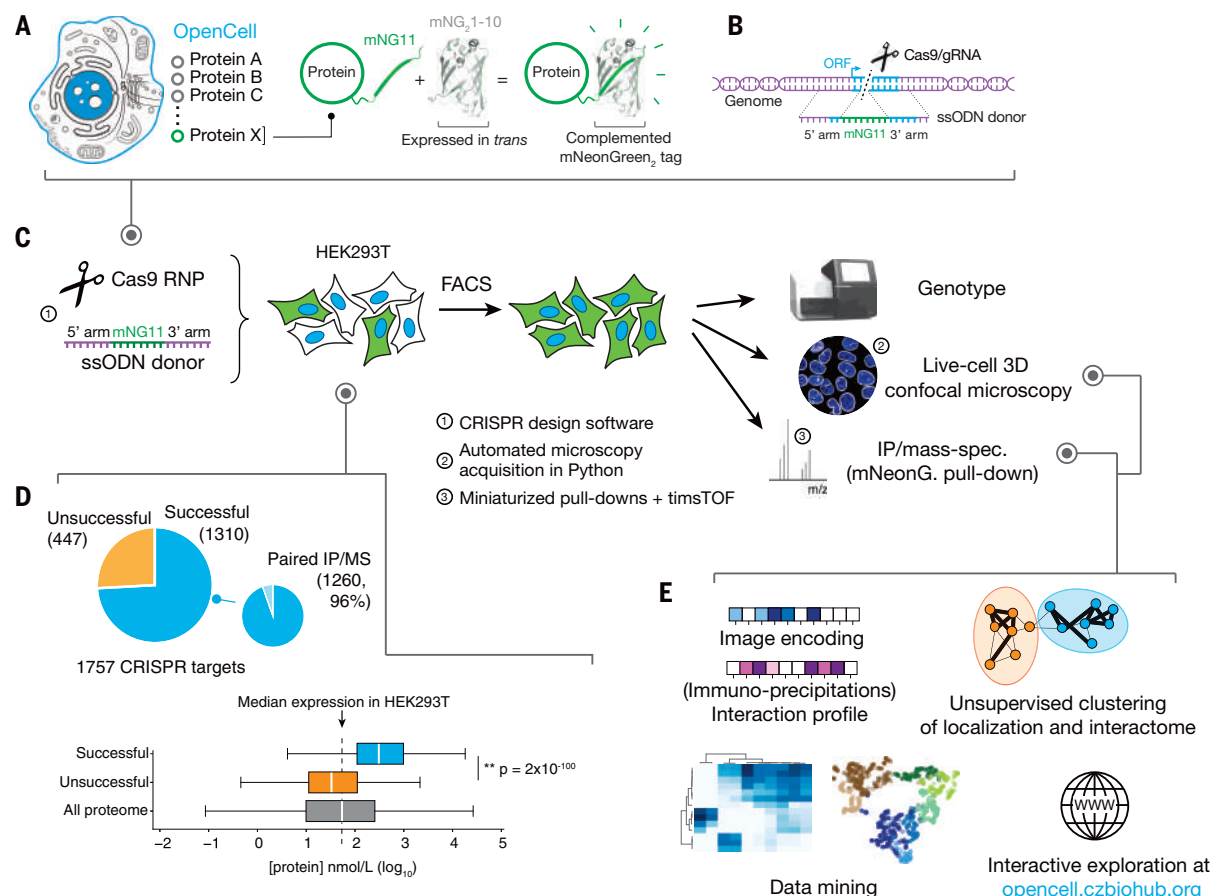†These authors contributed equally to this work.

Fig. 1. The OpenCell library. (**A**) Functional tagging with split-mNeonGreen$_2$. In this system, mNeonGreen$_2$ is separated into two fragments: a short mNG11 fragment, which is fused to a protein of interest, and a large mNG$_2$1-10 fragment, which is expressed separately in trans (that is, tagging is done in cells that have been engineered to constitutively express mNG$_2$1-10). (**B**) Endogenous tagging strategy: mNG11 fusion sequences are inserted directly within genomic open reading frames (ORFs) using CRISPR-Cas9 gene editing and homologous recombination with single-stranded oligodeoxynucleotide donors (ssODNs). (**C**) The OpenCell experimental pipeline. See text for details. (**D**) Successful detection of fluorescence in the OpenCell library. Top: Of 1757 genes that were originally targeted, fluorescent signal was successfully detected for 1310. Bottom: Low protein abundance is the main obstacle to successful detection. The graph shows the distribution of abundance for all proteins expressed in HEK293T versus successfully or unsuccessfully detected OpenCell targets; boxes represent 25th, 50th, and 75th percentiles, and whiskers represent 1.5 times the interquartile range. Median is indicated by a white line. *P* value: Student's *t* test. (**E**) The OpenCell data analysis pipeline.

on-the-fly computer vision and selection of desirable fields of view imaged in 96-well plates (github.com/czbiohub/2021-opencell-microscopy-automation). Our mass spectrometry protocols used the high sensitivity of trapped ion mobility spectrometry time-of-flight (timsTOF) instruments (*22*), which allowed miniaturization of IP-MS down to 0.8×10^6 cells of starting material [fig. S1C; about one-tenth of the material required in previous approaches (*7, 8*)].

In total, we targeted 1757 genes, of which 1310 (75%) could be detected by fluorescence imaging and form our current dataset (full library details in table S1). From these, we obtained paired IP-MS measurements for 1260 targets (96%; Fig. 1D). The 1310-protein collection includes a balanced representation of the pathways, compartments, and functions of the human proteome (fig. S1D), with the

exception of processes specific to mitochondria, organellar lumen, or extracellular matrix. Indeed, the split-FP system tags a gene of interest with a short sequence (mNG11) while a larger FP fragment (mNG$_2$1-10) is expressed separately (Fig. 1A). In the version used here, the mNG$_2$1-10 fragment is expressed in the nucleocytoplasm and prevents access to proteins inside organellar compartments. Membrane proteins can be tagged as long as one terminus extends into the nucleocytoplasm. In future iterations, other split systems that contain compartment-specific signal sequences could be used to target organellar lumen (*23*).

Fluorescent tagging was readily successful for essential genes, which suggests that FP fusions are well tolerated (fig. S2A). To evaluate other factors contributing to successful fluorescent detection, we measured RNA and protein concentration in HEK293T cells (fig.

S2B) using a 24-fraction scheme for deep proteome quantification (see fully annotated proteome in table S2). This revealed that protein abundance is the main limitation to detection (Fig. 1D and fig. S2C; see details for unsuccessful targets in table S3); most successful targets are among the top 50% most abundant (fig. S2D). Gene-editing efficiency was another important factor: Among well-expressed targets, failure was correlated with significantly lower rates of homologous recombination (fig. S2E), which would impair the selection of edited cells by fluorescence-activated cell sorting (FACS). Training a regression model revealed that the combination of protein abundance and editing efficiency could predict successful detection with 82% accuracy.

To maximize throughput, we used a polyclonal strategy to select genome-edited cells by FACS. Polyclonal pools contain cells with distinct
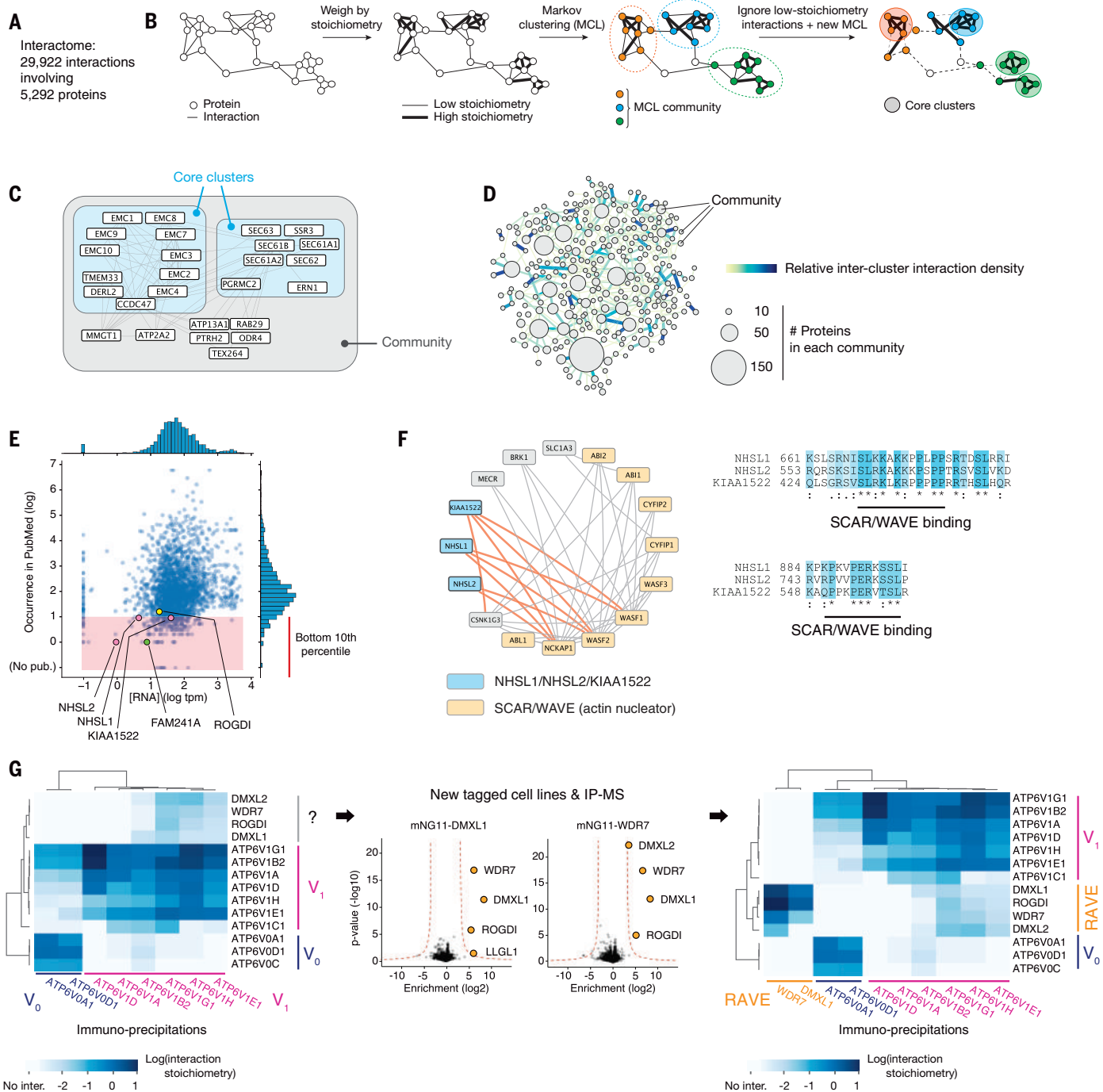
Fig. 2. Protein interactome. (**A**) Overall description of the interactome. (**B**) Unsupervised Markov clustering of the interactome graph. (**C**) Example of community and core cluster definition for the translocon/EMC community. (**D**) The complete graph of connections between interactome communities. The density of protein-protein interactions between communities is represented by increased edge width. The numbers of targets included in each community are represented by circles of increasing diameter. (**E**) Distribution of occurrence in PubMed articles versus RNA expression for all proteins found within interactome communities. The bottom 10th percentile of publication count (poorly characterized proteins) is highlighted. (**F**) NHSL1, NHSL2, and KIAA1522 are part of the SCAR/WAVE community and share amino acid sequence homology. (**G**) DMXL1/2, WDR7, and ROGDI form the human RAVE complex. Heatmaps represent the interaction stoichiometry of preys (lines) in the pull-downs of specific OpenCell targets (columns). See text for details.

genotypes. HEK293T cells are pseudo-triploid (*24*) and a single edited allele is sufficient to confer fluorescence. Moreover, various DNA repair mechanisms compete with homologous recombination for the resolution of CRISPR-induced genomic breaks (*25*) so that alleles containing nonfunctional mutations can be present in addition to the desired fusion alleles. However, such alleles do not support fluorescence and are therefore unlikely to have an impact on other measurements, especially in the context of a polyclonal pool. We developed a stringent selection scheme to significantly enrich for fluorescent fusion alleles (fig. S3A). Our final cell library has a median

61% of mNeonGreen-integrated alleles, 5% wild-type, and 26% other nonfunctional alleles (fig. S3B; full genotype information in table S1).

Finally, we verified that our engineering approach maintained the endogenous abundance of the tagged target proteins. For this, we quantified protein expression by Western blotting using antibodies specific to proteins targeted in 12 different cell pools (fig. S3C) and by single-shot mass spectrometry in 63 tagged lines (fig. S3D). Both approaches revealed a median abundance of tagged targets in engineered lines at ~80% of untagged HEK293T control, with five outliers (8% of total) identified by proteomics (fig. S3D, all within a factor of 3.5 of control). The overall proteome composition was unchanged in all tagged lines (fig. S3, E and F).

Overall, our gene-editing strategy preserves near-endogenous abundances and circumvents the limitations of ectopic overexpression (11, 26, 27), which include aberrant localization, changes in organellar morphology, and masking effects (see the examples of SPTLC1, TOMM20, and MAP1LC3B in fig. S3G). Therefore, OpenCell supports the functional profiling of tagged proteins in their native cellular context.

Interactome analysis and stoichiometry-driven clustering

Affinity enrichment coupled to mass spectrometry is an efficient and sensitive method for the systematic mapping of protein interaction networks (28). We isolated tagged proteins ("baits") from cell lysates solubilized in digitonin, a mild nonionic detergent that preserves the native structure and properties of membrane proteins (29). Specific protein interactors ("preys") were identified by proteomics from biological triplicate experiments [see fig. S4, A and B, and (21) for a detailed description of our statistical analysis, which builds upon established methods (7)]. In total, the full interactome from our 1260 OpenCell baits includes 29,922 interactions among a total of 5292 proteins (baits and preys, Fig. 2A; full interactome data in table S4).

To assess the quality of our interactome, we estimated its precision (the fraction of true positive interactions over all interactions) and recall (the fraction of interactions identified relative to a ground truth set) using reference data (fig. S4B). For recall analysis, we quantified the coverage in our data of interactions included in CORUM (30), a compendium of protein interactions manually curated from the literature. To estimate precision, we quantified how many of our interactions involved protein pairs expected to localize to the same broad cellular compartment (31) (fig. S4B). To benchmark OpenCell against other large-scale interactomes, we compared its precision and recall to Bioplex [overexpression of hemagglutinin-tagged baits (8, 32)], the yeast two-hybrid human reference interactome [HuRI (6)], and our own previous data [green fluorescent protein fusions expressed from bacterial artificial chromosomes (7)] (fig. S4, C to E). We also calculated compression rates for each dataset as a measure of the overall richness in network patterns and motifs distinguishable from noise, which correlates with overall network quality: Real-world networks contain redundant information that can be compressed, whereas pure noise is not compressible (33) (fig. S4F). Across all metrics, OpenCell outperformed previous approaches. OpenCell also includes many interactions not reported in previous datasets (fig. S4, E and G). Our interactome may better reflect biological interactions because it preserves near-endogenous protein expression.

A powerful way to interpret interactomes is to identify communities of interactors (8, 13). To this end, we applied unsupervised Markov clustering (MCL) (34) to the graph of interactions defined by our data (5292 proteins total, baits and preys). We first measured the stoichiometry of each interaction, using a quantitative approach we previously established (7). Interaction stoichiometry measures the abundance of a protein interactor relative to the abundance of the bait in a given immunoprecipitation sample. We have shown that stoichiometry can be interpreted as a proxy for interaction strength and that interactions can be classified between core (i.e., high) and low stoichiometries (7). In our current data, both high- and low-stoichiometry interactions were significantly enriched for protein pairs sharing Gene Ontology (GO) annotations (fig. S4H). Using stoichiometry to assign weights to the edges in the interaction graph (Fig. 2B), a first round of MCL delineated interconnected protein communities and led to better clustering performance than clustering based on connectivity alone (fig. S4I). To better delineate stable complexes, we further refined each individual MCL community by additional clustering while removing low-stoichiometry interactions. The resulting subclusters outline core interactions within existing communities (Fig. 2B). Figure 2C illustrates how this unsupervised approach enables delineation of functionally related proteins: All subunits of the machinery responsible for the translocation of newly translated proteins at the endoplasmic reticulum (ER) membrane (SEC61/62/63) and of the ER membrane complex (EMC) are grouped within respective core interaction clusters, but both are part of the same larger MCL community. This mirrors the recently appreciated cotranslational role of EMC for insertion of transmembrane domains at the ER (35). Additional proteins that have only recently been shown to act cotranslationally are found clustering with translocon or EMC subunits, including ERN1 (IRE1) (36) and CCDC47 (37, 38). Thus, clustering can facilitate

mechanistic exploration by grouping proteins involved in related pathways. Overall, we identified 300 communities including a total of 2096 baits and preys (full details in table S4). Ontology analysis revealed that these communities are significantly enriched for specific cellular functions, supporting their biological relevance (82% of all communities are significantly enriched for specific biological process or molecular function GO terms; see table S5 for complete analysis). A graph of interactions between communities reveals a richly interconnected network (Fig. 2D), the structure of which outlines the global architecture of the human interactome (discussed further below).

Interactome clustering can be directly applied to help elucidate the cellular roles of the many human proteins that remain poorly characterized (39). We identified poorly characterized proteins by quantifying their occurrence in article titles and abstracts from PubMed (Fig. 2E). Empirically, we determined that proteins in the bottom 10th percentile of publication count (corresponding to fewer than 10 publications) are very poorly annotated (Fig. 2E). This set encompasses a total of 251 proteins found in interaction communities for which our dataset offers potential mechanistic insights. For example, the proteins NHSL1, NHSL2, and KIAA1522 are all found as part of a community centered around SCAR/WAVE, a large multisubunit complex nucleating actin polymerization (Fig. 2F). All three proteins share sequence homology and are homologous to NHS (fig. S5A), a protein mutated in patients with Nance-Horan syndrome. NHS interacts with SCAR/WAVE components to coordinate actin remodeling (40). Thus, NHSL1, NHSL2, and KIAA1522 might also act to regulate actin assembly. A recent mechanistic study supports this hypothesis: NHSL1 localizes at the cell's leading edge and directly binds SCAR/WAVE to negatively regulate its activity, reducing F-actin content in lamellipodia and inhibiting cell migration (41). The authors identified NHSL1's SCAR/WAVE binding sites, and we found these sequences to be conserved in NSHL2 and KIA1522 (Fig. 2F). Therefore, our data suggest that both NHSL2 and KIAA1522 are also direct SCAR/WAVE binders and possible modulators of the actin cytoskeleton.

Our data also shed light on the function of ROGDI, whose variants cause Kohlschuetter-Toenz syndrome [a recessive developmental disease characterized by epilepsy and psychomotor regression (42)]. ROGDI appears in the literature because of its association with disease, but no study, to our knowledge, has specifically determined its molecular function. We first observed that ROGDI's interaction pattern closely matched that of three other proteins in our dataset: DMXL1, DMXL2, and WDR7 (Fig. 2G). This set exhibited a specific
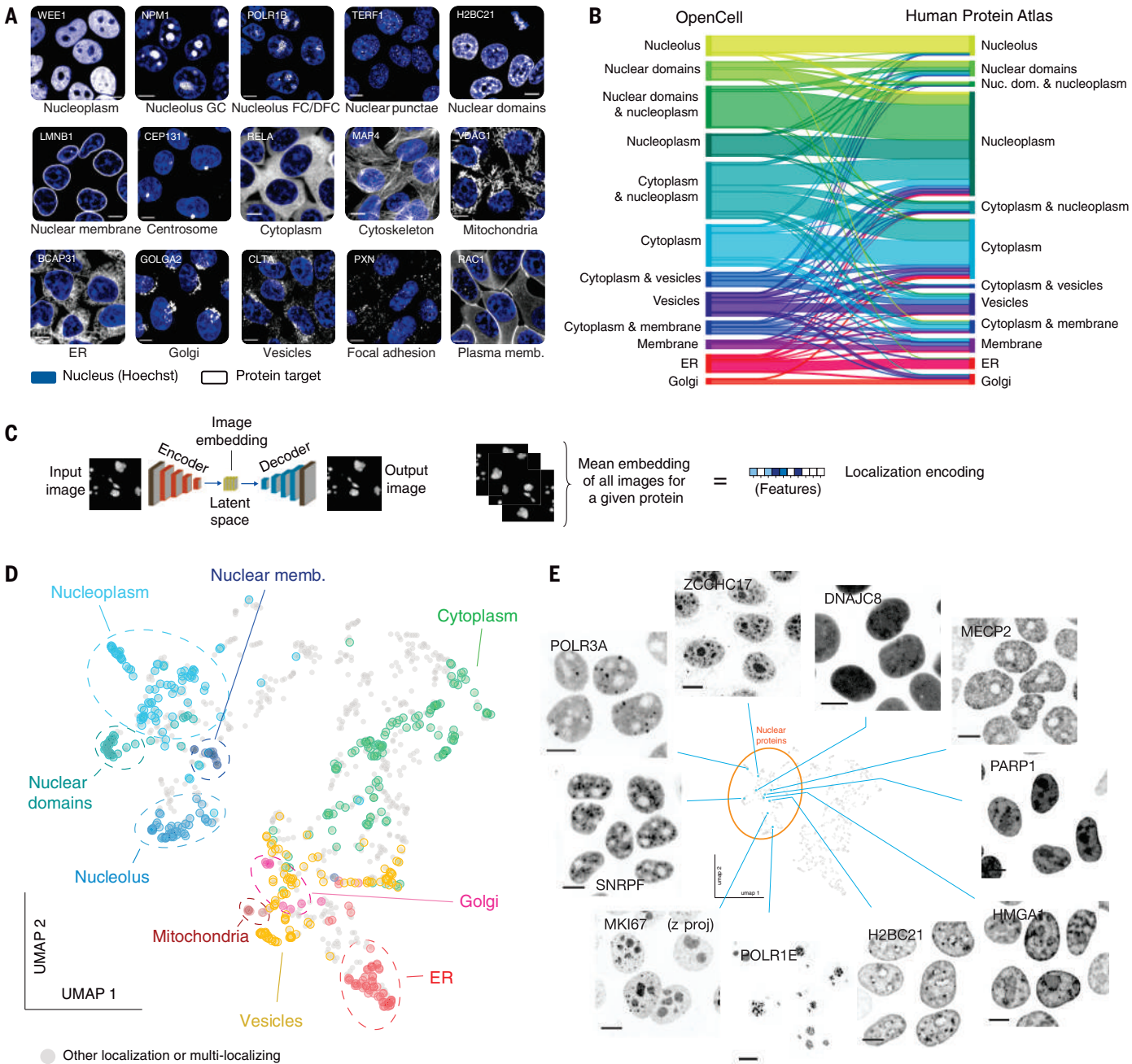
Fig. 3. Live-cell image collection. (**A**) The 15 cellular compartments segregated for annotation of localization patterns. The localization of a representative protein belonging to each group is shown (grayscale, gene names in top left corners; scale bar, 10 µm). Nuclear stain (Hoechst) is shown in blue. "Nuclear domains" designate proteins with pronounced nonuniform nucleoplasmic localization, such as chromatin-binding proteins. (**B**) Comparison of annotated localizations for proteins included in both OpenCell and Human Protein Atlas datasets. In this flow diagram, colored bands represent groups of proteins that share the same localization annotation in OpenCell, and the width of the band represents the number of proteins in each group. For readability, only the 12 most common localization groups are shown. Some multilocalization groups are included (e.g., "cytoplasm & nucleoplasm"). (**C**) Principle of localization encoding by self-supervised machine learning. See text for details. (**D**) UMAP representation of the OpenCell localization dataset, highlighting targets found to localize to a unique cellular compartment. (**E**) Representative images for 10 nuclear targets that exemplify the nuanced diversity of localization patterns across the proteome. Scale bars, 10 µm.

interaction signature with the v-ATPase lysosomal proton pump. All four proteins interact with soluble v-ATPase subunits (ATP6-V1), but not its intramembrane machinery (ATP6-V0). DMXL1 and WDR7 interact with V1 v-ATPase, and their

depletion in cells compromises lysosomal re-acidification (43). Sequence analysis showed that DMXL1 or 2, WDR7, and ROGDI are homologous to proteins from yeast and *Drosophila* involved in the regulation of assembly of the

soluble V1 subunits onto the V0 transmembrane ATPase core (44, 45) (fig. S5B). In yeast, Rav1 and Rav2 (homologous to DMXL1/2 and ROGDI, respectively) form the stoichiometric RAVE complex, a soluble chaperone that regulates

v-ATPase assembly (*45*). To assess the existence of a human RAVE-like complex, we generated new tagged cell lines for DMXL1 and 2, WDR7, and ROGDI. Because of the low abundance of these proteins, the localization of DMXL2 and ROGDI was not detectable, but pull-downs of DMXL1 and WDR7 confirmed a stoichiometric interaction between DMXL1 and 2, WDR7, and ROGDI (Fig. 2G, right panels). No direct interaction between DXML1 and DMXL2 was detected, suggesting that they might nucleate two separate subcomplexes. Therefore, our data reveal a human RAVE-like complex comprising DMXL1 or 2, WDR7, and ROGDI, which we propose acts as a chaperone for v-ATPase assembly based on its yeast homolog. Altogether, these results illustrate how our data can facilitate the generation of new mechanistic hypotheses by combining quantitative analysis and literature curation.

Image dataset: Localization annotation and self-supervised machine learning

A key advantage of our cell engineering approach is to enable the characterization of each tagged protein in live, unperturbed cells. To profile localization, we performed spinning-disk confocal fluorescence microscopy (63×, 1.47 NA objective) under environmental control (37°C, 5% CO_2) and imaged the 3D distribution of proteins in consecutive z-slices. Microscopy acquisition was fully automated in Python to enable scalability (fig. S6, A and B). In particular, we trained a computer vision model to identify fields of view (FOVs) with homogeneous cell density on-the-fly, which reduced experimental variation between images. Our dataset contains a collection of 6375 3D stacks (five different FOVs for each target) and includes paired imaging of nuclei with live-cell Hoechst 33342 staining.

We manually annotated localization patterns by assigning each protein to one or more of 15 separate cellular compartments such as the nucleolus, centrosome, or Golgi apparatus (Fig. 3A). Because proteins often populate multiple compartments at steady state (*9*), we graded annotations using a three-tier system: Grade 3 identifies prominent localization compartment(s), grade 2 represents less pronounced localizations, and grade 1 annotates weak localization patterns nearing our limit of detection (see fig. S7A for two representative examples; full annotations in table S6). Ignoring grade 1 annotations, which are inherently less precise, 55% of proteins in our library were detected in multiple locations consistent with known functional relationships. For example, clear connections were observed between secretory compartments (ER, Golgi, vesicles, plasma membrane), or between cytoskeleton and plasma membrane (fig. S7B and table S6). Many proteins were found in both nucleus and cytoplasm (21% of

our library), highlighting the importance of the nucleocytoplasmic import and export machinery in shaping global cellular function (*46, 47*). Because our split-FP system does not enable the detection of proteins in the lumen of organelles, multilocalization involving translocation across an organellar membrane (which is rare but does happen for mitochondrial or peroxisomal proteins) cannot be detected in our data.

To benchmark our dataset, we compared our localization annotations against the Human Protein Atlas (HPA), the reference antibody-based compendium of human protein localization (*9*). This revealed significant agreement between datasets: 75% of proteins shared at least one localization annotation in common (Fig. 3B; this includes 25% of all proteins that shared the exact same set of annotations, see full description in table S7A). Because HPA mostly reports on cell lines other than HEK293T, a perfect overlap was not expected, as proteins might differentially localize between related compartments in different cell types. However, the annotations for 147 proteins (11% of our data) were fully inconsistent between the two datasets (fig. S7C). An extensive curation of the literature on the localization of those proteins allowed us to resolve discrepancies for 115 proteins (i.e., 78% of that set; full curation in table S8). Of these, existing literature evidence supported the OpenCell results for 113 (98.3%) of the 115 cases (fig. S7D). This confirms the usefulness of endogenous tagging as an aid to refining the curation of localization in the human proteome. Finally, our dataset included 350 targets that have orthologs in *S. cerevisiae*. Comparison between OpenCell and yeast localization annotations (*48*) revealed a high degree of concordance (fig. S7E and table S7B; 81% of proteins share at least one annotation in common, including 36% perfect matches).

Although expert annotation remains the best-performing strategy to curate protein localization (*49, 50*), the low-dimensional description it allows is not well suited for quantitative comparisons. Recent developments in image analysis and machine learning offer new opportunities to extract high-dimensional features from microscopy images (*50, 51*). Therefore, we developed a deep learning model to quantitatively represent the localization pattern of each protein in our dataset (*52*). Briefly, our model is a variant of an autoencoder (Fig. 3C): a form of neural network that learns to vectorize an image through paired tasks of encoding (from an input image to a vector in a latent space) and decoding (from the latent space vector to a new output image). After training, a consensus representation for a given protein can be obtained from the average of the encodings from all its associated images. This generates a high-dimensional "localization encoding" (Fig. 3C) that captures the complex

set of features that define the spatial distribution of a protein at steady state and across many individual cells. One of the main advantages of this approach is that it is self-supervised. Therefore, as opposed to supervised machine learning strategies that are trained to recognize pre-annotated patterns [for example, manual annotations of protein localization (*50*)], our method extracts localization signatures from raw images without any a priori assumptions or manually assigned labels. To visualize the relationships between these high-dimensional encodings, we embedded the encodings for all 1310 OpenCell targets in two dimensions using UMAP, an algorithm that reduces high-dimensional datasets to two dimensions (UMAP 1 and UMAP 2) while attempting to preserve the global and local structures of the original data (*53*). The resulting map is organized in distinct territories that closely match manual annotations (Fig. 3D, highlighting monolocalizing proteins). This shows that the encoding approach yields a quantitative representation of the biologically relevant information in our microscopy data. The separation of different protein clusters in the UMAP embedding (discussed further below) mirrors the fascinating diversity of localization patterns across the full proteome. Images from nuclear proteins offer compelling illustrative examples of this diversity and reveal how fine-scale details can define the localization of proteins within the same organelle (Fig. 3E).

Functional specificity of protein localization in the human cell

Extracting functional insights directly from cellular images is a major goal of modern cell biology and data science (*54*). In this context, our image library and associated machine learning encodings enable us to explore what degree of functional relationship can be inferred between proteins solely based on their localization. For this, we first used an unsupervised Leiden clustering strategy commonly used to identify cell types in single-cell RNA sequencing datasets (*55*). Clusters group proteins that share similar localization properties (every protein in the dataset is included in a cluster); these groups can then be analyzed for how well they match different sets of ground-truth annotations (Fig. 4A). The average size of clusters is controlled by varying a hyperparameter called resolution (fig. S8A). Systematically varying clustering resolution in our dataset revealed that not only did low-resolution clusters delineate proteins belonging to the same organelles (Fig. 4, A and B), clustering at higher resolution also enabled us to delineate functional pathways and even molecular complexes of interacting proteins (Fig. 4, A to C). This demonstrates that the spatial distribution of each protein in the cell is highly specific, to the point that proteins sharing closely related
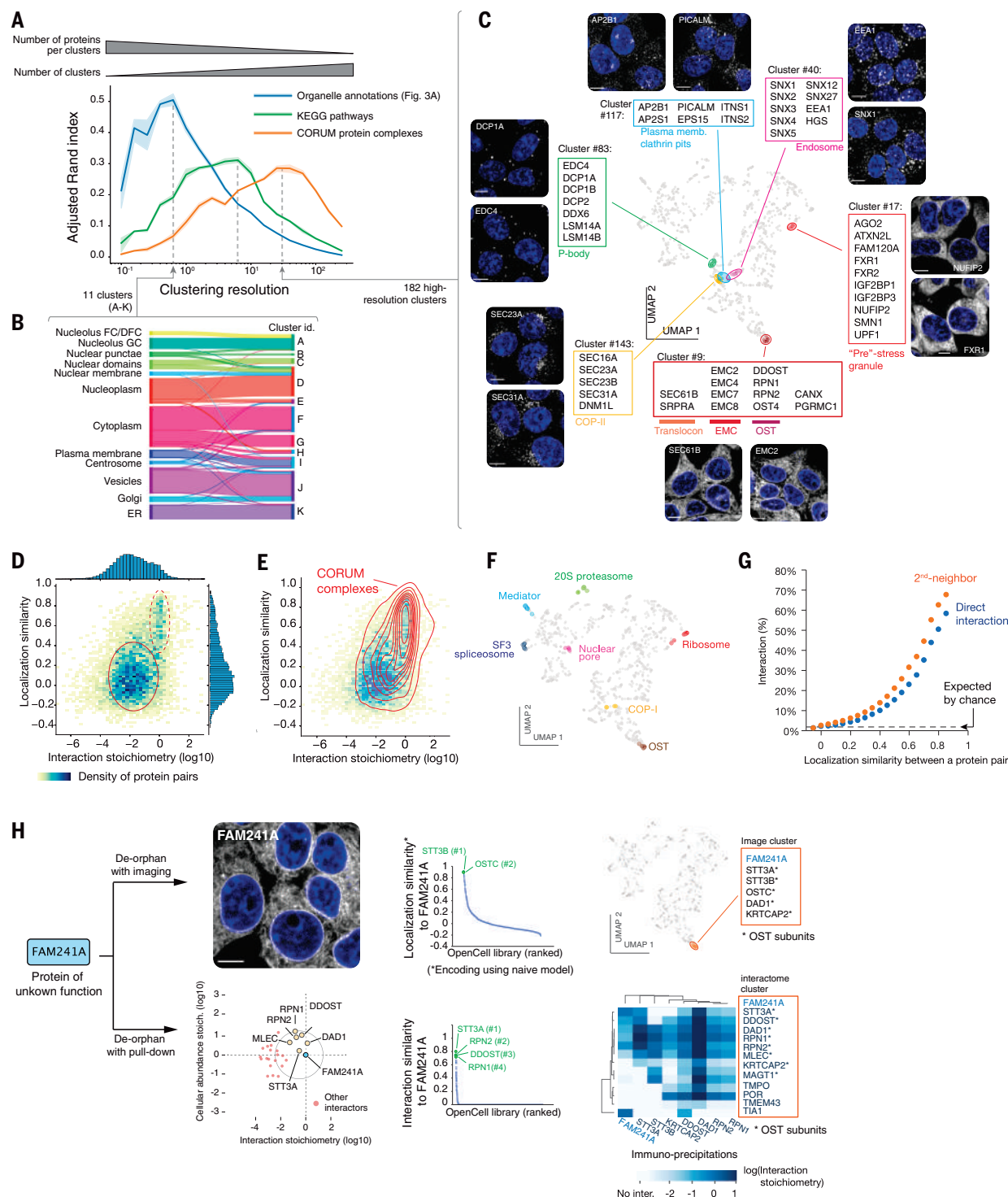
Fig. 4. Protein functional features derived from unsupervised image analysis. (**A**) Comparison of image-based Leiden clusters with ground-truth annotations. The adjusted Rand index [ARI (88)] of clusters relative to three ground-truth datasets is plotted as a function of the Leiden clustering resolution. ARI [a metric between 0 and 1 (21)] measures how well the groups from a given partition (in our case, the groups of proteins delineated at different clustering resolutions) match groups defined in a reference set. The amplitude of the ARI curves is approximately equal to the number of pairs of elements that partition similarly between sets; the resolution at which each curve reaches its maximum corresponds to the resolution that best captures the information in

each ground-truth dataset. At a low resolution, Leiden clustering delineates groups that recapitulate about half of the organellar localization annotations; at increasing resolutions, clustering recapitulates about one-third of pathways annotated in KEGG, or molecular protein complexes annotated in CORUM. Shaded regions show standard deviations calculated from nine separate repeat rounds of clustering, and average values are shown as solid lines. (**B**) High correspondence between low-resolution image clusters and cellular organelles. (**C**) Examples of functional groups delineated by high-resolution image clusters, highlighted on the localization UMAP. (**D**) Heatmap distribution of localization similarity (defined as the Pearson correlation between two deep learning–derived

encoding vectors) versus interaction stoichiometry between all interacting pairs of OpenCell targets. Two discrete subgroups are outlined: low stoichiometry/low localization similarity pairs (solid oval) and high stoichiometry/high localization similarity pairs (dashed oval). In this representation, the distributions of values across *x* and *y* axes have been binned, and the density of protein pairs within each bin is color-coded. (**E**) Probability density distribution of CORUM interactions mapped on the graph from (D). Contours correspond to isoproportions of density thresholds for each 10th percentile. (**F**) Localization patterns of different subunits from example stable protein complexes, represented on the localization UMAP. (**G**) Frequency of direct (first neighbor) or once-removed (second neighbor, having a direct interactor in common) protein-protein interactions between any two pairs of OpenCell targets sharing localization similarities above a given threshold (*x* axis). (**H**) Parallel identification of FAM241A as a new OST subunit by imaging or mass spectrometry. See text for details.

functions can be identified on the sole basis of the similarity between their spatial distributions. This is further illustrated by how finely high-resolution clusters encapsulate proteins specialized in defined cellular functions (Fig. 4C). For example, our analysis not only separated P-body proteins (cluster #83) from other forms of punctate cytoplasmic structures, but also unambiguously differentiated vesicular trafficking pathways despite their very similar localization patterns: The endosomal machinery (#40), plasma membrane endocytic pits (#117), and COP-II vesicles (#143) were all delineated with high precision (Fig. 4C). Among ER proteins, the translocon formed clusters with the SRP receptor, EMC subunits, and the OST glycosylation complex, all responsible for cotranslational operations (#9). This performance extended to cytoplasmic (fig. S8A) and nuclear clusters (fig. S8B), revealing that spatial patterning is not limited to membrane-bound organelles and that subcompartments also exist in the nucleocytoplasm. An illustrative example is a cytoplasmic cluster (#17) formed by a group of RNA binding proteins (including ATXN2L, NUFIP2, and FXR1; Fig. 4C) that separate into granules upon stress conditions (*56–59*). Stress granules are not formed under the standard growth conditions used in our experiments, but the ability of our analysis to cluster these proteins together reveals an underlying specificity to their cytoplasmic localization (i.e., "texture") even in the absence of stress.

A direct comparison between imaging and interactome data allowed us to further examine the extent to which molecular-level relationships (that is, protein interactions) can be derived from a comparison of localization patterns. For OpenCell targets that directly interact, we compared the correlation between their localization encodings derived from machine learning (defining a "localization similarity") and the stoichiometry of their interaction. This "localization similarity" measures the similarity between the global steady-state distributions of two proteins, as opposed to a direct measure of colocalization. We found that most proteins interact with low stoichiometry [as we previously described (*7*)] and without strong similarities in their spatial distribution (Fig. 4D, solid oval). This means that although low-stoichiometry interactors colocalize at least partially to interact,

their global distribution within the cell is different at steady state. On the other hand, high-stoichiometry interactors share very similar localization signatures (Fig. 4D, dashed oval). Indeed, proteins interacting within stable complexes annotated in CORUM fall into this category (Fig. 4E), and the localization signatures of different subunits from large complexes are positioned very closely in UMAP embedding (Fig. 4F). In an important correlate, we found that a high similarity of spatial distribution is a strong predictor of molecular interaction. Across the entire set of target pairs (predicted to interact or not), proteins that share high localization similarities are also very likely to interact (Fig. 4G). For example, target pairs with a localization similarity greater than 0.85 have a 58% chance of being direct interactors and a 68% chance of being second neighbors (i.e., sharing a direct interactor in common). This suggests that protein-protein interactions could be identified from a quantitative comparison of spatial distribution alone.

To test this, we focused on FAM241A (C4orf32), a protein of unknown function that was not part of our original library, and asked whether we could predict its interactions using imaging data alone, rather than the classical deorphaning approach that uses interaction proteomics. We thus generated a FAM241A endogenous fusion that was analyzed with live imaging and IP-MS separately. Encoding its localization pattern, using a "naïve" machine learning model that was never trained with images of this new target, revealed a very high localization similarity with two subunits of the ER oligosaccharyl transferase OST (>0.85 similarity to STT3B and OSTC), and high-resolution Leiden clustering placed FAM241A in an image cluster containing only OST subunits (Fig. 4H, top). This analysis suggested that FAM241A is a high-stoichiometry interactor of OST. IP-MS revealed that FAM241A was indeed a stoichiometric subunit of the OST complex (Fig. 4H, bottom). Although the specific function of FAM241A in protein glycosylation remains to be fully elucidated, this proof-of-concept example establishes that live-cell imaging can be used as a specific readout to predict molecular interactions.

Collectively, our analyses establish that the spatial distribution of a given protein contains highly specific information from which precise functional attributes can be extracted by

modern machine learning algorithms. In addition, we show that whereas high-stoichiometry interactors share very similar localization patterns, most proteins interact with low stoichiometry and share different localization signatures. This reinforces the importance of low-stoichiometry interactions for defining the overall structure of the cellular network, not only providing the "glue" that holds the interactome network together (*7*) but also connecting different cellular compartments.

RNA binding proteins form a unique group in both interactome and spatial networks

To gain insight into global signatures that organize the proteome, we further examined the structures of our imaging and interactome datasets. First, we reduced the dimensionality of each dataset by grouping proteins into their respective spatial clusters (as defined by the high-resolution localization-based clusters in Fig. 4, A and C) or interaction communities (as defined in Fig. 2B). We then separately clustered these spatial groups (fig. S9A) and interaction communities (fig. S9B) to formalize paired hierarchical descriptions of the human proteome organization. These hierarchies are highly structured and delineate clear groups of proteins (see comparison to hierarchies expected by chance, fig. S9C). In both hierarchies, groups isolated at an intermediate hierarchical layer outline "modules" that are enriched for specific cellular functions or compartments (fig. S9, A and B; full ontology analysis in tables S5 and S9). At a higher layer, each dataset is partitioned into three "branches," which represent core signatures that shape the proteome's architecture from a molecular or spatial perspective (fig. S9, A and B). The structure of the localization-based hierarchy (fig. S9A) recapitulates the human cell's architecture across its three key compartments (nucleus, cytoplasm, membrane-bound organelles; fig. S10, A and B), which reinforces the relevance of our unsupervised hierarchical analysis. This motivated a deeper examination of the hierarchical architecture of the interactome (fig. S9B; ontology analysis in table S5). We found that intermediate-layer modules of the interactome delineate specific cellular functions such as transcription or vesicular transport (fig. S9B), reflecting, as expected, that functional pathways are formed by groups of proteins that physically interact (*60, 61*). More

strikingly, the highest-layer structure showed that two of the three interactome branches were defined by clear functional signatures (fig. S10, C to E): Branch B is significantly enriched in proteins that reside in or interact with lipid membranes, whereas branch C is significantly enriched in RNA binding proteins (RNA-BPs) (Fig. 5B). This indicates that both membrane-related proteins and RNA-BPs interact more preferentially with each other than with other kinds of proteins in the cell.

That membrane-related proteins form a specific interaction group is perhaps not surprising, as the membrane surfaces that sequester them within the 3D cell will be partially maintained upon detergent solubilization. On the other hand, the fact that RNA-BPs also form a specific interaction group is unexpected, because our protein interactions were measured in nuclease-treated samples (21) in which most RNAs are degraded. This suggests that protein features beyond binding to RNAs themselves might drive the preferential interactions of RNA-BPs with each other. Therefore,
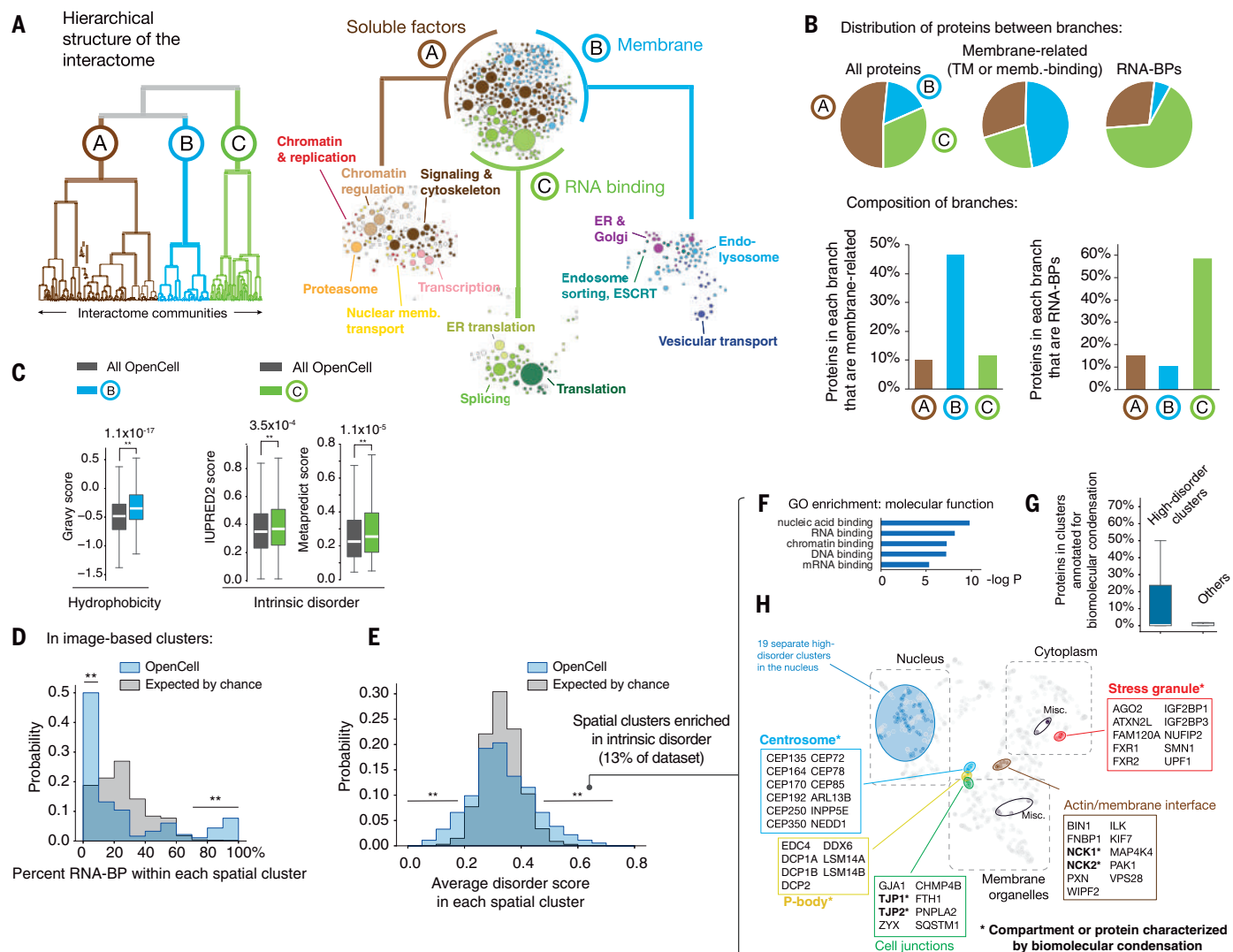
Fig. 5. Segregation of RNA-BPs in both interactome and imaging datasets.
(**A**) Hierarchical structure of the interactome dataset; see full description in fig. S9B. (**B**) Distribution of membrane-related (transmembrane or membrane-binding) proteins and RNA-BPs within the three interactome branches. (**C**) Distribution of hydrophobicity and intrinsic disorder in the membrane and RNA-BP branches of the interactome hierarchy, respectively (see full analysis in fig. S10). For intrinsic disorder, two separate scores are shown for completeness: IUPRED2 (89) and metapredict (90), a new aggregative disorder scoring algorithm. Boxes represent 25th, 50th, and 75th percentiles; whiskers represent 1.5 times the interquartile range. Median is represented by a white line. **$P < 10^{-3}$ (Student's t test); exact P values are shown. (**D**) Distribution of RNA-BP percentage across spatial clusters, comparing our data to a control in which the membership of proteins across clusters was randomized 1000 times. Lines indicate parts of the distribution overrepresented in our data versus control (**$P < 2 \times 10^{-3}$, Fisher's exact t test). (**E**) Distribution of disorder score (IUPRED2) across spatial clusters, comparing our data to a control in which the membership of proteins across clusters was randomized 1000 times. Lines indicate parts of the distribution overrepresented in our data versus control (**$P < 2 \times 10^{-3}$, Fisher's exact t test). (**F**) Ontology enrichment analysis of proteins contained in high-disorder spatial clusters (average disorder score >0.45). Enrichment compares to the whole set of OpenCell targets (P value: Fisher's exact test). (**G**) Prevalence of proteins annotated to be involved in biomolecular condensation in high-disorder versus other spatial clusters. Boxes represent 25th, 50th, and 75th percentiles; whiskers represent 1.5 times the interquartile range. Median is represented by a white line. Note that for both distributions, the median is zero. (**H**) Distribution of high-disorder spatial clusters in the UMAP embedding from Fig. 3D. Individual nuclear clusters are not outlined for readability. Multiple high-disorder spatial clusters include compartments or proteins characterized by biomolecular condensation behaviors, which are marked by an asterisk.

we reasoned that the biophysical properties of proteins within each interactome branch might underlie their segregation. Indeed, an analysis of protein sequence features revealed a separation of different biophysical properties in each branch (fig. S10, F and G). Branch B was enriched for hydrophobic sequences (Fig. 5C), consistent with its enrichment for membrane-related proteins, whereas branch C was enriched for intrinsic disorder (Fig. 5C). This is consistent with the fact that RNA-BPs are significantly more disordered than other proteins in the proteome (fig. S11A) (62). RNA-BPs are also among the most abundant in the cell (fig. S11B) and form a higher number of interactions than other proteins (fig. S11, C and D).

IP-MS measures protein interactions in vitro after lysis and therefore does not directly address the spatial relationship between interacting proteins. Thus, we sought to further examine how RNA-BPs distribute in our live-cell imaging data. If RNA-BPs segregate into interacting groups in vivo, this should also manifest at the level of their intracellular localization: They should enrich in the same spatial clusters derived from our unsupervised machine learning analysis. Indeed, the distribution of RNA-BP content within spatial clusters revealed a significant overrepresentation of clusters that were either strongly enriched or depleted for RNA-BPs (Fig. 5D). Because spatial clusters can be interpreted as defining "microcompartments" within the cell, both enrichment and depletion have functional implications: Not only are RNA-BPs enriched within the same microcompartments, they also tend to be excluded from others. Of the 26 spatial clusters (62%) that are highly enriched in RNA-BPs, 16 include at least one protein involved in biomolecular condensation [as curated in PhaSepDB (63)], which might reflect a prevalent role for biomolecular condensation in shaping the RNA-BP proteome. Collectively, both interactome and imaging data underscore that RNA-BPs (a prevalent group of proteins representing 13% of proteins expressed in HEK293T cells; see table S2) form a distinct subgroup within the proteome characterized by unique properties.

These results motivated a broader analysis of the contribution of intrinsic disorder to the spatial organization of the proteome in our dataset. Plotting the distribution of mean intrinsic disorder within spatial clusters revealed a significant overrepresentation of clusters both enriched and depleted in disordered proteins (Fig. 5E). Of 182 total spatial clusters, 26 were enriched for disordered proteins, covering 13% of the proteins in our imaging dataset. Overall, the extent to which disordered proteins segregated spatially was similar to the degree of segregation found for hydrophobic proteins: An analogous analysis revealed that 10% of proteins in our dataset are found within clus-

ters significantly enriched for high hydrophobicity (fig. S12E), which map to membrane-bound organelles (fig. S12F). This supports the hypothesis that intrinsic disorder is as important a feature as hydrophobicity in organizing the spatial distribution of the human proteome. Consistent with our previous analysis, high-disorder clusters were enriched for RNA-BPs (Fig. 5F), with 15 of these 26 clusters containing more than 50% of RNA-BPs. High-disorder clusters were also enriched for proteins annotated to participate in biomolecular condensation (Fig. 5G) and were predominantly found in the nucleus (19 clusters, 73% of total, Fig. 5H). Five of seven high-disorder clusters found in the cytosol delineate compartments for which biomolecular condensation has been proposed to play an important role (Fig. 5G), namely P-bodies (64), stress granules (59), centrosome (65), cell junctions (66), and the interface between cell surface and actin cytoskeleton (67).

Interactive data sharing at opencell.czbiohub.org

To enable widespread access to the OpenCell datasets, we built an interactive web application that provides side-by-side visualizations of the 3D confocal images and of the interaction network for each tagged protein, together with RNA and protein abundances for the whole proteome (Fig. 6). Our web interface is fully described in fig. S12.

Discussion

OpenCell combines three strategies to augment the description of human cellular architecture. First, we present an integrated experimental pipeline for high-throughput cell biology, fueled by scalable methods for genome engineering, live-cell microscopy, and IP-MS. Second, we provide an open-source resource of well-curated localization and interactome measurements, easily accessible through an interactive web interface at opencell.czbiohub.org. Third, we describe an analytical framework for the representation and comparison of interaction or localization signatures (including a self-supervised machine learning approach for image encoding). Finally, we demonstrate how our dataset can be used for fine-grained mechanistic exploration (to explore the function of multiple proteins that were previously uncharacterized) as well as for investigating the core organizational principles of the proteome.

Our current strategy that combines split FPs and HEK293T—a cell line that is heavily transformed but easily manipulable—is mostly constrained by scalability considerations. Technological advances are quickly broadening the set of cellular systems that can be engineered and profiled at scale. Advances in stem cell technologies enable the generation of libraries that can be differentiated in multiple cell types (11), while innovations in genome engineering [for example,

by modulating DNA repair (68)] pave the way for the scalable insertion of gene-sized payload, for the combination of multiple edits in the same cell, or for increased homozygosity in polyclonal pools. In addition, recent developments in high-throughput light-sheet microscopy (69) might soon enable the systematic description of 4D intracellular dynamics (70).

A central feature of our approach is the use of endogenous fluorescent tags to study protein function. Genome-edited cells enable protein function to be examined at near-native expression levels [which can circumvent some limitations of overexpression (71)] and enable the measurement of protein localization in live cells [which can avoid artifacts caused by fixation or antibody labeling (72)]. Comparing our data to the current reference datasets of protein-protein interactions (fig. S4, C to F) or localization (fig. S7, C and D) highlights the performance of our strategy. In addition, our high success rate tagging essential genes [fig. S2A; see also (73) in yeast] and the successful tagging of the near-complete yeast proteome (14, 73) indicate that fluorescent tagging generally preserves normal protein physiology.

However, limitations exist for specific protein targets. FPs are as big as an average human protein, and their insertion can impair function or localization—for example, by occluding important interaction interfaces or impairing subcellular targeting sequences. In other cases, tags can affect expression or degradation rates, which might explain why we find tagged proteins being expressed at 80% of their endogenous abundance, and 8% of targets in our dataset having outlier abundances at steady state (Fig. S3D). Further, tagging often cannot discriminate between different isoforms of a protein (such as splicing or posttranslationally modified variants). Finally, relying on endogenous expression can be an obstacle given the low concentration of most proteins in the human cell: Detection of poorly abundant proteins is difficult (especially those in the bottom half of the abundance distribution) even when using a very bright FP such as mNeonGreen (74) (fig. S2D). Solutions to this obstacle include using FP repeats to increase signal (18, 23) or using tags that bind chemical fluorophores [e.g., HaloTag (75)], which can be brighter than FPs or operate at wavelengths where cellular autofluorescence is decreased (76). Overall, the full description of human cellular architecture remains a formidable challenge that will require complementary methods being applied in parallel. The diversity of large-scale cell biology approaches is a solution to this problem (6, 8, 9, 11, 31, 70, 77–80). Mirroring the advances in genomics following the human genome sequence (2), open-source systematic datasets will likely play an important role in how the growth of cell biology measurements can be transformed
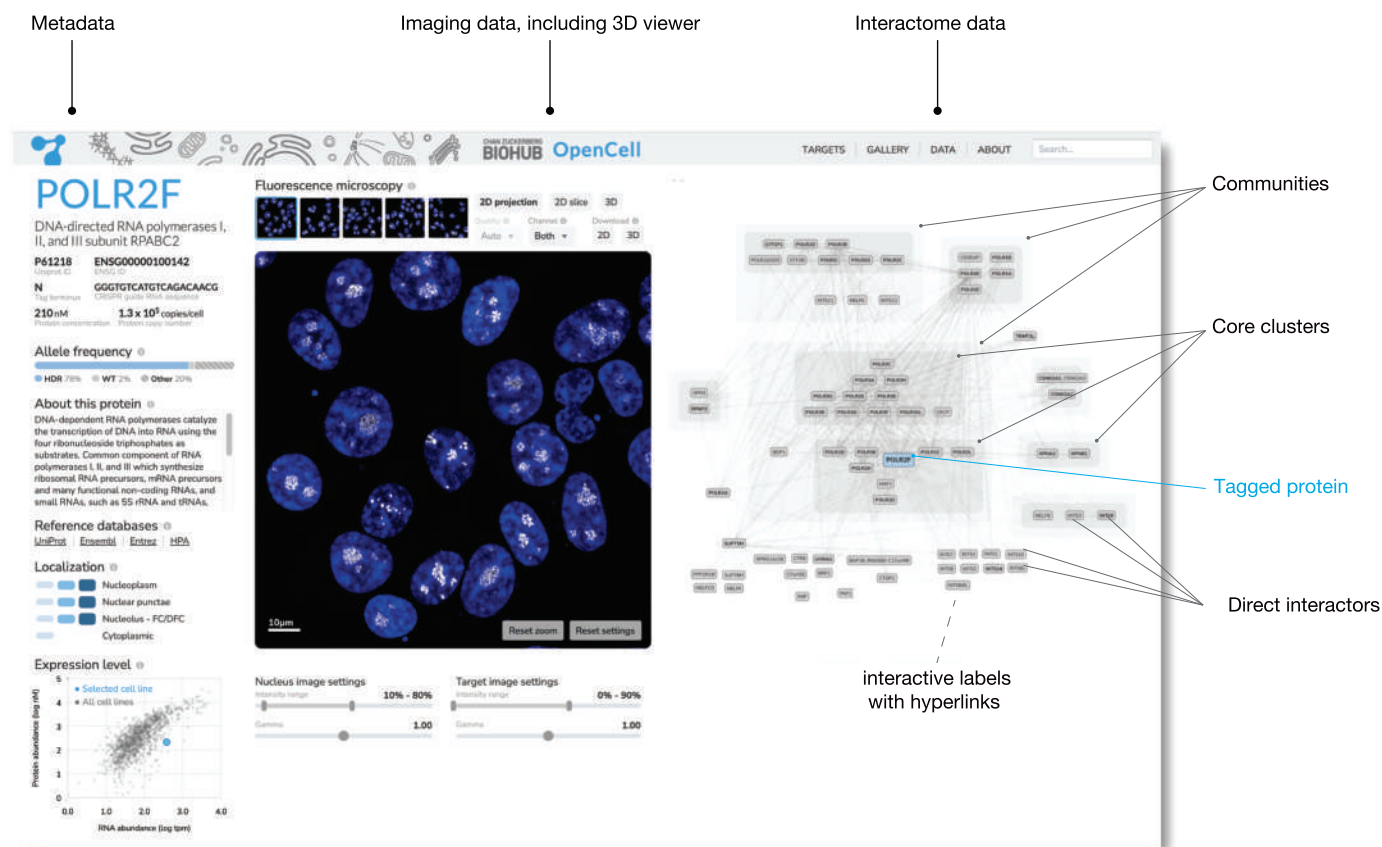
Fig. 6. The OpenCell website. Shown is an annotated screenshot from our web-app at https://opencell.czbiohub.org, which is described in more detail in fig. S12.

into fundamental discoveries by an entire community (*81*).

In addition to presenting a resource of measurements and protocols, we also demonstrate how our data can be used to study the global signatures that pattern the proteome. Our analysis reveals that RNA binding proteins, which form one of the biggest functional families in the cell, are characterized by a unique set of properties and segregate from other proteins in terms of both interactions and spatial distribution. It would be fascinating to explore the extent to which RNA itself might act as a structural organizer of the cellular proteome (*62*, *82*). This is, for example, the case for some noncoding RNAs whose main function is to template protein interactions to form nuclear bodies (*83*). High intrinsic disorder is one of the distinguishing features of RNA-BPs that likely contributes to their unique properties. Beyond RNA-BPs, our data support a general role for intrinsic disorder in shaping the spatial distribution of human proteins. For example, 13% of proteins in our dataset are found in spatial clusters that are significantly enriched for disordered proteins. This adds to the growing appreciation that intrinsic disorder, which is much more prevalent in eukaryotic than in prokaryotic proteomes (*84*, *85*), plays a key role in the functional sub-

compartmentalization of the eukaryotic nucleo- and cytoplasm in the context of biomolecular condensation (*86*).

Lastly, we show that the spatial distribution of each human protein is very specific, to the point that remarkably detailed functional relationships can be inferred on the sole basis of similarities between localization patterns, including the prediction of molecular interactions [which complements other studies (*87*)]. This highlights that intracellular organization is defined by fine-grained features that go beyond membership to a given organelle. Our demonstration that self-supervised deep learning models can identify complex but deterministic signatures from light microscopy images opens exciting avenues for the use of imaging as an information-rich method for deep phenotyping and functional genomics (*51*). Because light microscopy is easily scalable, can be performed live, and enables measurements at the single-cell level, this should offer rich opportunities for the full quantitative description of cellular diversity in normal physiology and disease.

Methods summary

See (*21*) for a complete description of methods for cell culture and CRISPR engineering, immunoprecipitation and mass spectrometry,

live-cell imaging, and data analysis of both interactome and imaging datasets. In brief, HEK-293T cells (ATCC CRL-3216) were engineered to express in-frame fluorescent gene fusions using the split-mNeonGreen2 system (see Fig. 1A). In total, we targeted 1757 human genes in separate experiments and could successfully detect fluorescence for 1310 of these gene targets, which constitute our current dataset. CRISPR-based genomic insertions were performed by nucleofection of purified Cas9 protein precomplexed with synthetic guide RNAs, delivered together with single-stranded oligodeoxynucleotide donors to template homologous recombination. Fluorescent cells were selected by flow cytometry as a polyclonal pool, which was genotyped by next-generation amplicon sequencing of the edited alleles. These selected cell pools were used for functional characterization by microscopy and IP-MS. To image protein localization in live cells, we performed 3D spinning disk confocal microscopy (63× objective, 1.47 NA) under environmental control (37°C, 5% CO_2). Microscopy acquisition was fully automated in 96-well plates using a custom acquisition script, written in Python (github.com/czbiohub/2021-opencell-microscopy-automation). To measure protein interactions, we performed IP-MS from cell lysates solubilized using digitonin detergent in the presence

of benzonase (nuclease for DNA and RNA). After immunoprecipitation of target proteins using anti-mNeonGreen nanobody resin, captured proteins were digested into peptides by LysC protease. Bottom-up mass spectrometry analysis was then performed on a timsTOF instrument. Mass spectrometry data were quantified using MaxQuant. Data analysis was performed in Python as detailed in (*21*). The code and data used to generate the figures can be found on GitHub at github.com/czbiohub/2021-opencell-figures.

REFERENCES AND NOTES

1. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004). doi: 10.1038/nature03001; pmid: 15496913

2. L. Hood, L. Rowen, The Human Genome Project: Big science transforms biology and medicine. *Genome Med.* **5**, 79 (2013). doi: 10.1186/gm483; pmid: 24040834

3. P. Nurse, J. Hayles, The cell in an era of systems biology. *Cell* **144**, 850–854 (2011). doi: 10.1016/j.cell.2011.02.045; pmid: 21414476

4. F. D. Mast, A. V. Ratushny, J. D. Aitchison, Systems cell biology. *J. Cell Biol.* **206**, 695–706 (2014). doi: 10.1083/jcb.201405027; pmid: 25225336

5. E. Lundberg, G. H. H. Borner, Spatial proteomics: A powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019). doi: 10.1038/s41580-018-0094-y; pmid: 30659282

6. K. Luck et al., A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020). doi: 10.1038/s41586-020-2188-x; pmid: 32296183

7. M. Y. Hein et al., A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015). doi: 10.1016/j.cell.2015.09.053; pmid: 26496610

8. E. L. Huttlin et al., Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017). doi: 10.1038/nature22366; pmid: 28514442

9. P. J. Thul et al., A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017). doi: 10.1126/science.aal3321; pmid: 28495876

10. H. Bukhari, T. Müller, Endogenous Fluorescence Tagging by CRISPR. *Trends Cell Biol.* **29**, 912–928 (2019). doi: 10.1016/j.tcb.2019.08.004; pmid: 31522960

11. B. Roberts et al., Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol. Biol. Cell* **28**, 2854–2874 (2017). doi: 10.1091/mbc.e17-03-0209; pmid: 28814507

12. S. Ghaemmaghami et al., Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003). doi: 10.1038/nature02046; pmid: 14562106

13. S. R. Collins et al., Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **6**, 439–450 (2007). doi: 10.1074/mcp.M600381-MCP200; pmid: 17200106

14. U. Weill et al., Genome-wide SWAp-Tag yeast libraries for proteome exploration. *Nat. Methods* **15**, 617–622 (2018). doi: 10.1038/s41592-018-0044-9; pmid: 29988094

15. A. Baudin, O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, C. Cullin, A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res.* **21**, 3329–3330 (1993). doi: 10.1093/nar/21.14.3329; pmid: 8341614

16. I. Poser et al., BAC TransgeneOmics: A high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5**, 409–415 (2008). doi: 10.1038/nmeth.1199; pmid: 18391959

17. A. Sigal et al., Generation of a fluorescently labeled endogenous protein library in living human cells. *Nat. Protoc.* **2**, 1515–1527 (2007). doi: 10.1038/nprot.2007.197; pmid: 17571059

18. M. D. Leonetti, S. Sekine, D. Kamiyama, J. S. Weissman, B. Huang, A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3501–E3508 (2016). doi: 10.1073/pnas.1606731113; pmid: 27274053

19. N. C. Hubner et al., Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189**, 739–754 (2010). doi: 10.1083/jcb.200911091; pmid: 20479470

20. S. Feng et al., Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* **8**, 370 (2017). doi: 10.1038/s41467-017-00494-8; pmid: 28851864

21. See supplementary materials.

22. F. Meier et al., Parallel Accumulation-Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **14**, 5378–5387 (2015). doi: 10.1021/acs.jproteome.5b00932; pmid: 26538118

23. D. Kamiyama et al., Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* **7**, 11046 (2016). doi: 10.1038/ncomms11046; pmid: 26988139

24. Y.-C. Lin et al., Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, 4767 (2014). doi: 10.1038/ncomms5767; pmid: 25182477

25. S. Lin, B. T. Staahl, R. K. Alla, J. A. Doudna, Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014). doi: 10.7554/eLife.04766; pmid: 25497837

26. J. B. Doyon et al., Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nat. Cell Biol.* **13**, 331–337 (2011). doi: 10.1038/ncb2175; pmid: 21297641

27. T. J. Gibson, M. Seiler, R. A. Veitia, The transience of transient overexpression. *Nat. Methods* **10**, 715–721 (2013). doi: 10.1038/nmeth.2534; pmid: 23900254

28. E. C. Keilhauer, M. Y. Hein, M. Mann, Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol. Cell. Proteomics* **14**, 120–135 (2015). doi: 10.1074/mcp.M114.041012; pmid: 25363814

29. J. A. Thomas, C. G. Tate, Quality control in eukaryotic membrane protein overproduction. *J. Mol. Biol.* **426**, 4139–4154 (2014). doi: 10.1016/j.jmb.2014.10.012; pmid: 25454020

30. M. Giurgiu et al., CORUM: The comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res.* **47**, D559–D563 (2018). doi: 10.1093/nar/gky973

31. D. N. Itzhak, S. Tyanova, J. Cox, G. H. Borner, Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5**, e16950 (2016). doi: 10.7554/eLife.16950; pmid: 27278775

32. E. L. Huttlin et al., Dual Proteome-scale Networks Reveal Cell-specific Remodeling of the Human Interactome. *Cell* **184**, 3022–3040.e28 (2021). doi: 10.1101/2020.01.19.905109

33. L. Royer, M. Reimann, A. F. Stewart, M. Schroeder, Network compression as a quality measure for protein interaction networks. *PLOS ONE* **7**, e35729 (2012). doi: 10.1371/journal.pone.0035729; pmid: 22719828

34. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002). doi: 10.1093/nar/30.7.1575; pmid: 11917018

35. M. J. Shurtleff et al., The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7**, e37018 (2018). doi: 10.7554/eLife.37018; pmid: 29809151

36. D. Acosta-Alvear et al., The unfolded protein response and endoplasmic reticulum protein targeting machineries converge on the stress sensor IRE1. *eLife* **7**, e43036 (2018). doi: 10.7554/eLife.43036; pmid: 30582518

37. P. T. McGilvray et al., An ER translocon for multi-pass membrane protein biogenesis. *eLife* **9**, e56889 (2020). doi: 10.7554/eLife.56889; pmid: 32820719

38. P. J. Chitwood, R. S. Hegde, An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature* **584**, 630–634 (2020). doi: 10.1038/s41586-020-2624-y; pmid: 32814900

39. T. Stoeger, M. Gerlach, R. I. Morimoto, L. A. Nunes Amaral, Large-scale investigation of the reasons why potentially important genes are ignored. *PLOS Biol.* **16**, e2006643 (2018). doi: 10.1371/journal.pbio.2006643; pmid: 30226837

40. S. P. Brooks et al., The Nance-Horan syndrome protein encodes a functional WAVE homology domain (WHD) and is important for co-ordinating actin remodelling and maintaining cell morphology. *Hum. Mol. Genet.* **19**, 2421–2432 (2010). doi: 10.1093/hmg/ddq125; pmid: 20332100

41. A.-L. Law et al., Nance-Horan Syndrome-like 1 protein negatively regulates Scar/WAVE-Arp2/3 activity and inhibits lamellipodia stability and cell migration. *Nat. Commun.* **12**, 5687 (2021). doi: 10.1101/2020.05.11.083030

42. A. Schossig et al., Mutations in ROGDI Cause Kohlschütter-Tönz Syndrome. *Am. J. Hum. Genet.* **90**, 701–707 (2012). doi: 10.1016/j.ajhg.2012.02.012; pmid: 22424600

43. M. Merkulova et al., Mapping the H+ (V)-ATPase interactome: Identification of proteins involved in trafficking, folding, assembly and phosphorylation. *Sci. Rep.* **5**, 14827 (2015). doi: 10.1038/srep14827; pmid: 26442671

44. Y. Yan, N. Denef, T. Schüpbach, The vacuolar proton pump, V-ATPase, is required for notch signaling and endosomal trafficking in Drosophila. *Dev. Cell* **17**, 387–402 (2009). doi: 10.1016/j.devcel.2009.07.001; pmid: 19758563

45. T. Vasanthakumar, J. L. Rubinstein, Structure and Roles of V-type ATPases. *Trends Biochem. Sci.* **45**, 295–307 (2020). doi: 10.1016/j.tibs.2019.12.007; pmid: 32001091

46. D. Görlich, U. Kutay, Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 607–660 (1999). doi: 10.1146/annurev.cellbio.15.1.607; pmid: 10611974

47. C. P. Lusk, M. C. King, The nucleus: Keeping it together by keeping it apart. *Curr. Opin. Cell Biol.* **44**, 44–50 (2017). doi: 10.1016/j.ceb.2017.02.001; pmid: 28236735

48. M. Breker, M. Gymrek, M. Schuldiner, A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J. Cell Biol.* **200**, 839–850 (2013). doi: 10.1083/jcb.201301120; pmid: 23509072

49. D. P. Sullivan et al., Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nat. Biotechnol.* **36**, 820–828 (2018). doi: 10.1038/nbt.4225; pmid: 30125267

50. W. Ouyang et al., Analysis of the Human Protein Atlas Image Classification competition. *Nat. Methods* **16**, 1254–1261 (2019). doi: 10.1038/s41592-019-0658-6; pmid: 31780840

51. S. N. Chandrasekaran, H. Ceulemans, J. D. Boyd, A. E. Carpenter, Image-based profiling for drug discovery: Due for a machine-learning upgrade? *Nat. Rev. Drug Discov.* **20**, 145–159 (2021). doi: 10.1038/s41573-020-00117-w; pmid: 33353986

52. H. Kobayashi, K. C. Cheveralls, M. D. Leonetti, L. A. Royer, Self-Supervised Deep-Learning Encodes High-Resolution Features of Protein Subcellular Localization. bioRxiv 437595 [preprint] (2021). doi: doi:10.1101/2021.03.29.437595

53. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 1802.03426 (2018).

54. E. Meijering, A. E. Carpenter, H. Peng, F. A. Hamprecht, J.-C. Olivo-Marin, Imagining the future of bioimage analysis. *Nat. Biotechnol.* **34**, 1250–1255 (2016). doi: 10.1038/nbt.3722; pmid: 27926723

55. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019). doi: 10.1038/s41598-019-41695-z; pmid: 30914743

56. S. Markmiller et al., Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell* **172**, 590–604.e13 (2018). doi: 10.1016/j.cell.2017.12.032; pmid: 29373831

57. J.-Y. Youn et al., High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* **69**, 517–532.e11 (2018). doi: 10.1016/j.molcel.2017.12.020; pmid: 29395067

58. H. Marmor-Kollet et al., Spatiotemporal Proteomic Analysis of Stress Granule Disassembly Using APEX Reveals Regulation by SUMOylation and Links to ALS Pathogenesis. *Mol. Cell* **80**, 876–891.e6 (2020). doi: 10.1016/j.molcel.2020.10.032; pmid: 33217318

59. P. Yang et al., G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325–345.e28 (2020). doi: 10.1016/j.cell.2020.03.046; pmid: 32302571

60. M. Costanzo et al., A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016). doi: 10.1126/science.aaf1420; pmid: 27708008

61. M. A. Horlbeck et al., Mapping the Genetic Landscape of Human Cells. *Cell* **174**, 953–967.e22 (2018). doi: 10.1016/j.cell.2018.06.010; pmid: 30033366

62. A. Balcerak, A. Trebinska-Stryjewska, R. Konopinski, M. Wakula, E. A. Grzybowska, RNA-protein interactions: Disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biol.* **9**, 190096 (2019). doi: 10.1098/rsob.190096; pmid: 31213136

63. K. You et al., PhaSepDB: A database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.* **48** (D1), D354–D359 (2020). doi: 10.1093/nar/gkz847; pmid: 31584089

64. Y. Luo, Z. Na, S. A. Slavoff, P-Bodies: Composition, Properties, and Functions. *Biochemistry* **57**, 2424–2431 (2018). doi: 10.1021/acs.biochem.7b01162; pmid: 29381060

65. J. B. Woodruff *et al.*, The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. *Cell* **169**, 1066–1077.e10 (2017). doi: 10.1016/j.cell.2017.05.028; pmid: 28575670

66. O. Beutel, R. Maraspini, K. Pombo-García, C. Martin-Lemaitre, A. Honigmann, Phase Separation of Zonula Occludens Proteins Drives Formation of Tight Junctions. *Cell* **179**, 923–936.e11 (2019). doi: 10.1016/j.cell.2019.10.011; pmid: 31675499

67. S. Banjade *et al.*, Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6426–E6435 (2015). doi: 10.1073/pnas.1508778112; pmid: 26553976

68. S. Riesenberg *et al.*, Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Res.* **47**, e116 (2019). doi: 10.1073/pnas.1508778112; pmid: 26553976

69. B. Yang *et al.*, Epi-illumination SPIM for volumetric imaging with high spatial-temporal resolution. *Nat. Methods* **16**, 501–504 (2019). doi: 10.1038/s41592-019-0401-3; pmid: 31061492

70. Y. Cai *et al.*, Experimental and computational framework for a dynamic protein atlas of human cell division. *Nature* **561**, 411–415 (2018). doi: 10.1038/s41586-018-0518-z; pmid: 30202089

71. C. von Mering *et al.*, Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002). doi: 10.1038/nature750; pmid: 12000970

72. U. Schnell, F. Dijk, K. A. Sjollema, B. N. G. Giepmans, Immunolabeling artifacts and the need for live-cell imaging. *Nat. Methods* **9**, 152–158 (2012). doi: 10.1038/nmeth.1855; pmid: 22290187

73. W.-K. Huh *et al.*, Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003). doi: 10.1038/nature02026; pmid: 14562095

74. N. C. Shaner *et al.*, A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10**, 407–409 (2013). doi: 10.1038/nmeth.2413; pmid: 23524392

75. G. V. Los *et al.*, HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008). doi: 10.1021/cb800025k; pmid: 18533659

76. L. D. Lavis, Chemistry Is Dead. Long Live Chemistry! *Biochemistry* **56**, 5165–5170 (2017). doi: 10.1021/acs.biochem.7b00529; pmid: 28704030

77. C. D. Go *et al.*, A proximity-dependent biotinylation map of a human cell. *Nature* **595**, 120–124 (2021). doi: 10.1038/s41586-021-03592-2; pmid: 34079125

78. G. Gut, M. D. Herrmann, L. Pelkmans, Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042 (2018). doi: 10.1126/science.aar7042; pmid: 30072512

79. J. R. A. Hutchins *et al.*, Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* **328**, 593–599 (2010). doi: 10.1126/science.1181348; pmid: 20360068

80. P. C. Havugimana *et al.*, A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012). doi: 10.1016/j.cell.2012.08.011; pmid: 22939629

81. J. Ellenberg *et al.*, A call for public archives for biological image data. *Nat. Methods* **15**, 849–854 (2018). doi: 10.1038/s41592-018-0195-8; pmid: 30377375

82. M. W. Hentze, A. Castello, T. Schwarzl, T. Preiss, A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018). doi: 10.1038/nrm.2017.130; pmid: 29339797

83. T. Chujo, T. Hirose, Nuclear Bodies Built on Architectural Long Noncoding RNAs: Unifying Principles of Their Construction and Function. *Mol. Cells* **40**, 889–896 (2017). doi: 10.14348/molcells.2017.0263; pmid: 29276943

84. Z. Peng *et al.*, Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **72**, 137–151 (2015). doi: 10.1007/s00018-014-1661-9; pmid: 24939692

85. W. Basile, M. Salvatore, C. Bassot, A. Elofsson, Why do eukaryotic proteins contain more intrinsically disordered regions? *PLOS Comput. Biol.* **15**, e1007186 (2019). doi: 10.1371/journal.pcbi.1007186; pmid: 31329574

86. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017). doi: 10.1126/science.aaf4382; pmid: 28935776

87. Y. Qin *et al.*, A multi-scale map of cell structure fusing protein images and interactions. *Nature* **600**, 536–542 (2021). doi: 10.1038/s41586-021-04115-9; pmid: 34819669

88. L. Hubert, P. Arabie, Comparing partitions. *J. Classif.* **2**, 193–218 (1985). doi: 10.1007/BF01908075

89. B. Mészáros, G. Erdős, Z. Dosztányi, IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018). doi: 10.1007/BF01908075

90. R. J. Emenecker, D. Griffith, A. S. Holehouse, Metapredict: A fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **120**, 4312–4319 (2021). doi: 10.1016/j.bpj.2021.08.039; pmid: 34480923

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

NEUROSCIENCE

Local connectivity and synaptic dynamics in mouse and human neocortex

Luke Campagnola†, Stephanie C. Seeman†, Thomas Chartrand, Lisa Kim, Alex Hoggarth, Clare Gamlin, Shinya Ito, Jessica Trinh, Pasha Davoudian, Cristina Radaelli, Mean-Hwan Kim, Travis Hage, Thomas Braun, Lauren Alfiler, Julia Andrade, Phillip Bohn, Rachel Dalley, Alex Henry, Sara Kebede, Alice Mukora, David Sandman, Grace Williams, Rachael Larsen, Corinne Teeter, Tanya L. Daigle, Kyla Berry, Nadia Dotson, Rachel Enstrom, Melissa Gorham, Madie Hupp, Samuel Dingman Lee, Kiet Ngo, Philip R. Nicovich, Lydia Potekhina, Shea Ransford, Amanda Gary, Jeff Goldy, Delissa McMillen, Trangthanh Pham, Michael Tieu, La'Akea Siverts, Miranda Walker, Colin Farrell, Martin Schroedter, Cliff Slaughterbeck, Charles Cobb, Richard Ellenbogen, Ryder P. Gwinn, C. Dirk Keene, Andrew L. Ko, Jeffrey G. Ojemann, Daniel L. Silbergeld, Daniel Carey, Tamara Casper, Kirsten Crichton, Michael Clark, Nick Dee, Lauren Ellingwood, Jessica Gloe, Matthew Kroll, Josef Sulc, Herman Tung, Katherine Wadhwani, Krissy Brouner, Tom Egdorf, Michelle Maxwell, Medea McGraw, Christina Alice Pom, Augustin Ruiz, Jasmine Bomben, David Feng, Nika Hejazinia, Shu Shi, Aaron Szafer, Wayne Wakeman, John Phillips, Amy Bernard, Luke Esposito, Florence D. D'Orazi, Susan Sunkin, Kimberly Smith, Bosiljka Tasic, Anton Arkhipov, Staci Sorensen, Ed Lein, Christof Koch, Gabe Murphy, Hongkui Zeng, Tim Jarsky*

INTRODUCTION: The mammalian neocortex is believed to act as the computational substrate for our highest cognitive abilities, particularly the ability to model the world around us and predict the effects of our actions. Many aspects of cortical structure are repeated across brain regions and conserved across species, suggesting a general-purpose approach to cognition. Within this repeating structure, neurons influence the formation of synaptic connections based on their cell type-specific biases. This results in a stereotyped network architecture in which synapse properties and connectivity are strongly influenced by cell type.

Synapses between cell types transmit information in a way that is highly stochastic and depends on the prior history of activity. The dynamic properties of synapses are also strongly dependent on both the pre- and postsynaptic cell types, suggesting an important role in cortical function. This provides a major source of computational diversity that is often absent in neuroscience modeling studies as well as modern machine-learning architectures.

Neurons are broadly grouped into excitatory and inhibitory classes, each of which can be divided into more specific subclasses. Cortical inhibitory neurons, for example, are commonly divided into Pvalb, Sst, and Vip subclasses and are distributed broadly across cortical layers. In contrast, most excitatory cell subclasses occupy narrower regions across cortical layers.

RATIONALE: Understanding the connectivity among cell subclasses and the computations performed at their synapses is an essential step to understanding cortical circuit function.

This has led to experiments in different species, brain regions, ages, etc. that focus on one or a few circuit elements. These efforts offer an excellent depth of insight to isolated regions of

Variance
Low ——— High
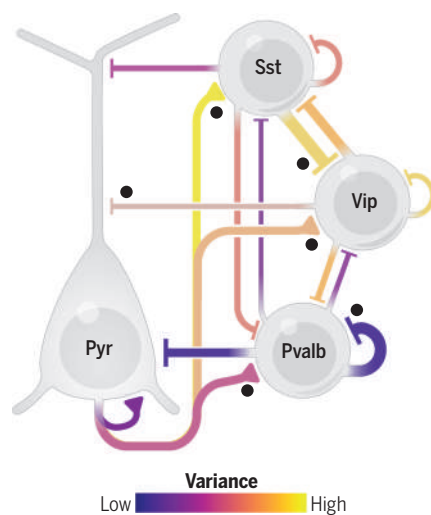
Intralaminar circuit diagram among major excitatory (Pyr) and inhibitory (Pvalb, Sst, and Vip) cell subclasses aggregated from all layers of mouse primary visual cortex. Line (axon) thickness depicts the relative weight (strength and probability of connection) of connections between subclasses. Black dots indicate connections that are stronger in layer 2/3 compared with layer 5. Axon color shows the spike-to-spike variance in amplitude of synaptic signaling, which is strongly cell subclass dependent. Excitatory synapse variance depends on the postsynaptic subclass. Pvalb cells project low-variance connections, whereas Sst and Vip project high-variance connections. More saturated axon colors indicate higher confidence measurements.

the circuit but offer a fragmented view of the circuit as a whole. Further, the difficulty of accessing these historical data discourages reuse and reanalysis. We saw an opportunity to expand upon this history and conduct a more comprehensive and standardized survey than has been attempted in the past. By publishing the analyses, tools, and data that characterize cortical connection properties, we encourage a more unified approach to describing cortical function.

RESULTS: We used microelectrodes to record the activity of 1731 synaptic connections across diverse cell types in living tissue samples from mouse and human neocortex. We characterized these connections with the aid of a synaptic release model and found that excitatory dynamics aligned with postsynaptic cell subclass, whereas inhibitory dynamics aligned with the presynaptic subclass in ways that were subclass specific. Synaptic variability was a primary driver of these cross-subclass differences in mouse cortex. Compared with the mouse, human excitatory connections were tuned toward stability and reliability pointing toward species differences in cortical function. We further introduced a method to estimate the rate of connectivity between cell types that accounts for differences between experimental preparations. With this approach, we compared connection probabilities across layer, cell subclass, and species. For instance, connectivity between excitatory cells and Vip inhibitory cells was present in layer 2/3 and absent in layer 5/6 of mouse cortex. Likewise, connection probability among layer 4 excitatory cells was high in mouse cortex and nearly absent in human cortex. Overall, we found that layer-specific circuit representations are necessary to capture the diversity of intralaminar connectivity among cortical cell subclasses.

CONCLUSION: We have generated a comprehensive dataset describing synaptic connections within each layer in the mouse and human cortex. Our deep characterization of synapses points toward important principles of cortical organization that relate to current topics in computational neuroscience and machine learning. The open distribution of our data, analyses, and tools enables greater realism in constraining network and synapse models. ∎

S **READ THE FULL ARTICLE AT**
https://doi.org/10.1126/science.abj5861

NEUROSCIENCE

Local connectivity and synaptic dynamics in mouse and human neocortex

Luke Campagnola[1]†, Stephanie C. Seeman[1]†, Thomas Chartrand[1], Lisa Kim[1], Alex Hoggarth[1], Clare Gamlin[1], Shinya Ito[1], Jessica Trinh[1], Pasha Davoudian[1], Cristina Radaelli[1], Mean-Hwan Kim[1], Travis Hage[1], Thomas Braun[2], Lauren Alfiler[1], Julia Andrade[1], Phillip Bohn[1], Rachel Dalley[1], Alex Henry[1], Sara Kebede[1], Alice Mukora[1], David Sandman[1], Grace Williams[1], Rachael Larsen[1], Corinne Teeter[1]‡, Tanya L. Daigle[1], Kyla Berry[1]§, Nadia Dotson[1], Rachel Enstrom[1], Melissa Gorham[1], Madie Hupp[1], Samuel Dingman Lee[1], Kiet Ngo[1], Philip R. Nicovich[1], Lydia Potekhina[1], Shea Ransford[1], Amanda Gary[1], Jeff Goldy[1], Delissa McMillen[1], Trangthanh Pham[1], Michael Tieu[1], La'Akea Siverts[1], Miranda Walker[1], Colin Farrell[1], Martin Schroedter[1], Cliff Slaughterbeck[1], Charles Cobb[3], Richard Ellenbogen[4], Ryder P. Gwinn[5], C. Dirk Keene[6], Andrew L. Ko[4,7], Jeffrey G. Ojemann[4,7], Daniel L. Silbergeld[4], Daniel Carey[1], Tamara Casper[1], Kirsten Crichton[1], Michael Clark[1], Nick Dee[1], Lauren Ellingwood[1], Jessica Gloe[1], Matthew Kroll[1], Josef Sulc[1], Herman Tung[1], Katherine Wadhwani[1], Krissy Brouner[1], Tom Egdorf[1], Michelle Maxwell[1], Medea McGraw[1], Christina Alice Pom[1], Augustin Ruiz[1], Jasmine Bomben[1], David Feng[1], Nika Hejazinia[1], Shu Shi[1], Aaron Szafer[1], Wayne Wakeman[1], John Phillips[1], Amy Bernard[1], Luke Esposito[1], Florence D. D'Orazi[1], Susan Sunkin[1], Kimberly Smith[1], Bosiljka Tasic[1], Anton Arkhipov[1], Staci Sorensen[1], Ed Lein[1], Christof Koch[1], Gabe Murphy[1], Hongkui Zeng[1], Tim Jarsky[1]*

We present a unique, extensive, and open synaptic physiology analysis platform and dataset. Through its application, we reveal principles that relate cell type to synaptic properties and intralaminar circuit organization in the mouse and human cortex. The dynamics of excitatory synapses align with the postsynaptic cell subclass, whereas inhibitory synapse dynamics partly align with presynaptic cell subclass but with considerable overlap. Synaptic properties are heterogeneous in most subclass-to-subclass connections. The two main axes of heterogeneity are strength and variability. Cell subclasses divide along the variability axis, whereas the strength axis accounts for substantial heterogeneity within the subclass. In the human cortex, excitatory-to-excitatory synaptic dynamics are distinct from those in the mouse cortex and vary with depth across layers 2 and 3.

The study of cortical connectivity in mammalian model systems established foundational circuit diagrams among neurons classified by their morphology and electrophysiology (1–3). Refinement of cell classes by their long-range projections and genetic markers facilitated more detailed microcircuit representations in rodents (4–8). By contrast, studies of human cortical connectivity (9–14) have been more limited but offer the opportunity to identify features that may contribute to our unique cognitive abilities. Although our knowledge of cell types and circuits in rodents and humans has advanced, a complete description of the connectivity and synaptic properties among cell subclasses in each cortical layer is still lacking (15).

Cortical synapses are dynamic, varying their response strength in ways that are highly stochastic and also modulated by the prior history of activity. The dynamical properties are influenced by both the presynaptic and postsynaptic cell types (16–20) and endow neuronal networks with essential sources of computational diversity (21, 22). As models in neuroscience push toward greater biological realism, there is a need for large-scale datasets with a deep description of synaptic dynamics among a diversity of cell types. Likewise, the incorporation of biologically inspired sources of computational diversity has shown promise in advancing machine-learning techniques (23). To meet the needs of the theoretical and experimental research communities, these large datasets need to be open and accessible.

Results

Synaptic physiology pipeline

We performed 1931 experiments (Fig. 1A) in acute brain slices targeting intralaminar connections in layer 2 (L2) to L6 of adult mouse (mean age 46.0 ± 4.6 days) primary visual cortex (VISp; 1715 experiments) and human frontotemporal cortex from neurosurgical excised tissue (216 experiments). We used transgenic mice that express unique reporters in two subclasses (24) (table S1). Six excitatory subclasses were layer or projection-class specific (Nr5a1 and Rorb, L4; Sim1 or mscRE4-FlpO AAV, L5 extratelencephalic [ET]; Tlx3, L5 intratelencephalic [IT]; Ntsr1, L6 corticothalamic [CT]), and three inhibitory subclasses (Pvalb, Sst, and Vip) were assessed in all targeted layers (Fig. 1A). We probed 23,620 potential connections (mouse: 20,949; human: 2671), of which 1731 were connected by chemical synapses, giving an overall connectivity rate of 7.3% (mouse: 1526, 7.3%; human: 205, 7.7%).

Details of the experiments are described in Fig. 1, B and C, and in the materials and methods. In each experiment, up to eight neurons were selected for simultaneous whole-cell patch-clamp recording primarily in current-clamp conditions, with a subset of stimuli administered in voltage-clamp conditions (Fig. 1B). Stimuli were elicited in each patched neuron while all others were recorded for evidence of a postsynaptic response. Dynamics (short-term plasticity [STP] and variability) of individual connections were assessed from stimulus trains at frequencies ranging from 10 to 200 Hz (Fig. 1B). Cells were stimulated with long current pulses to characterize their intrinsic physiology and later stained with biocytin to characterize their morphology (Fig. 1C).

Intralaminar connectivity in mouse VISp
Distance dependence of connectivity

The likelihood that two neurons are connected decreases with increasing intersomatic distance (25–27). We probed connectivity among cells up to ~200 μm apart (Fig. 2A) but could not ensure that intersomatic distances were sampled equally across different connection elements (specific pre-post combinations). To make reliable comparisons, we modeled the spatial profile of connectivity versus lateral somatic distance with a Gaussian (25) and estimated the peak (p_{max}) and lateral spread (sigma: σ) of connection probability using maximum likelihood estimation (see the materials and methods; fig. S1B). This analysis provided an estimate of the spatial profile of connectivity while being relatively robust to differences in the sampled intersomatic distribution. We initially classified cell pairs among the four combinations of excitatory (E) and inhibitory (I) cell classes (E→E, E→I, I→E, and I→I). These data were well fit by the Gaussian model (Fig. 2A, solid red line) and indicated low peak connectivity among excitatory cells (5%), with moderate connectivity rates among inhibitory cells (12%) and across E-I cell classes (E→I 12%, I→E 15%). In a tracing study of primary visual cortex

[1]Allen Institute for Brain Science, Seattle, WA, USA. [2]Byte Physics, Berlin, Germany. [3]The Ben and Catherine Ivy Center for Advanced Brain Tumor Treatment, Swedish Neuroscience Institute, Seattle, WA, USA. [4]Department of Neurological Surgery, University of Washington, Seattle, WA, USA. [5]Epilepsy Surgery and Functional Neurosurgery, Swedish Neuroscience Institute, Seattle, WA, USA. [6]Department of Pathology, University of Washington, Seattle, WA, USA. [7]Regional Epilepsy Center at Harborview Medical Center, Seattle, WA, USA.
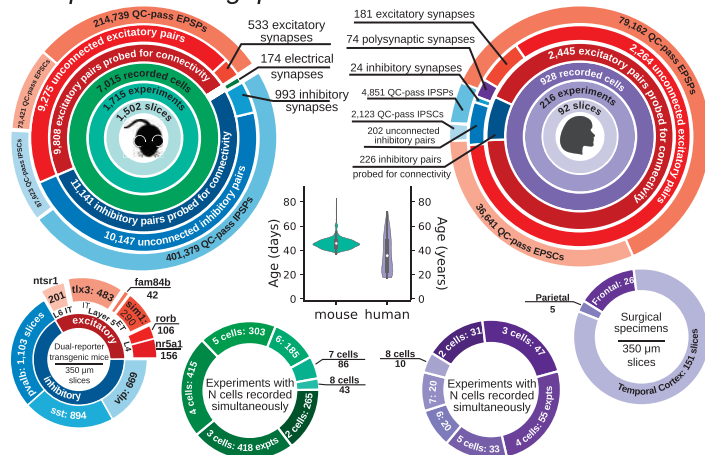*Corresponding author. Email: timj@alleninstitute.org
†These authors contributed equally to this work.
‡Present address: Sandia National Laboratories, Albuquerque, NM.
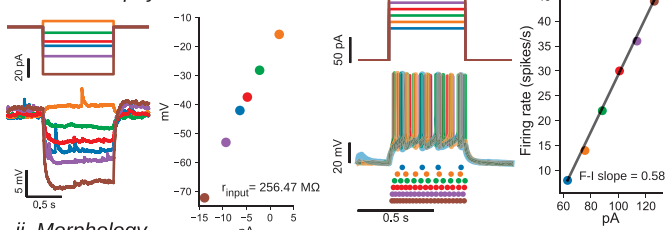§Present address: KB Cajal Neuroscience, Seattle, WA.
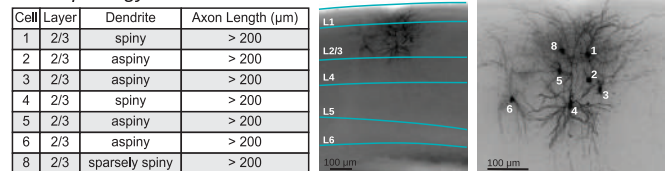
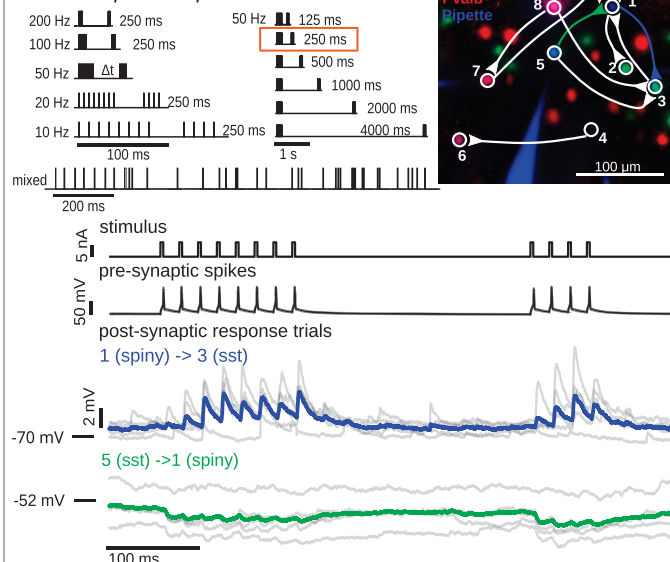## A Pipeline Throughput

B Connection Characterization

i. Multipatch Experiment

C Cell Characterization

i. Intrinsic Ephys

ii. Morphology

Cell	Layer	Dendrite	Axon Length (μm)
1	2/3	spiny	> 200
2	2/3	aspiny	> 200
3	2/3	aspiny	> 200
4	2/3	spiny	> 200
5	2/3	aspiny	> 200
6	2/3	aspiny	> 200
8	2/3	sparsely spiny	> 200

ii. Pair Processing
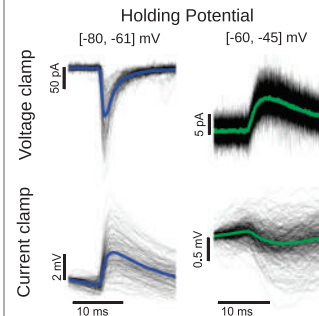
iii. Connection Analysis
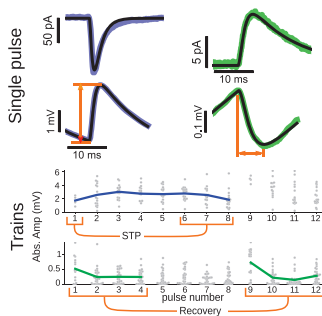
Fig. 1. Synaptic physiology pipeline. (**A**) Pipeline throughput summary. Large circles indicate data collection statistics (left, mouse primary visual cortex; right, human). Age distribution for each species is shown in the center. Distributions of recorded cells (outer circles) and experiment count with two through eight simultaneously recorded cells (inner circles) are shown at the bottom. (**B**) Connection characterization. (i) Multipatch experiment. Stimuli sets used to probe connectivity and dynamics are shown at the top. Fluorescent image (right) shows recorded cells with connectivity diagram overlaid. Example electrophysiology recordings of cells 1, 3, and 5 during the 50-Hz stimulus (orange box). (ii) Pair processing. For each recorded pair, presynaptic-spike aligned postsynaptic responses (black traces) were overlaid and averaged (colored trace corresponding to the same connection in the multipatch experiment panel) in both voltage-clamp (top row) and current-clamp (bottom row) conditions.

From this a connection could be identified. (iii) Connection analysis. When a connection was identified, the average (colored) was fit (black) and metrics such as amplitude and rise time (orange lines) were extracted from the fit. STP was quantified as a ratio from fits of individual PSPs. (**C**) (i) Intrinsic Electrophysiology. Long pulse steps were applied to quantify intrinsic cell properties and electrical connections (fig. S4). Hyperpolarizing steps (left) delivered to an example cell probed the subthreshold I–V relationship to quantify metrics such as input resistance and depolarizing suprathreshold sweeps (right) were used to measure spiking and firing rate properties. (ii) Morphology. Cells were filled with biocytin during recording and stained. Images (20×) of the full slice stained with DAPI allowed identification of the cortical layers, and 63× z-stack images were used to assess morphological properties such as dendritic type and axon length. Images and cells are the same as in (B).

L2/3, E→E connections had a wider lateral extent compared with I→E (*28*). We additionally found that E→E and I→I connections had a similar spatial extent that was larger than both the E→I and I→E connections.

An overabundance of bidirectional connections relative to unidirectional connections can be evidence for connectivity rules that promote the formation of bidirectional connections (*29*). A simpler explanation, however, is that bidirectionality is not specifically promoted in cortex but rather is an artifact of merging connectivity results across cell types

(*30*). We quantified the ratio of connected pairs with and without bidirectional connections and observed that reciprocal connections were three to five times more common than expected for a randomly connected network (Fig. 2A, bottom row, red line) among class-level connections (Fig. 2A, bottom row gray line).

Connection probability measurement: Slicing artifacts and detection limits

In the in vitro slice preparation, some connections may be severed, reducing the measured rate of connectivity (*25*, *31*). To mitigate this

bias, we used thick slices (350 μm), targeted cells deep in the slice (median cell depth = 74 μm; fig. S2B), and focused on local (<200 μm apart) intralaminar connections. By modeling the effects of cell depth and axon length on connection probability, we estimated the size of this bias and adjusted our p_{max} measurements accordingly. This yielded an ~3 to 6% increase in p_{max} when accounting for severed axons (fig. S2, A and D) and an ~2 to 20% increase in p_{max} when accounting for depth of the targeted cells from the slice surface (fig. S2, B and D).

Fig. 2. Mouse connectivity.
(**A**) Cells were divided into two main classes, excitatory and inhibitory, and pairs classified into the four combinations of those two classes. Top row: Connection probability as a function of intersomatic distance fit with a Gaussian (red line) and output parameters p_{max} and sigma (σ) describe the maximum connection probability and width of the Gaussian. Also shown is the connection probability as a function of intersomatic distance adjusted for presynaptic axon length, depth of the pair from the slice surface, and detection power of connections using a unified model (dashed red line) or by filtering of the data (dotted red line) (see the results section and the materials and methods). Gray line and area are 40-µm binned average connection probability and 95% CI. Raster below shows distance distribution of connections probed (bottom) and found (top). Bottom row: Normalized rate of reciprocal connections. Probed pairs are unordered and the number of reciprocal connections counted was normalized to the expected value of connection probability squared for a randomly connected network (solid red line). (**B**) Connection probability matrix for mouse. Connection probability is estimated using a unified model accounting for all corrections as determined from (A) (dashed red line, "model"). The shading of each element indicates the 95% CI of the data, with higher contrast indicating smaller CI and lower contrast (toward gray) indicating larger CI. The number of

connections found out of the number of connections probed are printed in each element. (**C** to **F**) Gaussian fit of connection probability versus intersomatic distance (with CI at p_{max}, shaded region) for two contrasting elements with connections found and connections probed raster below. Cross symbol denotes p_{max} with all adjustments.
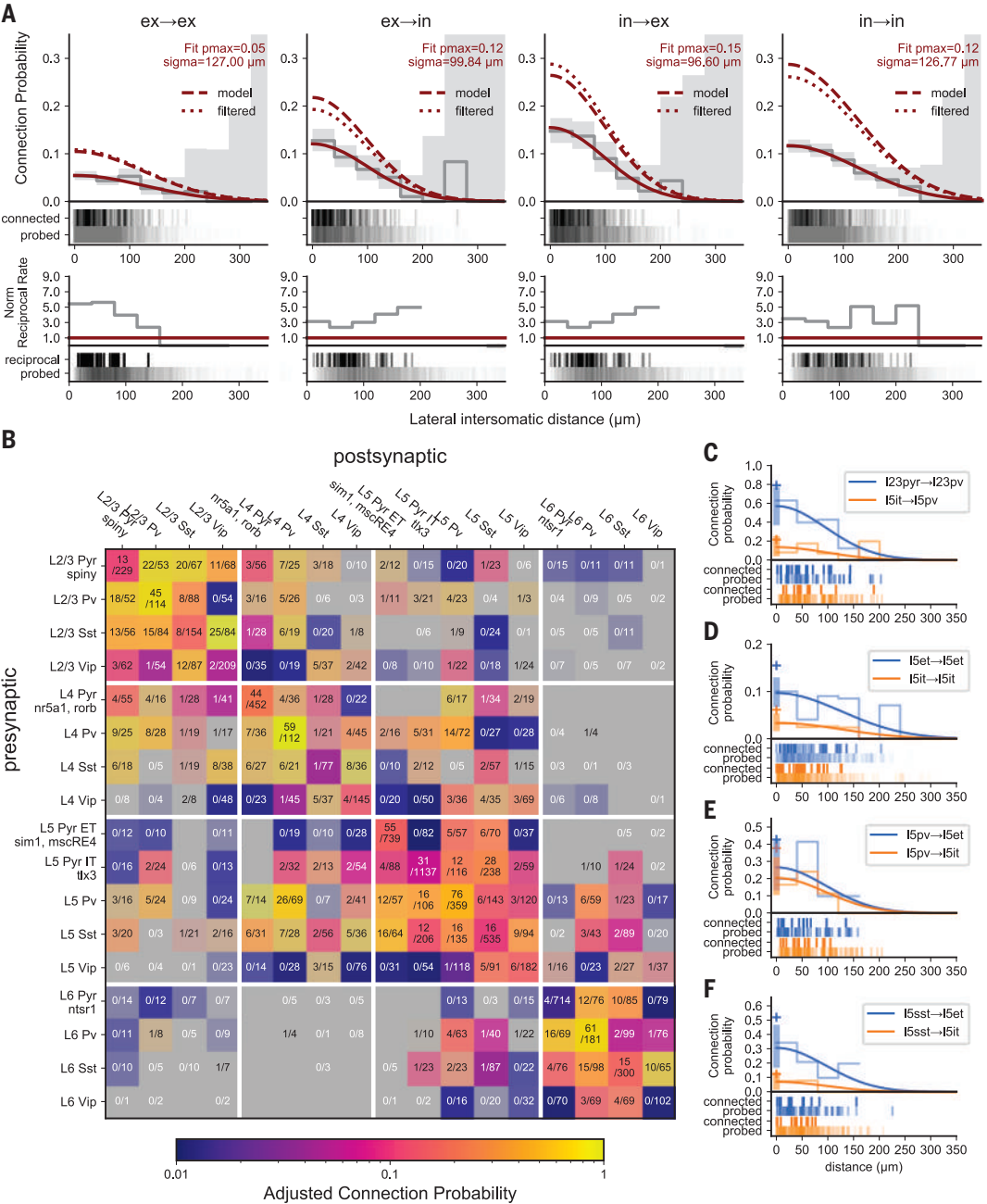
Detecting a connection in electrophysiological recordings is influenced by the background electrical noise, which may obscure synaptic responses in the postsynaptic cell. The effect of noise on detection is reduced by averaging the responses to multiple presynaptic action potentials (APs). Over the course of a typical experiment, most cells were stimulated to evoke hundreds of APs; however, this number varied significantly from cell to cell (485 [192, 623] APs, median [interquartile range, IQR]). To account for the opposing

effects of noise and averaging, we quantified the "detection power" (analogous to the signal-to-noise ratio; see the materials and methods) for each pair of cells that were probed for connectivity. We observed a range of detection powers across pairs, and these data were used to model the effect of detection power on connectivity. A model of the relationship between detection power and connection probability resulted in a 45% increase in estimated p_{max} for excitatory connections and up to a 90% increase for inhibitory connections (fig.

S2, D and E), suggesting that the observed connectivity rate is affected more by detection power than by slicing artifacts. Detection power is seldom reported and may explain cases in which the observed connectivity in vitro is higher than that in vivo for the same brain area and connections (*32, 33*).

We extended our model of connection probability on intersomatic distance (Fig. 2A, solid line) to include the effects of slicing and response detection outlined above (see the materials and methods). The model-adjusted

connectivity rate resulted in a one- to twofold increase in estimated p_{max} at the cell class level (Fig. 2A, dashed line). To confirm these results, we implemented a similar connectivity adjustment by filtering the data for axon lengths, pair depths, and detection power values above their respective median (Fig. 2A, dotted line), where the impact on connection probability is reduced (fig. S2, A to C). This yielded a comparable increase in estimated p_{max} (Fig. 2A, dotted versus dashed line), so we used the model to adjust peak connectivity estimates at the subclass level, where filtering would induce undersampling. Inhibitory connections received a higher detection power adjustment than excitatory connections. Adjustments for presynaptic axon length and pair depth were relatively uniform across subclasses (fig. S2E).

Connectivity among cell subclasses

The intralaminar connectivity among mouse VISp subclasses is summarized in Fig. 2B (also see fig. S1A). We found that intralaminar connection probability varied by layer (Fig. 2, B and C), long-range projection target (Fig. 2D), and cell subclass (Fig. 2, E and F). Connectivity was overall higher in superficial layers than in deep layers. For example, the pyramidal to Pvalb connection in L2/3 (p_{max} = 0.79 [0.57, 0.99], 95% confidence interval [CI] compared with L5 (IT p_{max} = 0.22 [0.12, 0.35]; χ^2 P = 7.2 × 10^{-54}) (Fig. 2C and fig. S1A). Recurrent connections between excitatory and Vip cells were common in L2/3 (E→Vip 0.38 [0.21, 0.60]; Vip→E 0.11 [0.03, 0.26]) but rare or absent in deeper layers (Fisher's P = 6 × 10^{-4}; fig. S1A).

Within L5, we found several differences between two excitatory projection classes, IT (labeled by Tlx3) and ET (labeled by Sim1 and mscRE4). ET cells overall have more input from local sources relative to IT cells. ET cells have higher recurrent connectivity (χ^2 P = 3.17 × 10^{-6}; Fig. 2D) and receive unidirectional input from IT cells, consistent with previous results (34). Sst cells also innervate ET cells at a higher rate than IT cells (χ^2 P = 3.18 × 10^{-5}; Fig. 2, E and F); a similar connectivity pattern was observed in the rat frontal cortex (35).

Sst is thought to avoid connecting with itself (4, 36, 37); however, we observed connections between Sst neurons in every layer (Fig. 2B). The Sst-IRES-Cre driver can sparsely label fast-spiking interneurons that also express Pvalb (38). Uniform manifold approximation and projection (UMAP) of intrinsic properties from inhibitory Cre types showed that slightly less than half of the recurrent Sst connections had both the pre- and postsynaptic cell in a cluster that was spatially distinct from Pvalb cells, suggesting that these connections came from cells that were intrinsically Sst like (fig. S3A). We confirmed that in at least one case, both cells in the pair had axons extending into L1 and sparsely spiny dendrites, consistent with

Martinotti neurons (fig. S3B). We performed experiments using the Patch-seq method (39) and further confirmed that cells that transcriptionally mapped to the Sst subclass do form connections with each other (five connections were found of 66 probed; fig. S3C).

Electrical connections

In addition to chemical synaptic transmission among cell subclasses, electrical connections formed by gap junctions were also found between inhibitory subclasses (fig. S4A). The likelihood of electrical connectivity as a function of lateral intersomatic distance could be approximated by a Gaussian but with a narrower profile (σ = 74 μm; fig. S4B) compared with chemical connections between inhibitory cells (σ = 127 μm, Fig. 2A; P < 0.001, fig. S4B). This is consistent with previous reports of electrical connections between nearby Pvalb cells showing that the average distance between electrically coupled cells is short, 40 to 80 μm (40, 41). Pvalb cells showed the highest rate of electrical connections (76/922, ~8.2%), whereas those among Vip cells were the most rare (15/908, 1.6%; χ^2 P = 1.8 × 10^{-10}), in contrast to a previous report of Vip electrical connections, which were more prevalent (7). Most of the electrical connections were found between like subclasses (164/170, 96%) and were bidirectional (148/170, 87%). The distance of the gap junction from the soma coupled with the potential for some rectification of electrical connections (42) could account for the few cases in which reciprocal electrical connections were not observed. The coupling coefficient of electrically coupled pairs was comparable across subclass (fig. S4C) largely due to the lower input resistance of Pvalb cells (fig. S4D). Estimating junctional conductance revealed stronger electrical connections between Pvalb cells (0.38 [0.20, 0.61] nS, median [IQR]) than either Sst (0.23 [0.19, 0.36] nS; Mann–Whitney [MW] P = 0.02) or Vip (0.07 [0.03, 0.19] nS; MW P = 5 × 10^{-4}) (fig. S4C).

Synaptic strength and kinetics

In addition to connectivity, synaptic properties such as strength, latency, and kinetics (rise time and decay tau) determine the impact of a connection on the postsynaptic neuron and, ultimately, on cortical processing. Although these properties have been described previously for some cortical connections (4, 27, 33, 43–45), measuring them systematically across many cell types enabled us to make direct comparisons.

At the class level, inhibitory connections showed short latencies (median = 1.07 ms), slow kinetics, and relatively strong postsynaptic potentials (PSPs). These trends were largely driven by the subclass of the presynaptic cell and Pvalb in particular. Pvalb connections were extremely fast, with submillisecond latencies (Fig. 3A) highlighted

by Pvalb→L6 excitatory connections (0.94 [0.86, 0.99] ms, median [IQR]). Pvalb cells also elicited large resting-state inhibitory postsynaptic potential (IPSP) amplitudes regardless of postsynaptic cell subclass. Presynaptic Sst cells stood out for having some of the slowest kinetics, independent of postsynaptic target (see Sst→Vip rise time: 7.02 [5.65, 9.75] ms, decay: 50.81 [26.97, 142.41] ms; Sst→L5 IT rise time: 7.46 [5.51, 9.92] ms, decay: 58.59 [27.76, 166.83] ms).

In contrast to inhibitory connections, excitatory connections generally had a long latency (median = 1.49 ms; Kolmogorov–Smirnov [KS] compared with inhibitory P = 1.4 × 10^{-34}), fast kinetics, and weak PSPs, all of which related more to the identity of the postsynaptic cell. E→I connections displayed faster rise times than recurrent excitatory connections (E→I 2.75 ms, E→E 3.88 ms, KS P = 2.97 × 10^{-12}; Fig. 3B). Recurrent excitatory connections showed some of the smallest amplitudes in the resting state (e.g., L5 ET→L5 ET, 0.27 [0.13, 0.49] mV), whereas E→I connections were stronger (Fig. 3, D and E) and generated the single biggest PSP (9.51 mV, L2/3 Pyr→L2/3 Sst). E→I connection properties can be further refined by postsynaptic cell subclass. Connections with postsynaptic Pvalb cells had faster kinetics (see L4 Pyr→Pvalb, Fig. 3; rise time: 1.45 [1.28, 1.94] ms, decay tau: 6.92 [5.4, 8.36] ms) than postsynaptic Sst cells (see L2/3 Pyr→Sst, Fig. 3; rise time: 3.91 [2.69, 4.79] ms, decay tau: 19.41 [14.88, 28.48] ms). A dichotomy between Pvalb and Sst was also apparent in the resting-state amplitude with larger excitatory postsynaptic potentials (EPSPs) to Pvalb cells (0.41 [0.22, 0.79] mV) than Sst cells (0.08 [0.05, 0.25] mV); however, the strongest excitatory connections were onto Vip cells found predominantly in superficial layers (L2/3 pyr→Vip 0.45 [0.26, 0.84] mV). Although resting-state excitation was weakest onto Sst cells, resting-state PSP amplitude is an underestimate of the potential impact on the postsynaptic cell, particularly for connections with strongly facilitating synapses. When we compared the 90th percentile amplitude (Fig. 3E), which measures near-maximal strength, E→Sst connections were comparatively stronger, and even surpassed E→Pvalb amplitudes in some cases (see L2/3 Pyr→Sst versus L4 Pyr→Pvalb). Facilitation onto inhibitory cells further contributes to the longer tails of 90th percentile amplitudes and the rightward shift of E→I amplitudes (0.73 mV) compared with E→E amplitudes (0.34 mV, KS P = 1.53 × 10^{-14}; Fig. 3E, histograms).

A recent survey of cortical connectivity found that connection strength positively correlated with connection probability (6). This result suggests an interesting principle of connectivity but may also result from the reduced detectability of weaker connections. When we adjusted for detection power, connection

Fig. 3. Synaptic strength and kinetics.
Left to right: E-I subclass matrix; excitatory and inhibitory minimum, median, and maximum average traces (light to dark colors); histograms for the major connection classes (E→E, E→I, I→I, I→E); and summary scatter plots for a subset of matrix elements for each metric PSP latency (**A**), PSP rise time (**B**), PSP decay tau (**C**), PSP resting-state amplitude (**D**), and PSP 90th percentile amplitude (**E**). In all matrices, inhibitory cells are merged across layers. All matrices are colorized by the median (text in each element) with the saturation scaled by the SE. Two or more pairs were required to fill in an element.
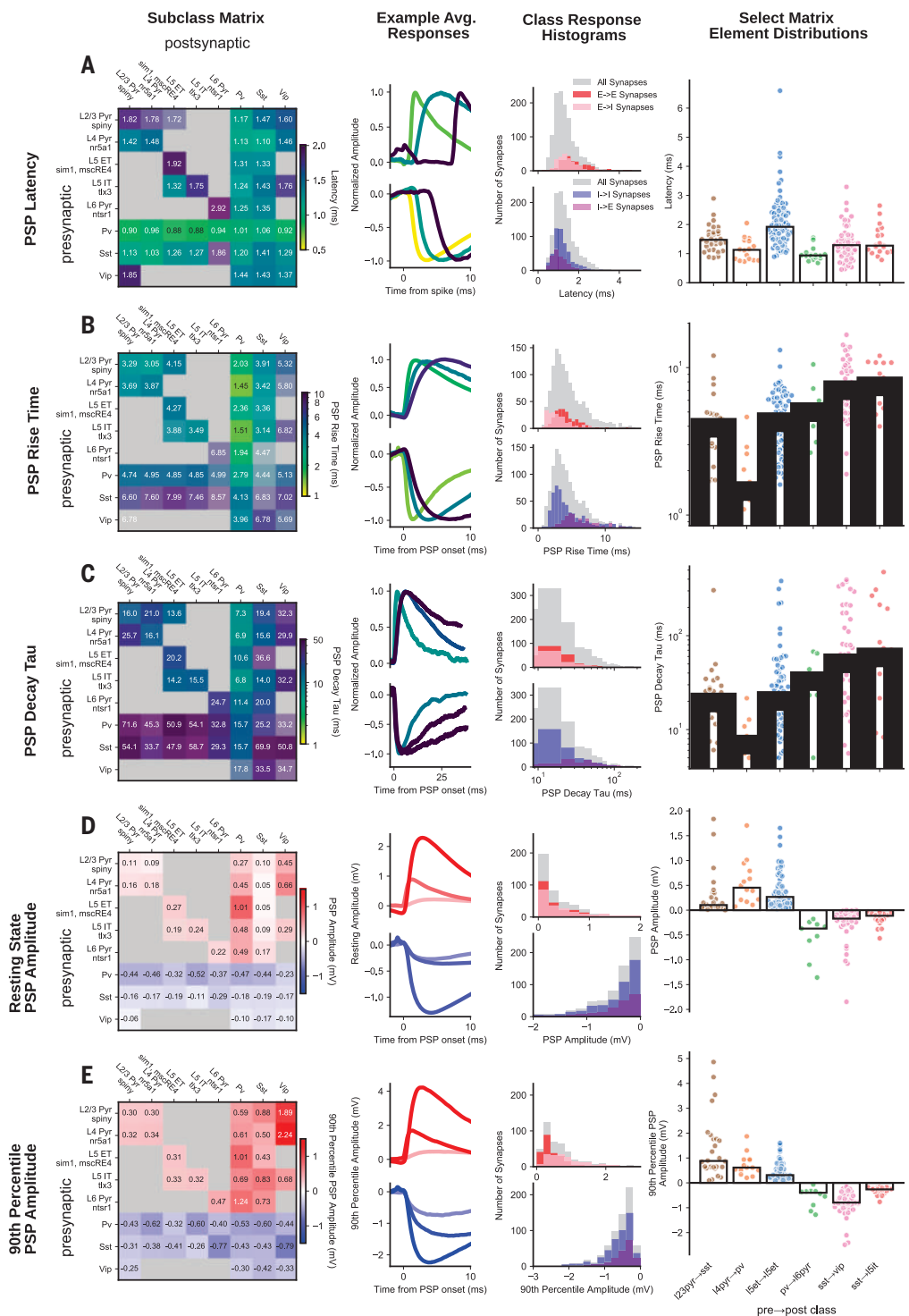
probability was independent of connection strength for both excitatory (weighted Huber regression r^2 = –0.1, P = 0.8) and inhibitory (r^2 = 0.2, P = 0.2) connections.

Synaptic dynamics

The strength and kinetic properties described above characterize the synapse in response to a single presynaptic spike; however, synapses are highly dynamic. PSP amplitude evolves in predictable ways over the course of milliseconds to seconds because of STP while also being highly stochastic from one response to the next, as quantified by the coefficient of variance. Overall, synaptic dynamics followed a similar pattern to synaptic strength, wherein excitatory connections were most strongly differentiated by the postsynaptic subclass

and inhibitory connections were differentiated by the presynaptic subclass. The STP of a synapse may result in a transient increase (facilitation), decrease (depression), or no change (pseudolinear) in PSP amplitude over the course of a stimulus train, as seen in our data (Fig. 4A) and as described previously (*16*, *20*, *46–48*). The time course of recovery from STP is an equally important property of synapses yet one that is

Fig. 4. Synaptic dynamics. (**A**) Depressing, facilitating, and pseudolinear excitatory connections (top to bottom) in 50-Hz train; gray/colored dots are individual PSP amplitudes; black traces are average PSP per pulse. Scatter points for pulse 1 (resting-state adjusted coefficient of variation [aCV]) and pulse 8 (STP-induced aCV) are colored according to the color scale in (D). (**B**) STP matrix. (**C**) Recovery (at 250 ms) matrix. (**D**) Resting-state aCV matrix. All matrices are colorized by the median (text in each element) with the saturation scaled by the SE. (**E**) Summary plots for paired-pulse ratio, STP induction ratio (average first pulse amplitude/average of the sixth to eighth pulse amplitude) normalized by the 90th percentile, resting-state aCV, and induced state aCV (top to bottom). Each dot corresponds to the average response from one connection. (**F**) Train-induced STP (top) at four different frequencies (10, 20, 50, and 100 Hz) for each of the elements in (E) (colors maintained). Each dot is the grand average of all connections in the element. For L5 ET→L5 ET, the blue shading highlights the 95% CI as an example. Lower plot shows recovery from STP at six different delays (125, 250, 500, 1000, 2000, and 4000 ms) in a similar manner to the plot above.
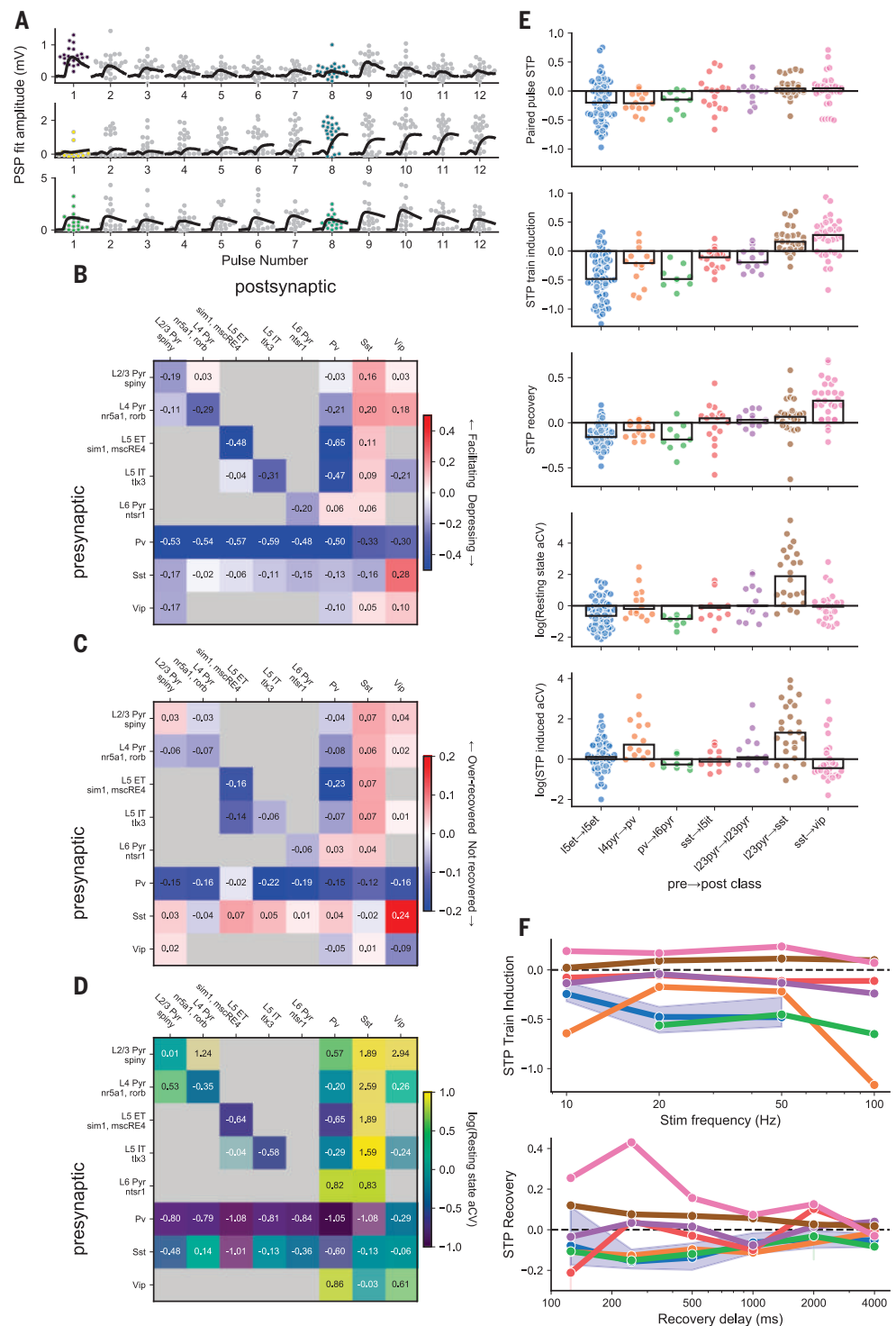
not well described. The variable delays that we imposed between the induction and recovery pulses (Fig. 1B, multipatch experiment) of our 50-Hz stimulus showed that at our earliest time point (125 ms), connections were still in their STP-induced state, but by 4 s, they were largely recovered.

Excitatory dynamics were strongly aligned with postsynaptic cell class (48) and further refined by layer in the case of excitatory targets and by subclass of inhibitory targets. Recurrent excitatory connections were largely depressing (Fig. 4, B and E, L5ET→L5ET) and showed increasing depression with stimulus frequency (Fig. 4F). Recurrent excitatory connections occupied a range of recovery and variability profiles that varied with layer. Superficial layers (e.g., L2/3→L2/3) tended to

recover more quickly (Fig. 4C) and showed a higher degree of variability (Fig. 4D). E→I dynamics depend on the subclass of the post-synaptic target. Excitatory connections to Sst cells were strongly facilitating, consistent with previous reports (16, 47). E→Sst connections were also highly variable in the resting state, likely because of a high initial failure rate (Fig. 4A, middle, pulse 1, Fig. 4E); however,

strongly facilitating synapses often become more reliable in the induced state (Fig. 4A, middle, pulse 8, Fig. 4E). Excitatory connections onto Pvalb (Fig. 4E, L4 pyr→Pvalb) were largely depressing on average, although a subset of connections in L2/3 showed pseudolinear STP that was similar to in vivo measurements in somatosensory cortex (45). Whereas these patterns of excitatory dynamics were apparent on average, multiple measurements of synaptic dynamics showed high heterogeneity from pair to pair within a connection type (Fig. 4E).

Dynamics of inhibitory connections showed patterns more related to the subclass identity of the inhibitory presynaptic cell. Pvalb connections onto other subclasses, both excitatory (Fig. 4E, Pvalb→L6 pyr) and inhibitory, were strongly depressing (Fig. 4B) and were still depressed at our earliest recovery time point (Fig. 4C). Depressing Pvalb connections showed high reliability at the beginning of a stimulus train (resting-state log variability = –1.01; Fig. 4D) and became more variable in their STP-induced state (log variability = –0.27, MW P = 3.65×10^{-31}; Fig. 4E). Connections from Sst and Vip cells had a mixture of dynamics from facilitation to moderate depression (Fig. 4E, Sst→L5 IT). These connections were also faster to recover from STP, particularly Sst connections, which tended to over-recover at the shortest interval (Fig. 4C).

Synaptic interactions between Sst and Vip were an exception to the trends highlighted above. Whereas most inhibitory connections were depressing, Sst→Vip (7) showed the highest degree of facilitation in our dataset (0.28 [0.0, 0.44]). The reciprocal Vip→Sst connection was weakly facilitating, as were recurrent Vip connections. These three connection types also over-recovered on short time scales (Fig. 4C) and took many seconds to fully recover (Fig. 4F). Given the facilitating nature of these connections, it is interesting that they had only a moderate degree of variance (resting-state log variability = –0.03; Fig. 4D) compared with other facilitating connections such as E→Sst (resting-state log variability = 1.66; MW P = -3.0×10^{-8}).

Human intralaminar connectivity

As a complement to the mouse visual cortex, our dataset includes synaptic physiology from human temporal cortex. Although our sampling of human connections covered all cortical layers, our analysis focused on the supragranular layers, which are expanded in anthropoid primate cortex (49). Deep L3 cells have distinct electrophysiology, morphology, and gene expression (including genes involved in connectivity and synaptic signaling), and many of these properties vary continuously with depth between L2 and L3b (50). Dense sampling of L2/3 allowed us to

define L2, L3a, and L3b pyramidal subclasses and demonstrate that these principles of cellular diversity have correlates in synaptic physiology. These subclasses showed distinct synaptic properties, including unique polysynaptic connections from L2 cells and STP that closely follow the continuous variability between L2/3 subclasses.

Distance dependence of connections was modeled and adjusted as with mouse connections but without distinguishing connections by cell class (most connections were recurrent excitatory). Connection probability was estimated to fall off with distance at a lateral spread (σ) of 130 μm (Gaussian model fit; fig. S5A), slightly larger than the comparable value in mouse (125 μm for within-class connections), reflecting that whereas cortical expansion is accompanied by the scaling of neuronal morphology, much of this scaling is axial rather than lateral. Examining the connectivity between subclasses (Fig. 5A), we tested for signatures of functional segregation within supragranular layers, finding a strong bias for recurrent connections over cross-connections between L3a and L3b (Fisher's P = 5.6×10^{-3}) and a bias for connections from L2 to L3a over L3b (Fisher's P = 3.4×10^{-3}).

Recurrent connectivity within human L4 (p_{max} = 0.01 [0.0, 0.04] 95% CI, 1/145) was significantly lower than observed in the mouse (p_{max} = 0.22 [0.16, 0.28], 44/452; Fisher's P = 1.4×10^{-4}). This contrast could be related to age (37), species, or brain area. Furthermore, we observed other connections involving L4 pyramidal cells (e.g., I→E and E→I), suggesting that low excitatory recurrence was not a technical limitation of our dataset but rather reveals a unique property of the human L4 circuit.

Human synaptic properties

The strength, kinetics, and STP properties of the human connections showed moderate differences across layers and large differences by cell class, largely resembling observations in mouse (Fig. 5B). Recurrent excitatory connections in human cortex (E→E) have longer latency than those with a pre- or postsynaptic inhibitory cell (E→E median 1.73 versus I→E 1.04 ms, KS test P = 2.2×10^{-5}; E→I 1.34 ms, P = 2.4×10^{-3}), and PSP rise times were faster for E→I than for E→E connections (2.42 versus 4.10 ms, P = 2.0×10^{-8}) but slower for I→E connections (5.72 ms; P = 5.3×10^{-3}, 4.9×10^{-6} versus E→E, E→I). We observe some differences between L2 and L3, including presynaptic L3 cells forming more depressing connections than L2 cells (STP ratios –0.39 versus –0.15, P = 9.2×10^{-3}). E→I connections were uniformly depressing, consistent with the identification of those inhibitory cells as fast-spiking Pvalb cells.

For certain properties, we did observe contrasts between human and mouse connec-

tions. The amplitudes of L2/3 excitatory PSPs were generally larger in human. For E→E connections, the contrast was stronger for the 90th percentile response (0.51 versus 0.30 mV, P = 0.019), whereas for E→I connections, the contrast was stronger for resting-state response (0.91 versus 0.26 mV, P = 2.2×10^{-3}) because of significantly more depressing E→I connections in human (–0.51 versus 0.05, P = 7.1×10^{-6}). We also observed faster recovery from STP in human than in mouse excitatory connections, with most fully recovered at 500 ms (Fig. 5C). This contrast has been previously noted in recurrent L2/3 excitatory connections (12), and our observations suggest that it also holds for L2/3 E→I connections.

Human polysynaptic events

We found short-latency (~3 ms) inhibition (Fig. 5D) after stimulation of excitatory cells in human cortex, indicating activation of polysynaptic connections. Plotting latency versus PSP amplitude of human connections (Fig. 5D) revealed a clear boundary where responses with a latency of ≥3 ms, evoked from a confirmed pyramidal cell, were almost exclusively inhibitory (median latency of IPSPs: 4.19 [3.79, 4.81] ms), compared with monosynaptic EPSPs, which had a latency of <3 ms (median EPSP latency: 1.71 [1.44, 2.22] ms). This potential disynaptic inhibition (dIPSPs) originated in L2 and projected to other L2 (n = 8) or L3 (n = 11) pyramidal cells, with just one ascending polysynaptic response originating in L3. This directionality was consistent with the directionality of monosynaptic excitation across layers 2 and 3 and supported by higher connection rates from L2 pyramidal cells to inhibitory cells.

Variation with depth in human L2 and L3

Another property of supragranular neurons observed in human, but not in mouse, is strong depth-driven variability of intrinsic electrophysiological properties (50, 51). Visualizing the electrophysiology feature space by a UMAP projection of 27 electrophysiology features (see the materials and methods), we found that L4-type cells (high input resistance) are situated around the perimeter, indicating distinct properties from L2/3-type cells (Fig. 5E). For the L2/3-type cells, projecting a normalized layer depth coordinate (relative to the L2+L3 thickness) onto this space showed a mostly smooth gradient of electrophysiological properties with layer depth, also verified by direct examination of depth correlations with sag and AP upstroke/downstroke ratio (P = 7.0×10^{-6}, 3.3×10^{-5}) (Fig. 5F). This correlation was not found in mouse L2/3 cells (P > 0.06 for both).

STP metrics revealed a similar linear variation with layer depth in connections from L2/3 pyramidal cells across both E→E (n = 86 human, n = 15 mouse) and E→I (n = 20 human,

Fig. 5. Human data. (**A**) Connection probability measured from human cortical tissue. Inhibitory cells are identified by morphology as aspiny or sparsely spiny cells, grouped across layer. (**B**) Kinetics, strength, and dynamics matrices are organized by layer for excitatory cells, with inhibitory cells grouped across layer. Each element is colorized by the median (text in each element) with the saturation scaled to the SE. Two or more pairs were required to fill in an element. Latency, rise tau, and resting-state amplitude are quantified from fits of the average PSP response. (**C**) Train-induced STP (top) across frequencies for a subset of connection types. Each dot is the median of all connections in the element, with shading for the 95% CI (bootstrapped) shown for a single example connection type. Recovery from STP at different delays is shown in the lower plot. (**D**) Example polysynaptic circuit from one experiment in which cell 1 forms a short-latency (~2 ms) monosynaptic excitatory connection to cell 2 and delayed (~4 ms) polysynaptic inhibitory connections to cells 3 and 4 (all cells confirmed morphologically spiny). Dashed lines indicate (from left to right) time of presynaptic spike and PSP onset. Polysynaptic connections from L2/3 pyramidal cells were inferred by response latency >3 ms versus PSP amplitude. (**E**) Structure of intrinsic electrophysiology feature space. UMAP projection colored by cell subclass (left) and by depth of L2/3-type excitatory cells (right). (**F**) Variation in L2/3 intrinsic properties is strongly correlated with depth in human but not mouse. Example traces show superficial

and deep human cells [top, colors by L2/3 depth as in (E)]. Left is the phase plane representation of the first spike in a depolarizing step response. Right is the sag in response to hyperpolarization. Bottom is the regression of corresponding electrophysiology features versus depth by species, with bootstrapped 95% CI. (**G**) STP of L2/3 excitatory connections is structured by depth in human and not mouse. Top shows the PSP responses to spike trains for example cells from (F). The larger response to the first spike is quantified by paired-pulse STP, plotted below in relation to presynaptic depth (left) and AP up/down ratio (right) (postsynaptic relationships are shown in fig. S5E).
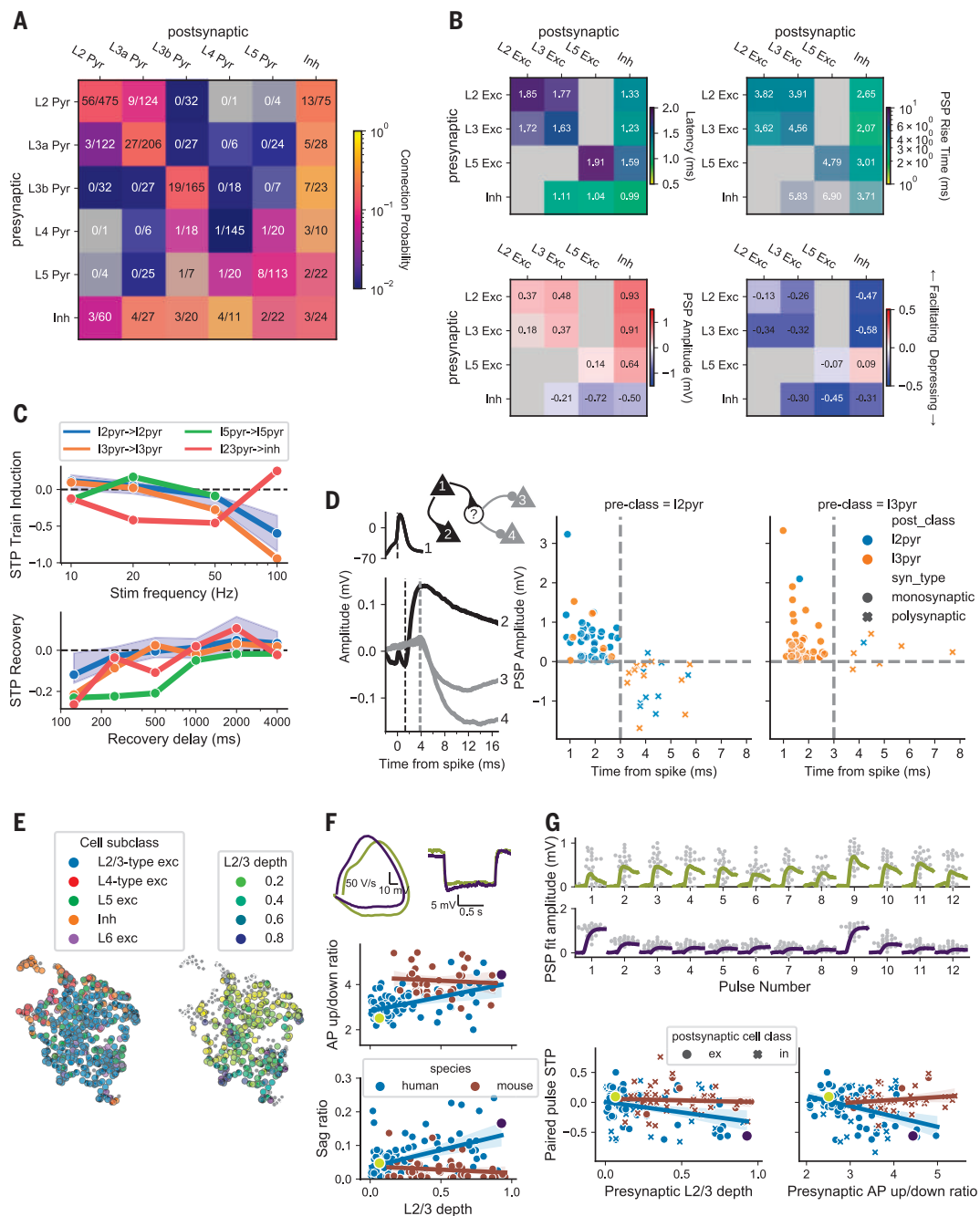
n = 72 mouse) connections. The paired-pulse STP showed a strong linear relationship with depth in the human data ($P = 5.4 \times 10^{-4}$), varying from weak facilitation for the most superficial cells (0.02 ± 0.03, mean ± SEM) to depression for the deepest (−0.32 ± 0.09). A strong correlation was also found with the AP upstroke/downstroke ratio of the presynaptic cell only ($P = 1.2 \times 10^{-4}$ versus $P > 0.3$ for post-

synaptic) (Fig. 5G). No corresponding trends were found in the mouse data ($P > 0.2$ for all regressions). Although lower sampling of L2/3 pyramidal cells in the mouse may contribute, regression coefficients for STP against the AP upstroke/downstroke ratio show a strong contrast (human −0.16 [−0.24, −0.08] CI; mouse 0.05 [−0.03, 0.13]), suggesting that there are real differences between these datasets in the

factors contributing to STP variability, whether explained by species, brain area, or other factors.

Modeling STP of mouse and human connections

The STP metrics introduced above were chosen for their ease of interpretation but have several drawbacks: They are sensitive to noise, they require successfully repeated stimuli that are not available for all connections, and they are

difficult to use in a biophysical modeling context. Ideally, we would like a description that incorporates all of the data available for each connection and can predict synapse behavior in response to any arbitrary stimulus. We developed a generative model of stochastic vesicle release and STP with several adjustable parameters (see the materials and methods) and investigated which combinations of parameters could best explain the responses recorded for each connection. In this way, we captured and described more of the dynamic behavior of each connection with a small number of parameters. Our approach is similar to other recently developed models (*52, 53*) in that it does not depend on any particular stimulus structure (aside from having a diversity of interspike intervals). Accordingly, the model results for each connection make use of all presynaptic spikes and are robust to spike failures and early experiment termination.

Model performance was evaluated by using the maximum likelihood parameter set for each connection to simulate experimental data. These simulated data were then used to generate the same STP metrics that were previously collected from synaptic data. Both resting-state PSP amplitude and 90th percentile amplitude were almost perfectly correlated between recorded and simulated data (fig. S6, A and B), indicating that the model does exceptionally well at capturing synaptic strength. STP and variability (fig. S6, C to F) were also strongly correlated but with more scatter relative to strength metrics. STP measured from the second pulse in 50-Hz trains was only half as large as that measured in synaptic responses, indicating that the model as parameterized was not able to fully capture STP on this time scale. Release probability was more highly correlated with variance than with the number of release sites, suggesting that synapses may control variability primarily through their release probability.

Most discussions of short-term depression in the cortical literature begin with the assumption that depression is caused by the depletion of vesicles from the readily-releasable pool. However, recent evidence suggests that calcium channel inactivation may be a more prominent mechanism in cortical depression (*54*). We ran the model on two separate parameter spaces, one that uses vesicle depletion and another that uses a release-independent depression mechanism. In most cases, the model maximum likelihood value was found in the release-independent parameter space (fig. S7E, left). Release-dependent depression mechanisms should result in negative correlation between consecutive PSP amplitudes; however, we found little evidence for such negative correlations in our mouse data. Likewise, we found little relationship between paired correlation values and the model preference

for release-dependent mechanisms (fig. S7E, right). These results are consistent with the proposal that cortical synapses in mouse use release-independent depression mechanisms and that vesicle depletion plays a relatively minor role in depression. By comparison, our data from human connections had a modest preference for negative correlation between paired event amplitudes.

Organization in mouse and human synaptic dynamics

With a large dataset describing synaptic properties of connections, it becomes possible to ask what patterns emerge from the data. What synaptic features correlate with one another, and do connections naturally split into clusters based on these features or do they form a continuum? What aspects of the synaptic feature space are driven by presynaptic versus postsynaptic cell type? Excitatory connections onto Pvalb and Sst cells have distinctly different dynamics, suggesting a general rule that excitatory dynamics depend primarily on the postsynaptic cell type (*16, 55*). By contrast, inhibitory dynamics depend mainly on the presynaptic type, particularly when comparing Pvalb with Sst (*47, 49*). Although our data often followed these rules, we also found exceptions and suggest some refinements.

We used the stochastic release model described above to characterize 1196 connections (1035 mouse, 161 human). Although it would have been possible to use the maximum likelihood parameter set to describe each connection, parameters derived in this way are sensitive to experimental noise (*56*). Thus, we generated a more robust representation of synaptic behavior by exhaustively searching a large parameter space for each connection and then used sparse principal components analysis followed by UMAP dimensionality reduction to summarize these results (Fig. 6). This analysis groups connections based on the similarity of their model results; therefore, it has access to any synaptic strength and dynamical properties that the model could capture but does not have access to information about kinetics, cell subclass, or other cell properties. Connections in this analysis formed clear excitatory and inhibitory clusters (Fig. 6A), with a continuum of synaptic properties within each cluster. Perhaps the most prominent feature of this organization was that excitatory connections were strongly differentiated by the postsynaptic E/I class (66% classifier accuracy gain relative to shuffled; see the materials and methods), whereas inhibitory connection properties were mostly independent of postsynaptic cell class (7% gain).

Several synaptic properties were found to have a clear relationship to the UMAP organization (Fig. 6B). Which of these properties best explains the separation of excitatory synapses

by postsynaptic class? We generated a list of features from the maximum likelihood models and ranked these by the strength of their relationship to cell subclasses (see the materials and methods). When distinguishing between E→E and E→I connections, features that describe quantal release and synapse variance made up eight of the top 10 features (table S2). By contrast, STP parameters had relatively low importance in this classification, and the effect of synaptic strength was negligible. This relationship was also apparent when classifying excitatory connections among excitatory, Pvalb, Sst, and Vip postsynaptic subclasses (48% gain; Fig. 6B). Again, this relationship appears to be driven by quantal release parameters and variability metrics, whereas STP metrics were somewhat less important. Only minor relationships were found between excitatory synapse properties and postsynaptic excitatory subclasses (13% gain; Fig. 6C) or inhibitory synapse properties and postsynaptic subclasses (excitatory and inhibitory subclasses, 14% gain; excitatory subclasses, 5% gain; Fig. 6B). To simplify, excitatory synapses have low variability when the postsynaptic cell is excitatory, high variability when the postsynaptic cell is inhibitory, and highest variability onto Sst cells.

Inhibitory synapses, by contrast, did not follow this pattern. Instead, inhibitory synapse properties were more predictive of the presynaptic subclass (41% gain). The parameters that most contributed to this relationship include STP, variability, and to a lesser degree synaptic strength. Excitatory synapses were more weakly affected by presynaptic subclass (24% gain), but this relatively small effect appeared to be similarly driven by differences in STP and variability.

Human E→E synapses were strongly differentiated from mouse E→E synapses (43% gain). This was driven by a more diverse set of properties, including STP (especially recovery), variability, facilitation time constant, and strength (table S2, last column).

Discussion

By probing >20,000 possible connections across 28 mouse lines, we have explored a large fraction of the subclass-specific intralaminar connectivity in the mouse visual cortex. In addition, we sampled connections in the human cortex using similar methods for comparison. Past surveys near this scale have focused on the connectivity and strength of connections; a major advance provided by our study is the depth of characterization and analysis for each connection, in the context of transgenically identified cell subclasses and species.

Proposed standardized model of connectivity

The likelihood that two neurons are locally connected depends on multiple factors such as cell type, cortical region, species, and animal
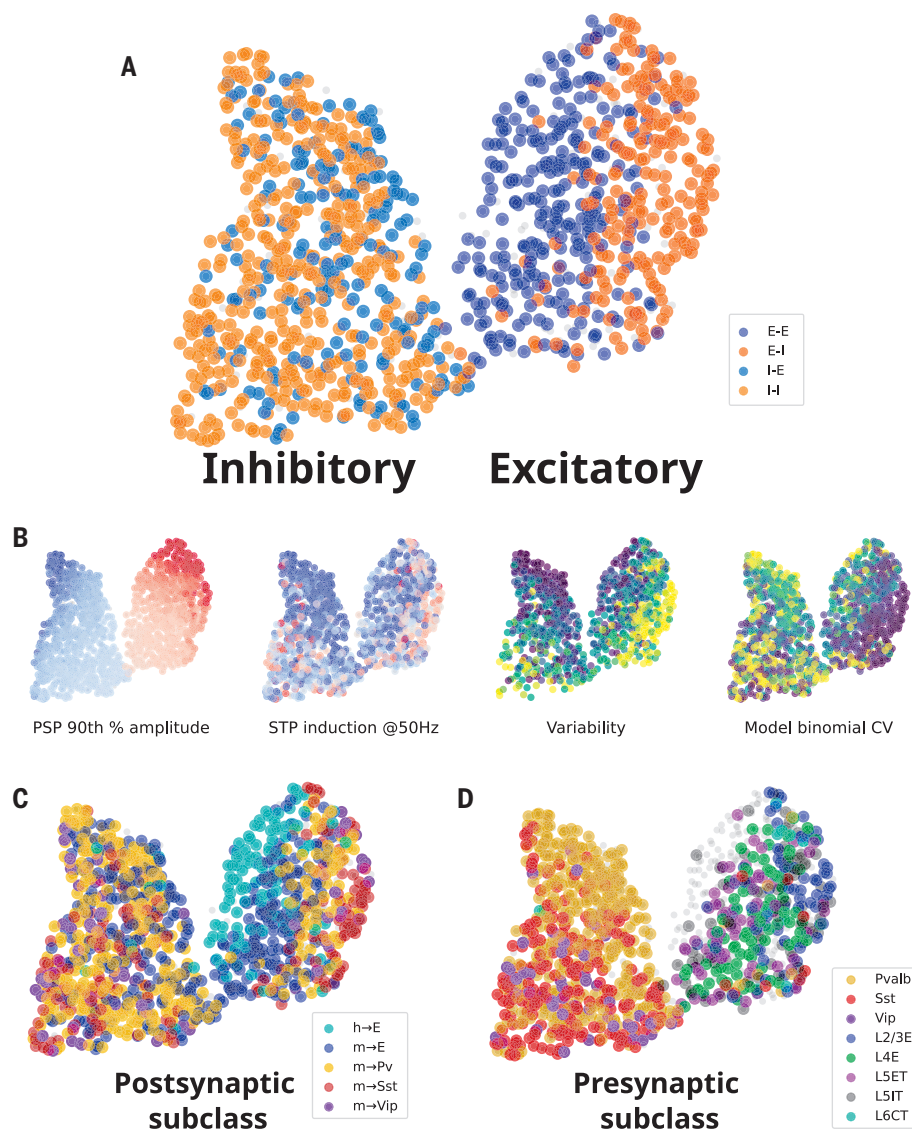
Fig. 6. Dimensionality reduction on connection properties. Relationships among connection properties and cell types revealed by dimensionality reduction. (**A**) All connections colored by postsynaptic E/I cell class. The UMAP output generates two clusters: inhibitory (left) and excitatory (right). (**B**) Four connection properties represented in reduced space, showing 90th percentile PSP amplitude (red, excitatory; blue, inhibitory); STP induced by 50 Hz trains (red, facilitating; blue, depressing), resting-state aCV during 50-Hz trains (purple, low variability; yellow, high variability), and the binomial CV derived from model parameters (release probability * number of release sites; purple, high CV; yellow, low CV). (**C**) Human and mouse connections colored by postsynaptic subclass. (**D**) Mouse connections colored by presynaptic subclass.

scribing the relationships between the excitatory and inhibitory subclasses and their functional relevance. The details of this picture vary somewhat between descriptions, but a few key elements appear consistently, especially in the systems and theoretical neuroscience literature (Fig. 7A). Pvalb interneurons strongly inhibit nearby pyramidal cells and other Pvalb cells, Sst interneurons broadly inhibit nearby cells but avoid other Sst cells, and Vip cells selectively inhibit Sst cells and receive feedback excitation, forming a disinhibitory circuit (*57*, *58*). We confirmed that each of these motifs is prominent across layers in the intralaminar cortical circuit.

Other circuit elements are equally prominent in our data but are more sparsely acknowledged in the literature. Many recent studies have focused on the importance of the Vip→Sst disinhibitory circuit. In the opposite direction, however, the connection from Sst to Vip has one of the highest connection probabilities, largest IPSP amplitudes, and strongest facilitation in our dataset. Although Sst→Vip connections have been described previously (*4*, *6*, *7*), they are often overlooked in consideration of the disinhibitory circuit. The synaptic features that we observed suggest an important functional ramification on the opposing Vip→Sst disinhibitory pathway (*57*, *59*, *60*).

Sst and Vip cells are often described as lacking recurrent connections (*36*, *37*, *48*, *61*) despite some evidence to the contrary (*6*, *62*). We confirmed that Sst and Vip do have the expected biases in their connectivity across all layers (Sst cells tend to avoid contacting other Sst cells, and Vip cells prefer to contact Sst cells). However, we also found sparse, recurrent connections within both interneuron populations approximately equal to the recurrent connectivity in excitatory populations. Furthermore, the strength of connections from Sst and Vip did not follow the same preferences, having roughly equal strength when connecting to preferred versus nonpreferred subclasses. Given their translaminar axon projection patterns, recurrent connections within these subclasses may be found more commonly across layer boundaries (*63*).

Laminar variations on the cortical circuit

Previous studies that sampled both L2/3 and L5 noted many similarities between the two layers (*4*, *6*). Although we confirmed a consistent set of connectivity rules describing the intralaminar circuit, we also found variations on these rules that could support different modes of cortical function. Differences in intralaminar circuitry may contribute to laminar differences in receptive field properties (*64*) or visually mediated behaviors (*58*).

Layer 2/3 has strong interconnections between pyramidal and inhibitory cells (Fig. 7B).

age. A longstanding goal has been to determine the governing principles of local circuit architecture. It is difficult to make direct comparisons among different studies because the observed rate of connectivity depends on several experimental details that are often inadequately reported or controlled for. Ideally, we would like a way to describe connectivity that allows more direct comparison between experiments regardless of their methodological differences. To facilitate such a comparison between connection subclasses in our own data, we developed a procedure for modeling connection probability as it relates to intersomatic

distance, axon truncation, cell depth, and signal detection power. With this model, we can estimate unbiased connection probabilities with confidence intervals that should be relatively robust to experimental bias. In principle, this approach is flexible enough to be replicated elsewhere in the field, and its adoption would substantially improve our ability to compare and reproduce results across studies.

Conserved and canonical elements in the mouse intralaminar circuit

As many previous studies have investigated the cortical circuit, a picture has emerged de-

Fig. 7. Intralaminar circuit diagram. The cortical intralayer circuit differs across layer and with activity. (**A**) Some commonly described elements of the intralaminar cortical circuit. Pvalb cells strongly inhibit pyramidal and other Pvalb cells, Sst cells provide broad inhibition, and Vip cells inhibit Sst cells to form a disinhibitory feedback pathway. (**B** and **C**) Circuit diagrams showing connections between major subclasses in mouse L2/3 (B) and L5 (C). The width of connecting lines roughly represents connection probability and PSP amplitude. Connections that are prominent in each layer compared with the other are highlighted in orange, and green lines indicate connections that are less prevalent in that layer. For simplicity, connections from IT pyramidal to inhibitory cells in L5 (C) are omitted. (**D**) Two complementary circuits that activate at different times. Red connections are facilitating and will be stronger during sustained activity. Blue connections are depressing and are strongest during quiescent periods.

In deep layers, these connections are either absent or greatly reduced, and the relative sparsity of Vip cells in particular in deep layers should further enhance these differences (*65*, *66*). Direct Vip inhibition of local pyramidal cells further complicates the prevailing view of the Vip→Sst disinhibitory pathway by, for example, allowing the possibility of feedback inhibition from higher cortical regions. Likewise, local excitatory inputs to Vip cells could be a source of feedforward disinhibition in layer 2/3.

Layer 5 excitatory subclasses differ in their visual responses and long-range projections, suggesting different functional roles in the circuit (*67*, *68*). Accordingly, we saw differences in the intralaminar connections of L5 ET and IT neurons. ET pyramidal cells were generally more highly connected, receiving more local excitation and inhibition than IT cells. We confirmed a much higher rate of recurrent connections among L5 ET cells compared with IT cells and also found that connections between these two subclasses were unidirectional from IT→ET (Fig. 7C) (*34*). Layer 5 ET cells also received more frequent inhibition from Sst cells, as previously observed in the rat frontal cortex (*35*).

Could laminar differences in connectivity indicate cell type divisions within subclasses? Two recent studies investigated the correspondence among morphological, electrophysiological, and transcriptomic (MET) features of inhibitory neurons in primary visual cortex and motor cortex (*63*, *69*). In visual cortex, different MET types had distinct patterns of local axonal innervation and dendritic morphologies (*63*), suggesting that their connec-

tivity will be different. MET types also exhibited layer localization, and thus some of the differences in connectivity that we observed as a function of layer may reflect differences in connectivity between different MET types.

Dynamic flexibility in the cortical circuit

Most studies in cortical synaptic physiology describe the circuit in its quiescent state. Ongoing activity in vivo, however, can transiently depress or facilitate synapses in a cell type–dependent way. In effect, the cortical circuit is modified from moment to moment based on the recent history of activity, so it may be misleading to conceptualize the cortex as a single, static circuit diagram. Although STP is variable across individual synapses, many subclass elements of the cortical circuit had a clear preference for either facilitation or depression (Fig. 7D). These patterns suggest an ability to dynamically switch between different network modalities in which intralaminar activity shifts from depressing interactions between pyramidal and Pvalb cells at the onset of activity to facilitating interactions among pyramidal, Sst, and Vip cells during sustained activity. Ultimately, activity patterns in vivo determine the extent of STP in each cell type, and the functional effects have yet to be determined.

Excitatory cells receive weak local excitation compared with inhibitory cells

Our previous study (*13*) observed low recurrent excitatory connectivity rates in all layers. We now find that recurrent excitatory connections have a set of features that distinguish them from excitatory inputs to inhibitory cells

and appear to limit their contribution to excitability. In addition to being sparse, they are relatively weak, get weaker with activity, and have slower PSP rise times. Slow PSP rise times limit excitability by raising the AP threshold (*70*). Slow rise times have been attributed to the PSPs being less synchronous at the multiple synapses that form E-E connections (*71*). Unlike the relatively high rate of intralaminar connectivity for E-I connections observed in all layers, recurrent excitatory connectivity is generally lower, and nearly absent in human L4 and in CT cells of mouse L6.

Unidirectional disynaptic inhibition in human

We observed disynaptic inhibition in human cortex between confirmed spiny pyramidal cells that is unidirectional, originating in L2 and targeting other L2 or L3 pyramidal cells. We did not see disynaptic inhibition in our mouse recordings, which may be due to the stronger excitation that we and others (*9*) have observed in human connections, particularly onto inhibitory cells. Disynaptic inhibition is often mediated by an interposed Sst cell (*72*, *73*) because these cells have a low spiking threshold and receive facilitating inputs from excitatory cells. The latency of Sst-mediated disynaptic inhibition is often long (>100 ms), but we saw disynaptic IPSPs with a much shorter latency (3 to 6 ms). This suggests the disynaptic inhibition is driven by an intermediate fast-spiking Pvalb cell, which has been observed in humans and can be recruited by very large excitatory events (*14*). The unidirectional nature of this disynaptic inhibition from more superficial to deeper cortex further

suggests a preferential routing of information by Pvalb cells in human cortex (*14*).

Connection types differ in variability

We used a new model of stochastic vesicle release and STP to estimate parameters that best describe each connection. A few governing principles emerge from this analysis. Cortical connection types could be organized into a two-dimensional feature space, with synaptic strength and variability forming two orthogonal gradients. Excitatory connections are grouped along the variability axis based on their postsynaptic subclass: These connections have lower variability when connecting to excitatory cells, higher variability onto Pvalb cells, and the highest variability onto Sst cells. By contrast, inhibitory connections partition more naturally based on their presynaptic subclass. This partition again separates connections based on variability, with Pvalb connections having overall lower variability than other inhibitory types. In most cases, variability correlates with STP, and the highest variance connections are more likely to be facilitating. The rules derived from these relationships are similar to rules described previously (*47, 48, 74*). However, we found synaptic variability to be a better predictor of cell subclass than STP or strength, particularly for excitatory connections.

Synaptic variability has been regarded as an undesirable consequence of signaling through metabolically expensive exocytosis. More recently, advances in machine learning that rely on stochasticity have supported the possibility that synaptic variability may offer computational benefits, such as a mechanism for regularization during learning (*21*). If that is the case, then it is further plausible that variability may be modulated by cell type and that these relationships are crucial features of cortical function. Indeed, our measurements of synaptic variability strongly differentiated cell type in a pattern that is largely, but not entirely, aligned with STP metrics, suggesting the possibility of cell type–specific tuning of variability. More broadly, analysis of our stochastic release model indicates that synaptic strength and variability form the two most significant parameters describing synapse behavior.

REFERENCES AND NOTES

1. A. M. Thomson, A. P. Bannister, Interlaminar connections in the neocortex. *Cereb. Cortex* **13**, 5–14 (2003). doi: 10.1093/cercor/13.1.5; pmid: 12466210
2. H. Markram *et al.*, Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* **5**, 793–807 (2004). doi: 10.1038/nrn1519; pmid: 15378039
3. R. J. Douglas, K. A. C. Martin, Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004). pmid: 15217339
4. C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, M. Scanziani, Inhibition of inhibition in visual cortex: The logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**, 1068–1076 (2013). doi: 10.1038/nn.3446; pmid: 23817549

5. K. D. Harris, G. M. G. Shepherd, The neocortical circuit: Themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015). doi: 10.1038/nn.3917; pmid: 25622573
6. X. Jiang *et al.*, Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**, aac9462 (2015). doi: 10.1126/science.aac9462; pmid: 26612957
7. M. M. Karnani *et al.*, Cooperative subnetworks of molecularly similar interneurons in mouse neocortex. *Neuron* **90**, 86–100 (2016). doi: 10.1016/j.neuron.2016.02.037; pmid: 27021171
8. R. Tremblay, S. Lee, B. Rudy, GABAergic interneurons in the neocortex: From cellular properties to circuits. *Neuron* **91**, 260–292 (2016). doi: 10.1016/j.neuron.2016.06.033; pmid: 27477017
9. G. Molnár *et al.*, Complex events initiated by individual spikes in the human cerebral cortex. *PLOS Biol.* **6**, e222 (2008). doi: 10.1371/journal.pbio.0060222; pmid: 18767905
10. G. Molnár *et al.*, Human pyramidal to interneuron synapses are mediated by multi-vesicular release and multiple docked vesicles. *eLife* **5**, e18167 (2016). doi: 10.7554/eLife.18167; pmid: 27536876
11. Y. Peng *et al.*, High-throughput microcircuit analysis of individual human brains through next-generation multineuron patch-clamp. *eLife* **8**, e48178 (2019). doi: 10.7554/eLife.48178; pmid: 31742558
12. G. Testa-Silva *et al.*, High bandwidth synaptic communication and frequency tracking in human neocortex. *PLOS Biol.* **12**, e1002007 (2014). doi: 10.1371/journal.pbio.1002007; pmid: 25422947
13. S. C. Seeman *et al.*, Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *eLife* **7**, e37349 (2018). doi: 10.7554/eLife.37349; pmid: 30256194
14. V. Szegedi *et al.*, High-precision fast-spiking basket cell discharges during complex events in the human neocortex. *eNeuro* **4**, ENEURO.0260-17.2017 (2017). doi: 10.1523/eneuro.0260-17.2017; pmid: 29034319
15. Y. N. Billeh *et al.*, Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron* **106**, 388–403.e18 (2020). doi: 10.1016/j.neuron.2020.01.040; pmid: 32142648
16. A. Reyes *et al.*, Target-cell-specific facilitation and depression in neocortical circuits. *Nat. Neurosci.* **1**, 279–285 (1998). doi: 10.1038/1092; pmid: 10195160
17. A. V. Blackman, T. Abrahamsson, R. P. Costa, P. Lalanne, P. J. Sjöström, Target-cell-specific short-term plasticity in local circuits. *Front. Synaptic Neurosci.* **5**, 11 (2013). doi: 10.3389/fnsyn.2013.00011; pmid: 24367330
18. R. S. Larsen, P. J. Sjöström, Synapse-type-specific plasticity in local circuits. *Curr. Opin. Neurobiol.* **35**, 127–135 (2015). doi: 10.1016/j.conb.2015.08.001; pmid: 26310110
19. S. Lefort, C. C. H. Petersen, Layer-dependent short-term synaptic plasticity between excitatory neurons in the C2 barrel column of mouse primary somatosensory cortex. *Cereb. Cortex* **27**, 3869–3878 (2017). doi: 10.1093/cercor/bhx094; pmid: 28444185
20. H. Markram *et al.*, Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015). doi: 10.1016/j.cell.2015.09.029; pmid: 26451489
21. M. Llera-Montero, J. Sacramento, R. P. Costa, Computational roles of plastic probabilistic synapses. *Curr. Opin. Neurobiol.* **54**, 90–97 (2019). doi: 10.1016/j.conb.2018.09.002; pmid: 30308457
22. L. F. Abbott, W. G. Regehr, Synaptic computation. *Nature* **431**, 796–803 (2004). doi: 10.1038/nature03010; pmid: 15483601
23. D. Burnham, E. Shea-Brown, S. Mihalas, Learning to predict in networks with heterogeneous and dynamic synapses. bioRxiv 444107 [Preprint] (2021). doi: 10.1101/2021.05.18.444107
24. T. L. Daigle *et al.*, A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* **174**, 465–480.e22 (2018). doi: 10.1016/j.cell.2018.06.035; pmid: 30007418
25. R. B. Levy, A. D. Reyes, Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *J. Neurosci.* **32**, 5609–5619 (2012). doi: 10.1523/JNEUROSCI.5158-11.2012; pmid: 22514322
26. R. Perin, T. K. Berger, H. Markram, A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5419–5424 (2011). doi: 10.1073/pnas.1016051108; pmid: 21383177
27. C. Holmgren, T. Harkany, B. Svennenfors, Y. Zilberter, Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol.* **551**, 139–153 (2003). doi: 10.1113/jphysiol.2003.044784; pmid: 12813147

28. L. F. Rossi, K. D. Harris, M. Carandini, Spatial connectivity matches direction selectivity in visual cortex. *Nature* **588**, 648–652 (2020). doi: 10.1038/s41586-020-2894-4; pmid: 33177719
29. S. Song, P. J. Sjöström, M. Reigl, S. Nelson, D. B. Chklovskii, Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLOS Biol.* **3**, e68 (2005). doi: 10.1371/journal.pbio.0030068; pmid: 15737062
30. F. Z. Hoffmann, J. Triesch, Nonrandom network connectivity comes in pairs. *Netw. Neurosci.* **1**, 31–41 (2017). doi: 10.1162/NETN_a_00004; pmid: 29601066
31. A. Stepanyants, L. M. Martinez, A. S. Ferecskó, Z. F. Kisvárday, The fractions of short- and long-range connections in the visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3555–3560 (2009). doi: 10.1073/pnas.0810390106; pmid: 19221032
32. S. Lefort, C. Tomm, J. C. Floyd Sarria, C. C. Petersen, The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* **61**, 301–316 (2009). doi: 10.1016/j.neuron.2008.12.020; pmid: 19186171
33. J.-S. Jouhanneau, J. Kremkow, A. L. Dorrn, J. F. Poulet, In vivo monosynaptic excitatory transmission between layer 2 cortical pyramidal neurons. *Cell Rep.* **13**, 2098–2106 (2015). doi: 10.1016/j.celrep.2015.11.011; pmid: 26670044
34. S. P. Brown, S. Hestrin, Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* **457**, 1133–1136 (2009). doi: 10.1038/nature07658; pmid: 19151698
35. M. Morishima, K. Kobayashi, S. Kato, K. Kobayashi, Y. Kawaguchi, Segregated excitatory–inhibitory recurrent subnetworks in layer 5 of the rat frontal cortex. *Cereb. Cortex* **27**, 5846–5857 (2017). doi: 10.1093/cercor/bhx276; pmid: 29045559
36. A. Naka *et al.*, Complementary networks of cortical somatostatin interneurons enforce layer specific control. *eLife* **8**, e43696 (2019). doi: 10.7554/eLife.43696; pmid: 30883329
37. F. Scala *et al.*, Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas. *Nat. Commun.* **10**, 4174 (2019). doi: 10.1038/s41467-019-12058-z; pmid: 31519874
38. H. Hu, J. Z. Cavendish, A. Agmon, Not all that glitters is gold: Off-target recombination in the somatostatin-IRES-Cre mouse line labels a subset of fast-spiking interneurons. *Front. Neural Circuits* **7**, 195 (2013). doi: 10.3389/fncir.2013.00195; pmid: 24339803
39. B. R. Lee *et al.*, Scaled, high fidelity electrophysiological, morphological, and transcriptomic cell characterization. *eLife* **10**, e65482 (2021). doi: 10.7554/eLife.65482; pmid: 34387544
40. M. Galarreta, S. Hestrin, A network of fast-spiking cells in the neocortex connected by electrical synapses. *Nature* **402**, 72–75 (1999). doi: 10.1038/47029; pmid: 10573418
41. M. Galarreta, S. Hestrin, Electrical and chemical synapses among parvalbumin fast-spiking GABAergic interneurons in adult mouse neocortex. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12438–12443 (2002). doi: 10.1073/pnas.192159599; pmid: 12213962
42. P. Alcamí, A. E. Pereda, Beyond plasticity: The dynamic impact of electrical synapses on neural circuits. *Nat. Rev. Neurosci.* **20**, 253–271 (2019). doi: 10.1038/s41583-019-0133-5; pmid: 30824857
43. F. Walker *et al.*, Parvalbumin- and vasoactive intestinal polypeptide-expressing neocortical interneurons impose differential inhibition on Martinotti cells. *Nat. Commun.* **7**, 13664 (2016). doi: 10.1038/ncomms13664; pmid: 27897179
44. A. M. Thomson, D. C. West, Y. Wang, A. P. Bannister, Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2-5 of adult rat and cat neocortex: Triple intracellular recordings and biocytin labelling in vitro. *Cereb. Cortex* **12**, 936–953 (2002). doi: 10.1093/cercor/12.9.936; pmid: 12183393
45. A. Pala, C. C. H. Petersen, In vivo measurement of cell-type-specific synaptic connectivity and synaptic transmission in layer 2/3 mouse barrel cortex. *Neuron* **85**, 68–75 (2015). doi: 10.1016/j.neuron.2014.11.025; pmid: 25543458
46. A. M. Thomson, Activity-dependent properties of synaptic transmission at two classes of connections made by rat neocortical pyramidal axons in vitro. *J. Physiol.* **502**, 131–147 (1997). doi: 10.1111/j.1469-7793.1997.131bl.x; pmid: 9234202
47. M. Beierlein, J. R. Gibson, B. W. Connors, Two dynamically distinct inhibitory networks in layer 4 of the neocortex. *J. Neurophysiol.* **90**, 2987–3000 (2003). doi: 10.1152/jn.00283.2003; pmid: 12815025
48. Y. Ma, H. Hu, A. Agmon, Short-term plasticity of unitary inhibitory-to-inhibitory synapses depends on the presynaptic interneuron subtype. *J. Neurosci.* **32**, 983–988 (2012). doi: 10.1523/JNEUROSCI.5007-11.2012; pmid: 22262896

49. P. Balaram, J. H. Kaas, Towards a unified scheme of cortical lamination for primary visual cortex across primates: Insights from NeuN and VGLUT2 immunoreactivity. *Front. Neuroanat.* **8**, 81 (2014). doi: 10.3389/fnana.2014.00081; pmid: 25177277

50. J. Berg *et al.*, Human neocortical expansion involves glutamatergic neuron diversification. *Nature* **598**, 151–158 (2021). doi: 10.1038/s41586-021-03813-8; pmid: 34616067

51. B. E. Kalmbach *et al.*, h-channels contribute to divergent intrinsic membrane properties of supragranular pyramidal neurons in human versus mouse cerebral cortex. *Neuron* **100**, 1194–1208.e5 (2018). doi: 10.1016/j.neuron.2018.10.012; pmid: 30392798

52. A. D. Bird, M. J. Wall, M. J. E. Richardson, Bayesian inference of synaptic quantal parameters from correlated vesicle release. *Front. Comput. Neurosci.* **10**, 116 (2016). doi: 10.3389/fncom.2016.00116; pmid: 27932970

53. A. Barri, Y. Wang, D. Hansel, G. Mongillo, Quantifying repetitive transmission at chemical synapses: A generative-model approach. *eNeuro* **3**, ENEURO.0113-15.2016 (2016). doi: 10.1523/ENEURO.0113-15.2016; pmid: 27200414

54. E. Nanou, W. A. Catterall, Calcium channels, synaptic plasticity, and neuropsychiatric disease. *Neuron* **98**, 466–481 (2018). doi: 10.1016/j.neuron.2018.03.017; pmid: 29723500

55. H. Markram, Y. Wang, M. Tsodyks, Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5323–5328 (1998). doi: 10.1073/pnas.95.9.5323; pmid: 9560274

56. O. Bykowska *et al.*, Model-based inference of synaptic transmission. *Front. Synaptic Neurosci.* **11**, 21 (2019). doi: 10.3389/fnsyn.2019.00021; pmid: 31481887

57. A. J. Keller *et al.*, A Disinhibitory circuit for contextual modulation in primary visual cortex. *Neuron* **108**, 1181–1193.e8 (2020). doi: 10.1016/j.neuron.2020.11.013; pmid: 33301712

58. Y. Fu *et al.*, A cortical circuit for gain control by behavioral state. *Cell* **156**, 1139–1152 (2014). doi: 10.1016/j.cell.2014.01.050; pmid: 24630718

59. D. J. Millman *et al.*, VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli. *eLife* **9**, e55130 (2020). doi: 10.7554/eLife.55130; pmid: 33108272

60. M. Dipoppa *et al.*, Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *Neuron* **98**, 602–615.e8 (2018). doi: 10.1016/j.neuron.2018.03.037; pmid: 29656873

61. H. Hu, Y. Ma, A. Agmon, Submillisecond firing synchrony between different subtypes of cortical interneurons connected chemically but not electrically. *J. Neurosci.* **31**, 3351–3361 (2011). doi: 10.1523/JNEUROSCI.4881-10.2011; pmid: 21368047

62. J. R. Gibson, M. Beierlein, B. W. Connors, Two networks of electrically coupled inhibitory neurons in neocortex. *Nature* **402**, 75–79 (1999). doi: 10.1038/47035; pmid: 10573419

63. N. W. Gouwens *et al.*, Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell* **183**, 935–953.e19 (2020). doi: 10.1016/j.cell.2020.09.057; pmid: 33186530

64. C. M. Niell, M. P. Stryker, Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008). doi: 10.1523/JNEUROSCI.0623-08.2008; pmid: 18650330

65. Z. Álmási, C. Dávid, M. Witte, J. F. Staiger, Distribution patterns of three molecularly defined classes of GABAergic neurons across columnar compartments in mouse barrel cortex. *Front. Neuroanat.* **13**, 45 (2019). doi: 10.3389/fnana.2019.00045; pmid: 31114486

66. Y. Gonchar, Q. Wang, A. Burkhalter, Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining. *Front. Neuroanat.* **1**, 3 (2008). doi: 10.3389/neuro.05.003.2007; pmid: 18958197

67. G. Lur, M. A. Vinck, L. Tang, J. A. Cardin, M. J. Higley, Projection-specific visual feature encoding by layer 5 cortical subnetworks. *Cell Rep.* **14**, 2538–2545 (2016). doi: 10.1016/j.celrep.2016.02.050; pmid: 26972011

68. E. J. Kim, A. L. Juavinett, E. M. Kyubwa, M. W. Jacobs, E. M. Callaway, Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. *Neuron* **88**, 1253–1267 (2015). doi: 10.1016/j.neuron.2015.11.002; pmid: 26671462

69. F. Scala *et al.*, Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021). doi: 10.1038/s41586-020-2907-3; pmid: 33184512

70. R. Azouz, C. M. Gray, Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8110–8115 (2000). doi: 10.1073/pnas.130200797; pmid: 10859358

71. S. J. Barnes *et al.*, Delayed and temporally imprecise neurotransmission in reorganizing cortical microcircuits. *J. Neurosci.* **35**, 9024–9037 (2015). doi: 10.1523/JNEUROSCI.4583-14.2015; pmid: 26085628

72. J. Obermayer *et al.*, Lateral inhibition by Martinotti interneurons is facilitated by cholinergic inputs in human and mouse neocortex. *Nat. Commun.* **9**, 4101 (2018). doi: 10.1038/s41467-018-06628-w; pmid: 30291244

73. G. Silberberg, H. Markram, Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. *Neuron* **53**, 735–746 (2007). doi: 10.1016/j.neuron.2007.02.012; pmid: 17329212

74. A. Gupta, Y. Wang, H. Markram, Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science* **287**, 273–278 (2000). doi: 10.1126/science.287.5451.273; pmid: 10634775

Where Science Gets Social.

AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
▲AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

RESEARCH ARTICLE

GLOBAL PRODUCTIVITY

Global biosphere primary productivity changes during the past eight glacial cycles

Ji-Woong Yang[1]*†, Margaux Brandon[1,2]‡, Amaëlle Landais[1], Stéphanie Duchamp-Alphonse[2], Thomas Blunier[3], Frédéric Prié[1], Thomas Extier[1]§

Global biosphere productivity is the largest uptake flux of atmospheric carbon dioxide (CO_2), and it plays an important role in past and future carbon cycles. However, global estimation of biosphere productivity remains a challenge. Using the ancient air enclosed in polar ice cores, we present the first 800,000-year record of triple isotopic ratios of atmospheric oxygen, which reflects past global biosphere productivity. We observe that global biosphere productivity in the past eight glacial intervals was lower than that in the preindustrial era and that, in most cases, it starts to increase millennia before deglaciations. Both variations occur concomitantly with CO_2 changes, implying a dominant control of CO_2 on global biosphere productivity that supports a pervasive negative feedback under the glacial climate.

Atmospheric carbon dioxide (CO_2) is a potent greenhouse gas that, together with orbital changes, is a primary determinant of Earth's global climate. Measurements of the CO_2 mixing ratio of air trapped in ice cores over the past 800 thousand years (ka) reveal clear glacial-interglacial cycles (*1*, *2*) and show a good correlation with global sea-level changes (*3*). The Southern Ocean (SO) is thought to have played a major role in these CO_2 variations through changes in sea-ice cover, overturning circulation, and biological productivity (*4*–*7*), and there is growing evidence that the terrestrial vegetation may be an important contributor as well [e.g., (*8*)]. Nevertheless, there are periods during which CO_2 concentration decouples from sea level, particularly during full-glacial periods when CO_2 shows relatively stable or slightly rising trends while sea level continues to decline (*9*, *10*).

To explain this CO_2 evolution during full-glacial times, Galbraith and Eggleston (*10*) hypothesized a negative feedback by which global photosynthesis becomes limited by low atmospheric CO_2 concentrations, prohibiting further CO_2 drawdown. This hypothesis has not been proven yet by observations because the reconstructions of past biosphere produc-tivity are based on geochemical (e.g., organic and inorganic biomarkers) and micropaleontological (e.g., pollen, coccolith, diatoms) data from sediment archives, which provide indirect and only qualitative reconstructions [e.g., (*11*)]. Furthermore, they sometimes show contrasting changes, making it difficult to estimate global variations. As an example, terrestrial vegetation records (e.g., arboreal pollen fraction) show a drastic decrease during glacial times, whereas phytoplankton indicators (mainly total organic carbon and alkenones) in the Subantarctic Zone (SAZ), a key part of the SO (*6*, *12*), suggest intensified marine productivity.

The above limitations can be alleviated by using measurements of the triple isotope composition (^{16}O, ^{17}O, and ^{18}O) of atmospheric oxygen (O_2), a marker of global gross productivity expressed in O_2 flux (*13*). The triple isotopic composition of O_2 is primarily affected by O-isotope exchange with CO_2 during photochemical reactions in the stratosphere and biological reactions of photosynthesis and respiration (*13*, *14*). The heavy isotopes (^{17}O and ^{18}O) are discriminated relative to the light one (^{16}O) in a mass-dependent way during most biological reactions. On the contrary, O_2-CO_2 isotope exchange in the stratosphere fractionates in a mass-independent manner (*13*–*15*). Accordingly, to estimate the relative contribution of biosphere and stratosphere fluxes, the ^{17}O anomaly of O_2 is defined as $^{17}\Delta \equiv \ln(\delta^{17}O + 1) - \lambda_{ref} \cdot \ln(\delta^{18}O + 1)$ (*16*), where λ_{ref} is the mass-dependent reference slope of 0.516 calculated from the modern O_2 isotope fractionations within the global biospheric cycle (*16*, *17*). By definition, tropospheric $^{17}\Delta$ is not modified much by the mass-dependent fractionation within the biosphere. Closed terrarium experiments showed that without stratospheric exchange, biospheric processes induce a positive $^{17}\Delta$ signal relative to that of the present atmosphere (*13*). By contrast, the stratospheric air measurements from rocket (*15*), aircraft [e.g., (*18*, *19*)], and balloons [e.g., (*19*, *20*)] showed a highly positive $^{17}\Delta$ of CO_2 as a result of photolysis of O_2 and ozone (O_3) and atomic O exchange with CO_2 (*21*), which is counterbalanced by a slight depletion of $^{17}\Delta$ of the O_2 reservoir in the stratosphere ($^{17}\Delta_{strat}$) because of the small abundance (mixing ratio) of O_3 relative to O_2 (*22*). In the present atmosphere, the input of positively fractionated $^{17}\Delta$ from the biosphere ($^{17}\Delta_{bio}$) is equilibrated by a massive flux of O_2 with slightly negative $^{17}\Delta$ from the stratosphere. The sizes of the two endmember fluxes from the biosphere and the stratosphere hence drive the $^{17}\Delta$ variations, so by knowing the magnitudes of the O_2 isotopic fractionations in the stratosphere and O_2 fluxes from the stratosphere (F_{strat}), $^{17}\Delta$ can be used to reconstruct the gross O_2 flux from Earth's biosphere (F_{bio}), or global gross primary productivity in terms of O_2 (GPP-O_2).

The stratosphere influences tropospheric $^{17}\Delta$ through different factors, including F_{strat}, stratosphere temperature, photochemical reaction rates, and O_3 abundance in the stratosphere. Numerical simulations show that greenhouse gases (predominantly CO_2) play an important role in controlling the above-mentioned changes in the stratosphere [e.g., (*23*–*26*)], allowing the stratosphere effect to be scaled to CO_2 changes such that a high CO_2 mixing ratio in the troposphere induces a strong depletion in $^{17}\Delta$ [see (*27*) for more details]. It should be noted that O_2 photolysis in the mesosphere may cause additional fractionation (*28*). However, model simulations suggest that the entrainment of mesospheric air into the stratosphere is small (~0.02%) (*28*), and its effect on the tropospheric $^{17}\Delta$ is expected to be negligible (*27*).

As a consequence, the imprint of the stratospheric mass-independent fractionation is reflected in the general anticorrelation between CO_2 and $^{17}\Delta$. However, CO_2 and $^{17}\Delta$ are not always anticorrelated because they decouple when changes in biosphere productivity occur. For example, Brandon *et al.* (*8*) pointed out that the notable decoupling over the Termination V (TV)–Marine Isotope Stage (MIS) 11 interval may be interpreted as an imprint of exceptionally high biosphere productivity. Therefore, the decoupling between CO_2 and $^{17}\Delta$ is the key to inferring the past evolution of global biosphere productivity, which provides distinctive insights into the past global carbon cycle.

The previous ice-core record of $^{17}\Delta$ extends over the past 444.8 ka (*8*, *29*). During the past four glacial cycles, which were characterized by a large glacial-interglacial amplitude and a periodicity of ~100 ka, the inferred global
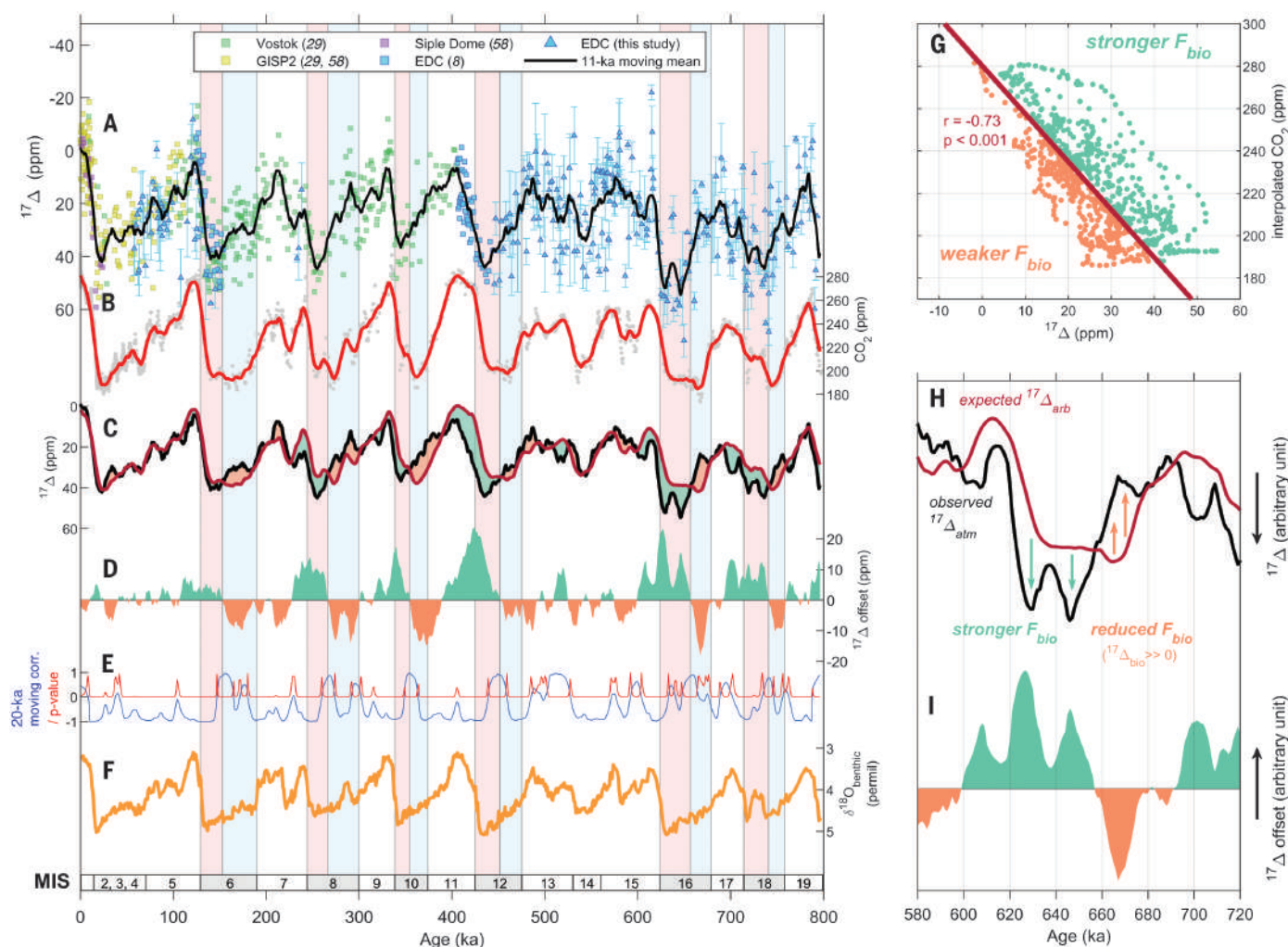
Fig. 1. The 800-ka ice-core composite of $^{17}\Delta$ records. (**A**) Compilation of $^{17}\Delta$ values for multiple ice-core records with previously published data (8, 29, 58). The new EDC data produced in this study are plotted in blue triangles with 1σ uncertainty ranges. The smoothed curve of the compilation record is shown as a solid black line. (**B**) Ice-core CO_2 compilation (gray dots) and the smoothed curve (red) (1, 2). (**C**) Smoothed $^{17}\Delta$ (black) and $^{17}\Delta_{arb}$ time series (red). (**D**) $^{17}\Delta$ offset between the two time series in (C). (**E**) The 20-ka moving correlation coefficient (blue) and p values (red) between the smoothed $^{17}\Delta$ and $^{17}\Delta_{arb}$ time series in (C). (**F**) LR04 benthic $\delta^{18}O$ stack (59). The blue and red shadings denote the negative and positive offsets, respectively, during glacial periods within the even-numbered MIS stages as defined by (59). permil, parts per thousand. (**G**) Scatter plot of $^{17}\Delta$ and CO_2, both smoothed by an 11-ka moving average (fig. S2). The CO_2 composite data in (B) are interpolated to the ages of the $^{17}\Delta$ data. The linear regression of CO_2 to $^{17}\Delta_{arb}$ is shown in dark red. At given CO_2 concentrations, scatter points to the right (green) and left (orange) of the regression line imply enhanced and reduced F_{bio}, respectively. (**H** and **I**) The enlarged view of (C) and (D) over the MIS 16 interval. All of the plots are based on, or transferred to, Antarctic ice core chronology (AICC) 2012 (27, 60, 61). GISP2, Greenland Ice Sheet Project 2.

biosphere productivity was systematically lower during glacial than during interglacial times (29). The amplitude of temperature and CO_2 changes over glacial-interglacial transitions was smaller before 450 ka and the glacial periods were shorter than after 450 ka (2). Knowing the global biosphere productivity over the period of 800 to 450 ka is hence of uttermost importance to studying the interactions between Earth's biosphere and CO_2 level during the glacial-interglacial cycles. Here, we extend the $^{17}\Delta$ records back to ~796 ka by analyzing samples from the European Project for Ice Coring in Antarctica Dome C (EDC) ice core (27). The data were corrected for the

fractionations by gas loss during ice storage, gravitational settling, and bubble close-off (27).

Decoupling between $^{17}\Delta$ and CO_2

The fully corrected EDC $^{17}\Delta$ data are plotted with CO_2 in Fig. 1. We observe a general anti-correlation between CO_2 and $^{17}\Delta$ over the past 796 ka [correlation coefficient (r) = −0.73; Fig. 1G and fig. S2]. However, this anticorrelation does not hold true during glacial intervals (Fig. 1, A to C): The moving correlation coefficients between CO_2 and $^{17}\Delta$ show a significant positive correlation in the middle of the glacial intervals of MISs 6, 8, 10, 12, 16, and 18 (Fig. 1E). Indeed, CO_2 concentrations show

stable or slightly increasing trends, whereas $^{17}\Delta$ signals gradually increase to glacial maxima (Fig. 1C). We illustrate this decoupling between $^{17}\Delta$ and CO_2 through the $^{17}\Delta$ offset (Fig. 1D), which is the difference between the ice-core $^{17}\Delta$ and a hypothetical $^{17}\Delta$ ($^{17}\Delta_{arb}$) purely driven by stratospheric fractionation (Fig. 1C and fig. S3) (30). Negative $^{17}\Delta$ offsets (orange in Fig. 1, D, G, and I) result from reduced biosphere productivity, whereas positive ones (green in Fig. 1, D, G, and I) can be explained by enhanced biosphere productivity. The negative offsets (reduced productivity) are found at earlier stages of glacial periods, whereas the positive ones (increased
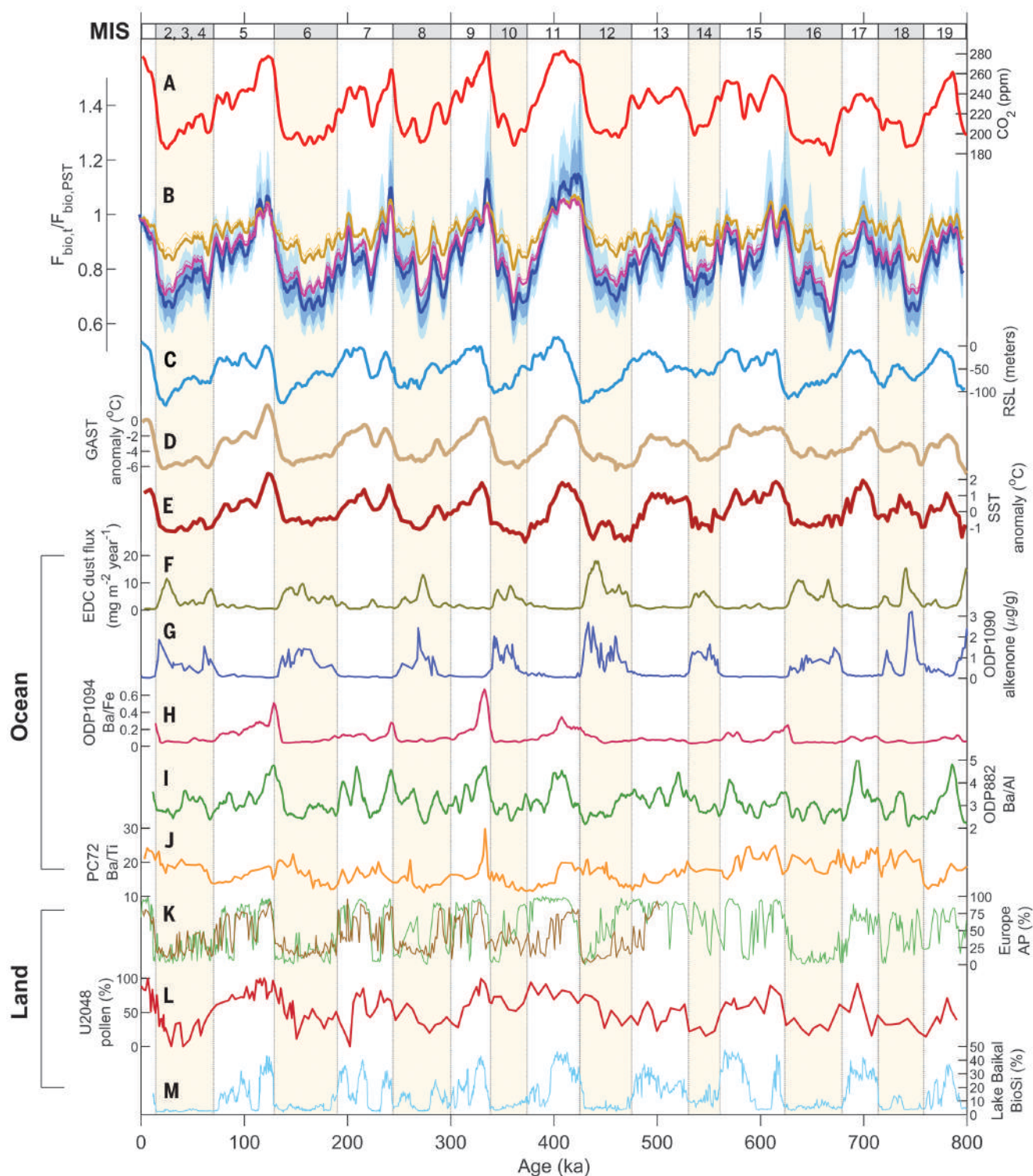
Fig. 2. Comparison of the reconstructed GPP-O₂ with global and regional paleoclimate records. (**A**) Ice-core CO_2 composite (*1, 2*) smoothed by a 5-ka moving average. (**B**) GPP-O₂ reconstructed by the AL (blue) and TB (pink and yellow) models. For the AL model, the dark and light blue shadings represent 68 and 95% ranges of Monte-Carlo sensitivity solutions, respectively (*27*). The TB model solutions for a LGM C4 plant contribution of 0.7 (pink) and 0.4 (yellow) are plotted with corresponding curves with the Holocene-LGM ^{17}O anomaly offset of 10 ppm (thin pink and thin yellow lines, respectively). All of the model reconstructions are smoothed by a 5-ka moving average. (**C**) RSL reconstruction (*53*). (**D**) GAST anomaly from present (0 to 5 ka) (*52*). (**E**) Global SST stack (*9*). (**F**) EDC dust flux (*62*) smoothed by a 5-ka moving mean. (**G**) Alkenone concentration from the ODP 1090 core (*6*). (**H**) Ba/Fe ratio from the ODP 1094 core (*18*) smoothed by a 5-ka moving average. (**I**) Ba/Al ratio at the ODP 882 site (*33*) smoothed by a 5-ka moving average. (**J**) Ba/Ti ratio from the TT013-PC72 core (*34*). (**K**) Arboreal pollen (AP) fraction from Lake Ohrid (brown) (*41*) and Tenaghi Philippon (dark green) (*40*). (**L**) Fractional abundances of the pollen end-members at the MD96-2048 core, southeastern Africa (*42*). (**M**) Biogenic silica from Lake Baikal (*43*). Yellow shadings indicate the glacial periods.
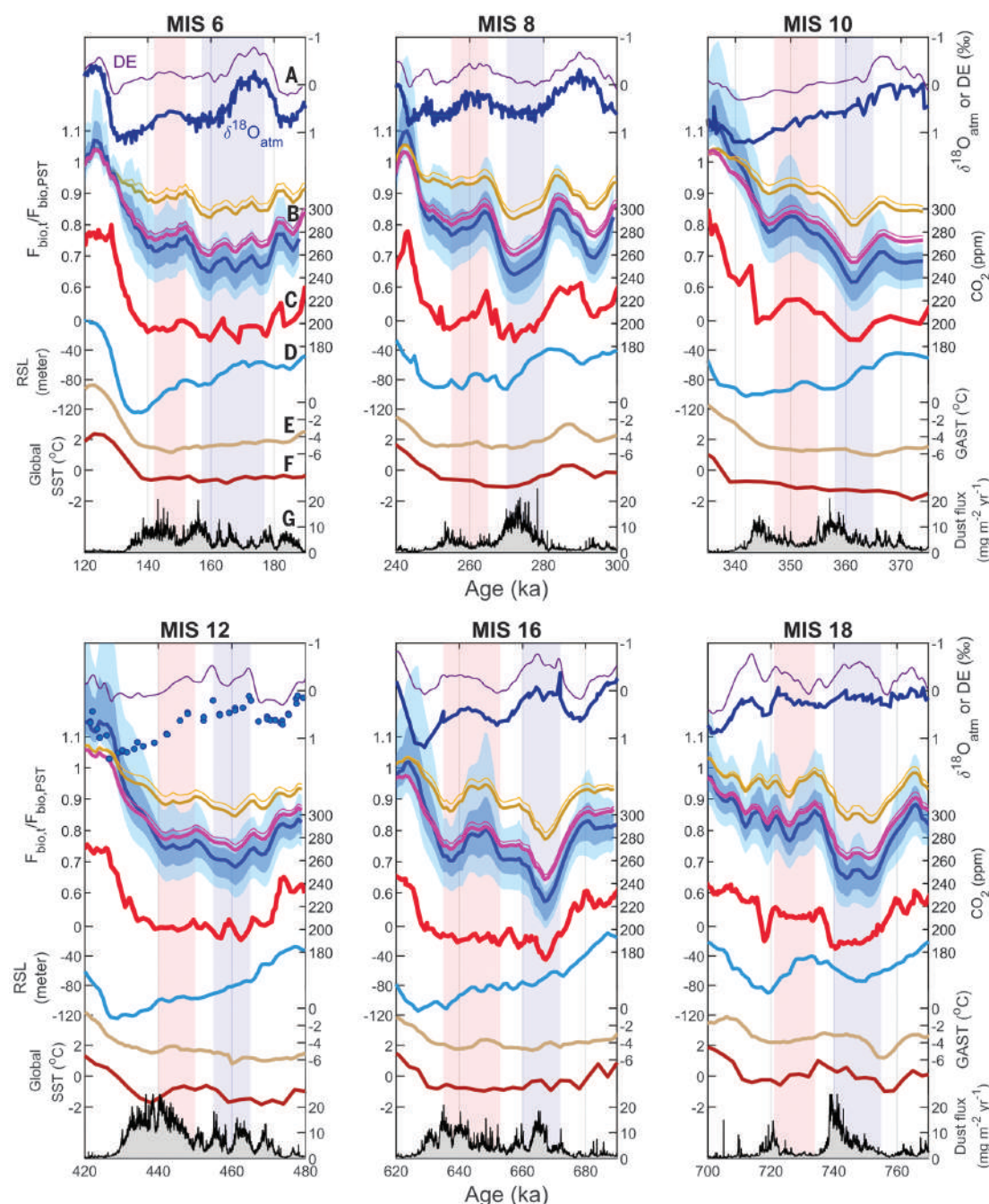
Fig. 3. Close-up of GPP-O_2 evolution during glacial intervals with different climate proxies. (**A**) Ice-core $\delta^{18}O_{atm}$ composite (dark blue) (63) and DE (purple) (56). (**B**) GPP-O_2 reconstructions using the AL and TB models in the same color schemes as in Fig. 2. (**C**) Ice-core CO_2 composite (1, 2). (**D**) RSL reconstruction (53). (**E**) GAST anomaly from present (0 to 5 ka) (52). (**F**) Global SST stack (9). (**G**) EDC dust flux (62). Blue and red shadings indicate the mid- and full-glacial stages, respectively.

productivity) prevail from later stages of glacial to interglacial periods.

GPP-O_2 reconstructions using box models

To obtain a more quantitative assessment of past changes in global biosphere productivity, we applied two different box models [the T. Blunier (TB) (29) and A. Landais (AL) models (31)] that describe triple O_2 isotope budgets in the biosphere, troposphere, and stratosphere. Assuming steady state, both models calculate the biosphere O_2 flux so that the biosphere input of $^{17}\Delta$ is balanced by O_2 flux

from the stratosphere, which is estimated from the CO_2 concentration in the troposphere (13). The different assumptions made by the two models and their limitations are described in the supplementary materials (27).

The reconstructed GPP-O_2 from the two different models are presented in Fig. 2B in terms of the ratio between the global biosphere O_2 flux of the past ($F_{bio,t}$) and the global biosphere O_2 flux of the preindustrial condition ($F_{bio,PST}$). Although the two model results are not necessarily identical, they share common features

that confirm our qualitative inferences based on $^{17}\Delta$ offset (Fig. 1): (i) They reveal clear glacial (low productivity)–interglacial (high productivity) cycles during the past 796 ka, (ii) the GPP-O_2 minima during each glacial period occurred under intermediate sea level and concomitantly with glacial CO_2 minima (mid-glacial stage) (Fig. 2, A and B), and (iii) following the glacial minima, GPP-O_2 increases while global ice volume continues to grow until the glacial maxima (full-glacial stage), millennia before the glacial terminations (Figs. 2B and 3B).
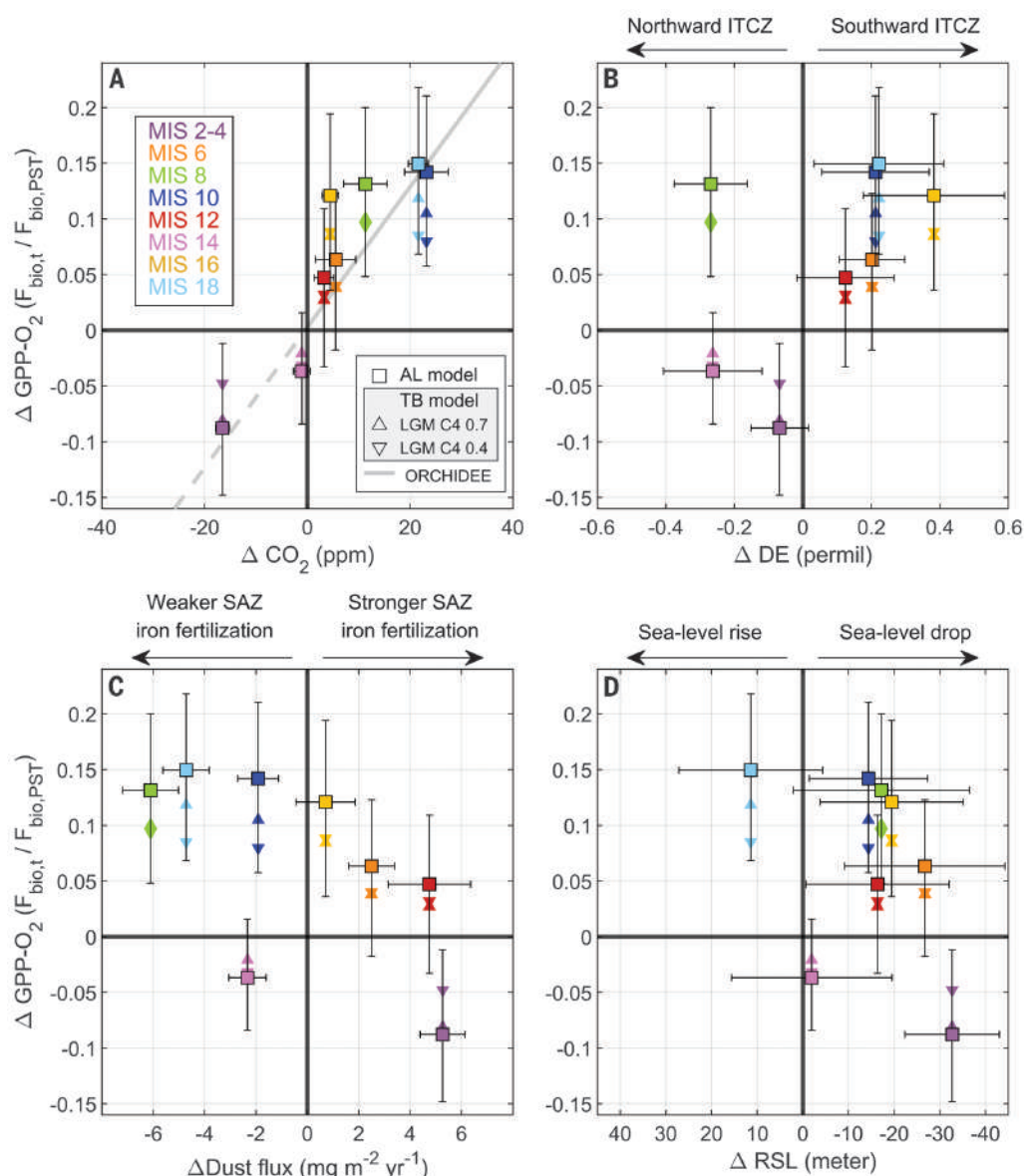
Fig. 4. Comparison of the magnitude of changes in GPP-O₂ and potential climatic controls between the two glacial stages. The differences (expressed in Δ) between the average values of the two stages (full-glacial minus mid-glacial stages) as defined in Fig. 3. Both ΔGPP-O₂ results from the AL (square) and TB (triangle) models are plotted. (**A**) Comparison with CO_2 changes. The solid gray line stands for the ΔGPP-O₂ predicted by second-order regression of CO_2-sensitivity simulations using the ORCHIDEE model (*44*), assuming constant ocean productivity and terrestrial-to-marine GPP ratio. The ΔGPP-O₂ uncertainties are taken as 16 to 84 percentiles of the Monte-Carlo simulations (n = 1000) of 26 sensitivity scenarios. CO_2 uncertainties are taken from (*2*). (**B**) Comparison with DE changes (*56*). The positive changes indicate southward displacement of the tropical rainbelt. The uncertainties for ΔDE are based on the 1σ errors reported in (*56*). (**C**) Comparison with changes in EDC dust flux (*62*). The positive offset implies stronger iron fertilization in the SO during full-glacial stages compared with mid-glacial stages. ΔDust flux uncertainty is estimated by assuming the maximum analytical error reported as 10% (*62*). (**D**) Comparison with RSL changes (*53*). The ΔRSL uncertainty is based on the 1σ uncertainty ranges in (*53*).

Glacial-interglacial changes

Both box-model results indicate concomitant GPP-O₂ and CO_2 minima during each of the eight glacial intervals (Fig. 2), with glacial productivity minima that are estimated to lie between 55 and 87% of modern productivity. In the following sections, we explore the possible contributions of marine and terrestrial productivity to this large reduction of global GPP-O₂ under glacial times.

The only global compilation of marine sedimentary records of marine productivity that exist for glacial times focuses on the Last Glacial Maximum (LGM) and indicates stronger export production and probably higher marine primary productivity (PP) during the LGM than during the Holocene (*12*). Over the past 800 ka, the very few available marine export production records that are suggested

to reflect PP reveal spatially different patterns during glacial times. The alkenone concentration at the Ocean Drilling Program (ODP) 1090 site located in the SAZ shows an increased PP during glacial periods (*6*), whereas biogenic barium (Ba/Fe) from ODP 1094 in the Antarctic Zone of the SO (*32*) and the Ba/Al ratio from the ODP 882 core in the subarctic Pacific (*33*) indicate reduced PP (Fig. 2, G to I). In addition, the Ba/Ti record from the core TT013-PC72 in the Equatorial Pacific exhibits no clear glacial-interglacial pattern (*34*) (Fig. 2J). None of those records explain our reconstructed lower global GPP-O₂ during glacial periods. Global ocean PP in glacial times is difficult to predict, and ocean biogeochemical models have shown contrasting results: Some predict greater global PP in the LGM (*35*), whereas others represent the opposite (*36*)

or no clear change (*37*). On the one hand, a decrease in ocean PP is expected because colder sea surface temperature (SST) reduces the metabolic rate of marine phytoplankton [e.g., (*38*)], shoaling of overturning circulation reduces the intake of nutrient-rich deep water (*39*), and highly productive continental shelf areas are lost by sea-level decline. On the other hand, an increase in ocean PP could arise from greatly enhanced aeolian dust deposition that supplies iron to high-nutrient low-chlorophyll oceans such as the SO [e.g., (*6*)].

No direct proxies for terrestrial GPP exist, but we do have indirect vegetation cover proxies, some of which reflect glacial shrinking and interglacial expansion patterns. Two long-term pollen assembly records in Europe—Tenaghi Philippon (*40*) and Lake Ohrid (*41*)—clearly indicate a near vanishing arboreal-type pollen

during glacial periods (Fig. 2K). In addition, woodland- and mountain forest–type vegetations from the MD96-2048 core off southeastern Africa exhibit glacial-interglacial variations (42) (Fig. 2L). Similar glacial-interglacial patterns are observed in biogenic silica records from the BDP-96 core from Lake Baikal (43) (Fig. 2M).

Several arguments favor a stronger decrease in terrestrial GPP than in marine PP during glacial intervals. A nearly 90 parts per million (ppm) lower concentration of CO_2 in the glacial atmosphere would have a strong negative fertilizing effect on photosynthesis of terrestrial vegetation (44, 45), whereas its impact on marine PP is expected to be minor (46). A modeling study using a dynamic global vegetation model (DGVM) shows that with a CO_2 increase of 185 to 285 ppm, the vegetation GPP increases by more than a factor of two GPP (44). The effects of temperature and precipitation on terrestrial GPP are less clear. Although modern observations suggest that temperature and precipitation are important controls for terrestrial GPP [e.g., (47)], model studies show that climate change alone (without changes in CO_2) has a limited impact on glacial-interglacial vegetation GPP changes (45, 48). A glacial sea-level low stand might have two opposing effects, with the exposure of continental shelves promoting increased amounts of new vegetation (49, 50) and ice-sheet expansion prohibiting photosynthesis in the ice-covered area. Previous DGVM simulations suggested that the relative sea-level (RSL) decline would increase LGM vegetation net primary productivity (NPP) by ~8% relative to constant RSL results (51). Taken together, it is therefore likely that the glacial-interglacial changes in global GPP-O_2 are largely driven by terrestrial productivity, whose evolution is strongly influenced by CO_2.

Glacial productivity changes

The glacial productivity minima occur at mid-glacial stages, followed by the systematic increase in GPP-O_2 from mid- to full-glacial stages. Such an increase in productivity between mid- and full-glacial stages is not easy to detect from paleoproductivity proxies because of a lack of high-resolution records and/or chronological issues. In the ocean, neither ODP 1090 alkenone concentrations nor ODP 882 Ba/Al records exhibit a detectable shift between the two stages, with the exceptions of MISs 6 and 12 (ODP 1090) and MIS 8 (ODP 882) (Fig. 2, G and I). In parallel, the TT-13-PC72 Ba/Ti record starts to rise several millennia before maximum productivity at deglacial terminations (Fig. 2J). In the terrestrial realm, the arboreal pollen fractions at Tenaghi Phillippon and Lake Ohrid and biogenic silica at Lake Baikal indicate nearly vanished productivity during glacial times, showing no clear trends during the mid- to full-glacial changes (Fig. 2,

K and M). By contrast, the MD96-2048 pollen assemblage records show that the fractions of woodland and mountain forest species increased in the late-glacial stages during certain glacial intervals such as MISs 6, 8, 10, and 12 (Fig. 2L). Therefore, despite the complexity of comparing our GPP-O_2 records with local records, the GPP-O_2 increase between mid- and full-glacial stages is confirmed by some local records.

There are several ways in which the possible contributions of the main potential drivers of GPP-O_2 might help to explain the increasing GPP-O_2 signal between full- and mid-glacial stages. As mentioned previously, the CO_2 fertilization effect may play a major role, and there is a clear correlation between GPP-O_2 and CO_2 changes over the mid- to full-glacial stages (Fig. 4). The observed GPP-O_2 versus CO_2 changes agree with a second-order fitting of Chen et al.'s (44) sensitivity simulations (Fig. 4A). However, over MISs 8 and 16, the GPP-O_2 versus CO_2 relationship is slightly different, suggesting that factors other than CO_2 concentration affect the GPP-O_2 evolution during glacial periods (Fig. 4A).

First, the global temperature change between mid- and full-glacial stages is minor. The global air surface temperature (GAST) reconstructions (52) and a global SST stack (9) show no warming or only limited changes (Fig. 3).

Second, RSL reconstructions indicate sea-level declines of 10 to 30 m between mid- and full-glacial stages in most glacial periods (53) (Fig. 3). Recalling that there was an ~8% increase in terrestrial NPP at the LGM for an RSL change of ~120 m (51, 53), the sea-level decline between two glacial stages should have only a small effect. Moreover, the comparison between GPP-O_2 and RSL records shows no strong relationship (Fig. 4D). Therefore, we suggest that the sea-level changes have a minor impact on glacial GPP-O_2 changes.

Third, low-latitude hydrological changes could play a role as well, because global vegetation is expected to be shifted southerly during glacial periods owing to Northern Hemisphere ice-sheet expansion [e.g., (54)]. Climate model experiments for the LGM show an increase in precipitation in Southern Hemisphere low-latitude regions, such as Amazonia and South Africa, for which modeled vegetation NPP is increased in tropical and temperate forests as compared to present (45). Proxies for past evolution of the low-latitude water cycle exist, such as the Dole effect (DE), that is, the $\delta^{18}O$ offset between air O_2 and sea water (14), and speleothem $\delta^{18}O_{calcite}$ from Chinese caves (55). The long DE records (56) indicate a mid- to full-glacial enrichment over most glacial periods when GPP-O_2 increases, which implies a further southward shift of the Intertropical Convergence Zone (ITCZ) from

mid- to full-glacial stages (Fig. 4B). This is supported by $\delta^{18}O_{calcite}$ records from Chinese caves over MISs 6 and 10 (55). Therefore, this evidence suggests that a southward shift of the ITCZ at the full-glacial stages might have stimulated terrestrial GPP-O_2.

Finally, marine productivity may also contribute to this GPP-O_2 increase, especially in the SO, where PP is usually limited by iron [e.g., (57)]. This iron limitation is alleviated by greater dust deposition into the SAZ, which is caused by stronger winds, together with glacial aridity and the meridional shift of westerly winds (6). However, the changes in EDC dust flux are not systematically positively correlated with GPP-O_2 increases (Fig. 4C): The full-glacial increases in EDC dust flux are observed only in MISs 6, 12, and 16, where the GPP-O_2 increases are relatively small (Fig. 4C).

Summary

Our $^{17}\Delta$ data provide a complete view of global biosphere productivity evolution during the past 800 ka, which confirms the pervasive glacial (low GPP)–interglacial (high GPP) cycles, and demonstrate an important feature of intraglacial GPP shift. The GPP reconstructions and proxy evidence discussed here suggest that much of the GPP-O_2 changes over glacial-interglacial cycles and between two glacial stages are dominantly attributable to CO_2 changes and that, in both cases, terrestrial GPP might have played an important role. Our findings also demonstrate the close interactions of global photosynthesis with CO_2 over the past 800 ka, providing observational evidence of the pervasive negative feedback between global photosynthesis and CO_2 (10).

REFERENCES AND NOTES

1. D. Lüthi et al., Nature 453, 379–382 (2008).
2. B. Bereiter et al., Geophys. Res. Lett. 42, 542–549 (2015).
3. G. L. Foster, E. J. Rohling, Proc. Natl. Acad. Sci. U.S.A. 110, 1209–1214 (2013).
4. B. B. Stephens, R. F. Keeling, Nature 404, 171–174 (2000).
5. D. M. Sigman, M. P. Hain, G. H. Haug, Nature 466, 47–55 (2010).
6. A. Martínez-Garcia et al., Paleoceanography 24, PA1207 (2009).
7. S. Duchamp-Alphonse et al., Nat. Commun. 9, 2396 (2018).
8. M. Brandon et al., Nat. Commun. 11, 2112 (2020).
9. J. D. Shakun, D. W. Lea, L. E. Lisiecki, M. E. Raymo, Earth Planet. Sci. Lett. 426, 58–68 (2015).
10. E. D. Galbraith, S. Eggleston, Nat. Geosci. 10, 295–298 (2017).
11. K. B. Averyt, A. Paytan, Paleoceanography 19, PA4003 (2004).
12. K. E. Kohfeld, C. Le Quéré, S. P. Harrison, R. F. Anderson, Science 308, 74–78 (2005).
13. B. Luz, E. Barkan, M. L. Bender, M. H. Thiemens, K. A. Boering, Nature 400, 547–550 (1999).
14. M. Bender, T. Sowers, L. Labeyrie, Global Biogeochem. Cycles 8, 363–376 (1994).
15. M. H. Thiemens, T. Jackson, E. C. Jipf, P. W. Erdman, C. van Egmond, Science 270, 969–972 (1995).
16. B. Luz, E. Barkan, Geochim. Cosmochim. Acta 69, 1099–1110 (2005).
17. Recent laboratory incubation studies suggested either a lower λ of ~0.510 (64) or a higher λ of ~0.522 (65) during dark respiration; however, sensitivity tests performed with a λ of 0.510 and 0.520 lead to $^{17}\Delta$ variations within the uncertainty range. See (27) for more details.
18. K. A. Boering et al., Geophys. Res. Lett. 31, L03109 (2004).
19. A. A. Wiegel et al., Proc. Natl. Acad. Sci. U.S.A. 110, 17680–17685 (2013).
20. B. Alexander, M. K. Vollmer, T. Jackson, R. F. Weiss, M. H. Thiemens, Geophys. Res. Lett. 28, 4103–4106 (2001).

21. Y. L. Yung, W. B. DeMore, J. P. Pinto, *Geophys. Res. Lett.* **18**, 13–16 (1991).
22. M. H. Thiemens, M. Lin, *Rev. Mineral. Geochem.* **86**, 35–95 (2021).
23. N. Butchart *et al.*, *Clim. Dyn.* **27**, 727–741 (2006).
24. K. P. Shine *et al.*, *Q. J. R. Meteorol. Soc.* **129**, 1565–1588 (2003).
25. G. Chiodo *et al.*, *J. Clim.* **31**, 3893–3907 (2018).
26. E. L. Fleming, C. H. Jackman, R. S. Stolarski, A. R. Douglas, *Atmos. Chem. Phys.* **11**, 8515–8541 (2011).
27. Materials and methods are available as supplementary materials.
28. M.-C. Liang, G. A. Blake, B. R. Lewis, Y. L. Yung, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 21–25 (2007).
29. T. Blunier, M. L. Bender, B. Barnett, J. C. von Fischer, *Clim. Past* **8**, 1509–1526 (2012).
30. $^{17}\Delta_{arb}$ is the expected $^{17}\Delta$ in the troposphere if the biosphere (F_{bio} and $^{17}\Delta_{bio}$) remains constant over time; hence, it is hypothesized to be affected by the stratosphere only. The stratosphere changes are assumed to be limited by CO_2 (*13, 31*). $^{17}\Delta_{arb}$ is therefore calculated as tropospheric $^{17}\Delta$ in the steady-state equations S2 and S4 in the three-box model (AL model) by assuming constant F_{bio} and $^{17}\Delta_{bio}$ [see (*27*) for more details].
31. A. Landais, J. Lathiere, E. Barkan, B. Luz, *Global Biogeochem. Cycles* **21**, GB1025 (2007).
32. S. L. Jaccard *et al.*, *Science* **339**, 1419–1423 (2013).
33. S. L. Jaccard, E. D. Galbraith, D. M. Sigman, G. H. Haug, *Quat. Sci. Rev.* **29**, 206–212 (2010).
34. R. W. Murray, C. Knowlton, M. Leinen, A. C. Mix, C. H. Polsky, *Paleoceanography* **15**, 570–592 (2000).
35. V. Brovkin, A. Ganopolski, D. Archer, S. Rahmstorf, *Paleoceanography* **22**, PA4202 (2007).
36. A. Tagliabue *et al.*, *Clim. Past* **5**, 695–706 (2009).
37. L. Bopp, K. E. Kohfeld, C. Le Quéré, O. Aumont, *Paleoceanography* **18**, 1046 (2003).
38. R. W. Eppley, *Fish Bull.* **70**, 1063 (1972).
39. J. B. Palter, J. L. Sarmiento, A. Gnanadesikan, J. Simeon, R. D. Slater, *Biogeosciences* **7**, 3549–3568 (2010).
40. P. C. Tzedakis, H. Hooghiemstra, H. Pälike, *Quat. Sci. Rev.* **25**, 3416–3430 (2006).
41. L. Sadori *et al.*, *Biogeosciences* **13**, 1423–1437 (2016).
42. L. M. Dupont, T. Caley, I. S. Castaneda, *Clim. Past* **15**, 1083–1097 (2019).
43. A. A. Prokopenko, L. A. Hinnov, D. F. Williams, M. I. Kuzmin, *Quat. Sci. Rev.* **25**, 3431–3457 (2006).
44. W. Chen *et al.*, *Quat. Sci. Rev.* **218**, 293–305 (2019).
45. M. N. Woillez *et al.*, *Clim. Past* **7**, 557–577 (2011).
46. M. Hein, K. Sand-Jensen, *Nature* **388**, 526–527 (1997).
47. A. Anav *et al.*, *Rev. Geophys.* **53**, 785–818 (2015).
48. M. Martin Calvo, I. C. Prentice, *New Phytol.* **208**, 987–994 (2015).
49. T. Hanebuth, K. Stattegger, P. M. Grootes, *Science* **288**, 1033–1035 (2000).
50. X. Wang, X. Sun, P. Wang, K. Stattegger, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **278**, 88–97 (2009).
51. B. A. A. Hoogakker *et al.*, *Clim. Past* **12**, 51–73 (2016).
52. C. W. Snyder, *Nature* **538**, 226–228 (2016).
53. R. M. Spratt, L. E. Lisiecki, *Clim. Past* **12**, 1079–1092 (2016).
54. I. C. Prentice, S. P. Harrison, P. J. Bartlein, *New Phytol.* **189**, 988–998 (2011).
55. H. Cheng *et al.*, *Nature* **534**, 640–646 (2016).
56. E. Huang *et al.*, *Sci. Adv.* **6**, eaba4823 (2020).
57. M. P. Hain, D. M. Sigman, G. H. Haug, in *Treatise on Geochemistry, Volume 8: The Oceans and Marine Geochemistry*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, ed. 2, 2014), pp. 485–517.
58. T. Blunier, B. Barnett, M. L. Bender, M. B. Hendricks, *Global Biogeochem. Cycles* **16**, 1029 (2002).
59. L. E. Lisiecki, M. E. Raymo, *Paleoceanography* **20**, PA1003 (2005).
60. L. Bazin *et al.*, *Clim. Past* **9**, 1715–1731 (2013).
61. D. Veres *et al.*, *Clim. Past* **9**, 1733–1748 (2013).
62. F. Lambert, M. Bigler, J. P. Steffensen, M. Hutterli, H. Fischer, *Clim. Past* **8**, 609–623 (2012).
63. T. Extier *et al.*, *Quat. Sci. Rev.* **185**, 244–257 (2018).
64. J. L. Ash, H. Hu, L. Y. Yeung, *ACS Earth Space Chem.* **4**, 50–66 (2020).
65. D. A. Stolper, W. W. Fischer, M. L. Bender, *Geochim. Cosmochim. Acta* **240**, 152–172 (2018).
66. J.-W. Yang *et al.*, Triple isotopic composition of atmospheric oxygen (Δ17O of O2) over 58.0-150.0, 233.2-238.1, and 445.6-796.3 ka from EPICA Dome C ice core. PANGAEA (2022); https://doi.org/10.1594/PANGAEA.941483.

REPORTS

CORONAVIRUS

Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel

Ottavia Prunas[1,2]*, Joshua L. Warren[2,3]†, Forrest W. Crawford[2,3,4,5,6]†, Sivan Gazit[7], Tal Patalon[7], Daniel M. Weinberger[1,2]‡, Virginia E. Pitzer[1,2]‡

The effectiveness of vaccines against COVID-19 on the individual level is well established. However, few studies have examined vaccine effectiveness against transmission. We used a chain binomial model to estimate the effectiveness of vaccination with BNT162b2 [Pfizer-BioNTech messenger RNA (mRNA)-based vaccine] against household transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Israel before and after emergence of the B.1.617.2 (Delta) variant. Vaccination reduced susceptibility to infection by 89.4% [95% confidence interval (CI): 88.7 to 90.0%], whereas vaccine effectiveness against infectiousness given infection was 23.0% (95% CI: −11.3 to 46.7%) during days 10 to 90 after the second dose, before 1 June 2021. Total vaccine effectiveness was 91.8% (95% CI: 88.1 to 94.3%). However, vaccine effectiveness is reduced over time as a result of the combined effect of waning of immunity and emergence of the Delta variant.

The COVID-19 pandemic has led to major disruptions worldwide. The rapid development and deployment of vaccines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has provided an opportunity to control the outbreak in populations with access to vaccination. Multiple vaccines against SARS-CoV-2 have been developed that effectively prevent clinical disease and reduce disease severity in those who do become infected (*1–3*); this direct protection against disease is critical. However, additional population-level benefits can be derived if vaccines also reduce transmission of the virus, thereby providing protection to those who are still vulnerable to infection (*1, 4*).

To date, there is little direct real-world evidence about the effects of vaccination on SARS-CoV-2 transmission. A few studies have investigated the reduction in transmission in households and among healthcare workers (*3, 5, 6*). Other studies have found indirect evidence for a likely effect of the vaccine on transmission by demonstrating reduced viral load in the upper respiratory tract of infected

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT, USA. [2]Public Health Modeling Unit, Yale School of Public Health, Yale University, New Haven, CT, USA. [3]Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT, USA. [4]Department of Statistics and Data Science, Yale School of Public Health, Yale University, New Haven, CT, USA. [5]Department of Ecology and Evolutionary Biology, Yale School of Public Health, Yale University, New Haven, CT, USA. [6]Yale School of Management, Yale University, New Haven, CT, USA. [7]Maccabi Institute for Research and Innovation, Maccabi Healthcare Services, Tel Aviv, Israel.
*Corresponding author. Email: ottavia.prunas@yale.edu
†These authors contributed equally to this work.
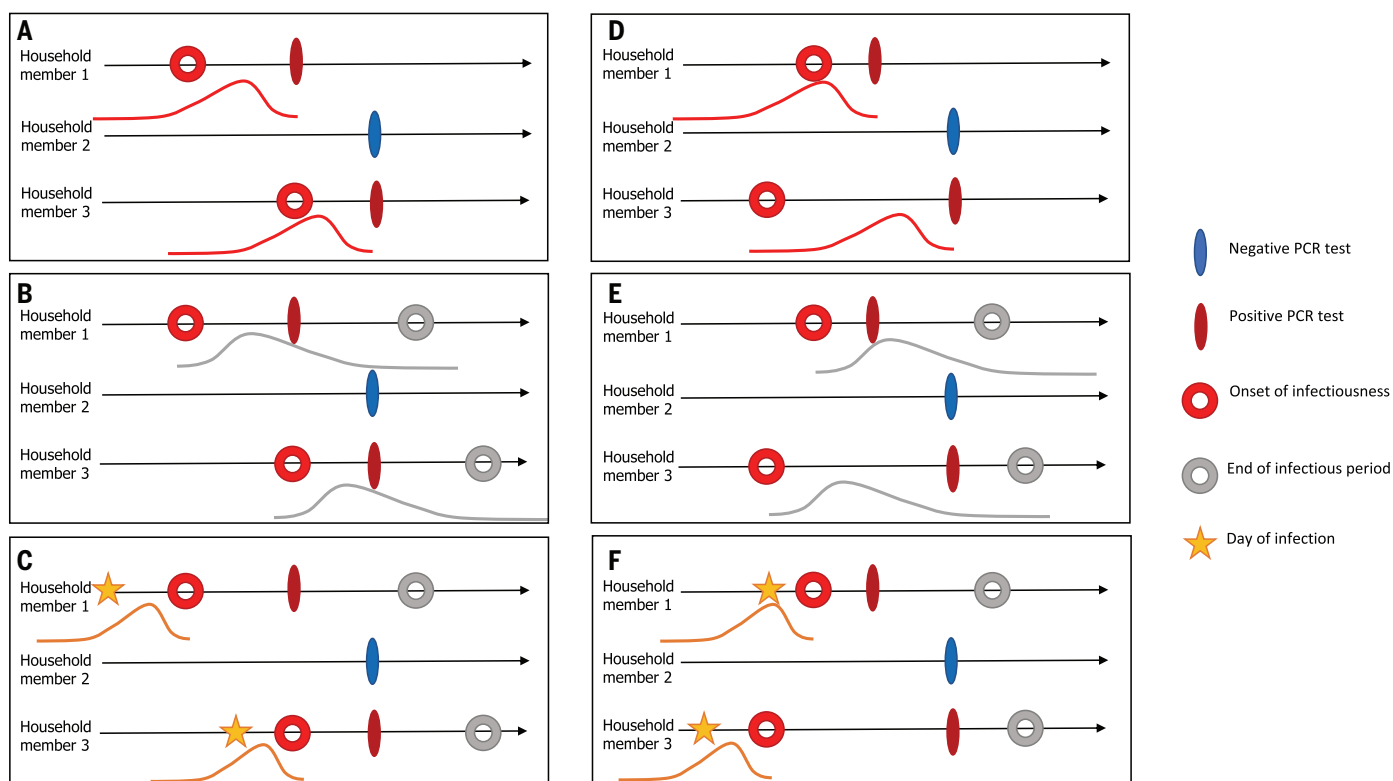‡These authors contributed equally to this work.

Fig. 1. Schematic representation of the multiple imputation process for an example household. Each infected household member is associated with: (**A** and **D**) a distribution for time from onset of infectiousness to testing; (**B** and **E**) a distribution for the infectious period; and (**C** and **F**) a distribution for the latent period to infer the time of infection. The filled ovals represent observed events, and the circles and stars represent unobserved events in the infection timeline. Panels (A) to (C) and (D) to (F) represent two possible sample sets from the delay distributions, each with a different index case, who is not necessarily the first person to test positive in the household. We generated 100 samples of the latent data for each infected individual.

individuals (*7–11*). These studies have mostly focused on the period when the B.1.1.7 (Alpha) variant was the dominant strain and have not examined effects on transmission after emergence of the B.1.617.2 (Delta) variant (*12*).

Households are an ideal setting for evaluating transmission of the virus and the effects of vaccination as a result of the high secondary attack rate (SAR) among household members (*3, 13*). Detailed data on household structure and timing of infections can be used to quantify the risk of transmission. We aimed to assess the effectiveness of vaccination with the Pfizer-BioNTech mRNA-based vaccine (BNT162b2) against susceptibility to infection and against infectiousness given infection with SARS-CoV-2, comparing the pre- and post-Delta periods. We accomplished this by means of a chain binomial model—a common approach for reconstruction of transmission in household settings (*14*)—applied to data from the second largest healthcare organization in Israel. The rapid and early rollout of mass vaccination in Israel provides a notable opportunity to evaluate the effectiveness of vaccination against transmission.

We used data from the centralized database of Maccabi Healthcare Services (MHS), which captures all information on the demographics and healthcare-related interactions of members. MHS is a nationwide, state-mandated, not-for-profit healthcare delivery organization in Israel with 2.5 million members, representing a quarter of the Israeli population. The full dataset covered the period from 1 June 2020 to 28 July 2021 and included information on 2,472,502 individuals from 1,327,647 households. Among these, 1,471,386 individuals had received two doses of BNT162b2 as of 28 July 2021 (before the widespread introduction of booster doses). There were 202,298 detected infections caused by SARS-CoV-2 (8.2% of the total population), with 6483 infections in fully vaccinated individuals at the time of their polymerase chain reaction (PCR) test date and 186,975 infections in unvaccinated individuals (unadjusted risk ratio = 0.066%) (table S1 and fig. S1).

The majority of households (60%) had a single household member; this individual was infected in 62,295 (7.8%) of 797,170 households. Information on the number of households and proportion of infections occurring in households of varying size can be found in table S2. The naïve SAR, based on the vaccination status of the "index case" (defined as the first person to test positive in a household), was lower when the index case was vaccinated during the pre-Delta period (table S3).

We used a chain binomial model for household transmission to estimate how the probability of infection per day depended on the characteristics of susceptible individuals and their household contacts (*14, 15*). An individual's infection probability is modeled as the risk of escaping infection from the community and any or all infectious household members on each day of exposure (see materials and methods). We used multiple imputation to generate latent data for when a person with a positive PCR test was infected and infectious. This was accomplished by using random samples from three different Gamma distributions representing the delay between onset of infectiousness and the date of the PCR test, the date of infection and onset of infectiousness (i.e., the latent period), and the onset of infectiousness to the end of infectiousness (i.e., the infectious period) (Fig. 1 and table S4; materials and methods). We performed sensitivity analyses to confirm the robustness of our results to variability in the delay distributions; we also performed a simulation study to validate our approach (figs. S2 to S5 and tables S5 to S7; materials and methods).

Table 1. Vaccine effectiveness against susceptibility to infection (VE$_S$); vaccine effectiveness against infectiousness given infection (VE$_I$); total vaccine effectiveness (VE$_T$) at different time ranges since vaccination, both before and after the emergence of the Delta variant (1 June, 2021).

Vaccine effectiveness measure	Time since vaccination	Estimate pre Delta [95% confidence interval]	Estimate post Delta [95% confidence interval]
Vaccine effectiveness against susceptibility to infection			
VE$_{S1}$	≥10 days after dose 1 and <10 days after dose 2	62.7% [61.5% to 63.8%]	72.1% [66.7% to 75.6%]
VE$_{S2}$	≥10 days after dose 2 and <90 days after dose 2	89.4% [88.7% to 90.0%]	72.0% [65.9% to 77.0%]
VE$_{S3}$	≥90 days after dose 2	58.3% [45.8% to 67.9%]	40.2% [37.6% to 42.6%]
Vaccine effectiveness against infectiousness given infection			
VE$_{I1}$	≥10 days after dose 1 and <10 days after dose 2	−15.9% [−27.9% to −5.0%]	38.3% [−24.2% to 69.3%]
VE$_{I2}$	≥10 days after dose 2 and <90 days after dose 2	23.0% [−11.3% to 46.7%]	−27.9% [−248.9% to 53.1%]
VE$_{I3}$	≥90 days after dose 2	6.9% [−124.8% to 61.4%]	−27.9% [−53.7% to −6.5%]
Total vaccine effectiveness			
VE$_{T1}$	≥10 days after dose 1 and <10 days after dose 2	56.8% [52.2% to 60.9%]	82.8% [64.8% to 91.6%]
VE$_{T1}$	≥10 days after dose 2 and <90 days after dose 2	91.8% [88.1% to 94.3%]	65.6% [4.9% to 87.6%]
VE$_{T3}$	≥90 days after dose 2	61.1% [5.2% to 84.1%]	24.2% [9.0% to 36.9%]

The pairwise daily probability of infection from the community and from each infected household member was modeled as a function of the time-varying number of SARS-CoV-2–positive individuals in the population, the characteristics of the susceptible individual (including age and vaccination status), and the vaccination status of their household contacts. We considered four categories of vaccination: (i) unvaccinated; (ii) ≥10 days from dose 1 to <10 days from dose 2; (iii) ≥10 days to <90 days from dose 2; and (iv) ≥90 days from dose 2 to account for partial vaccination, full vaccination, and waning of vaccine-induced immunity, respectively. Vaccine effectiveness against susceptibility to infection was estimated from the coefficient of the susceptible individual's vaccination status, whereas vaccine effectiveness against infectiousness given infection was estimated from the coefficient of the vaccination status of each infectious household member. To determine the effect of the Delta variant, we allowed the vaccine effects to vary before and after 1 June 2021 (i.e., the pre- and post-Delta period, respectively). We estimated the effects by averaging over 100 draws from the delay distributions used in the multiple imputation process; the variance of the estimates across these 100 draws was estimated with the law of total variance (figs. S6 to S8).

For the period before 1 June 2021 (before emergence of the Delta variant), receipt of two doses of the vaccine was associated with a vaccine effectiveness against susceptibility to infection (VE$_S$) of 89.4% [95% confidence interval (CI): 88.7 to 90.0%] within 10 to 90 days of receiving the second dose, and 58.3% (95% CI: 45.8 to 67.9%) more than 90 days after receiving the second dose. The vaccine effectiveness against infectiousness given infection (VE$_I$) was 23.0% (95% CI: −11.3 to 46.7%) within 10

to 90 days and 6.9% (95% CI: −124.8 to 61.4%) more than 90 days after the second dose (Table 1). The total vaccine effectiveness (VE$_T$), which combines the reduction in the risk of infection and the risk of infectiousness given infection among vaccinated individuals, was estimated to be 91.8% (95% CI: 88.1 to 94.3%) within 10 to 90 days, and 61.1% (95% CI: 5.2 to 84.1%) more than 90 days after the second dose. Evidence of waning protection after vaccination was apparent for the ≥90-day time period after the second dose for all vaccine effects (Table 1).

After the emergence of the Delta variant, we observed a marked reduction in the vaccine effectiveness against susceptibility to infection compared with the pre-Delta period. During this period, the VE$_S$ was 72.0% (95% CI: 65.9 to 77.0%) within 10 to 90 days and 40.2% (95% CI: 37.6 to 42.6%) more than 90 days after the second dose. A similar finding was observed for total vaccine effectiveness: VE$_T$ = 65.6% (95% CI: 4.9 to 87.6%) within 10 to 90 days and 24.2% (95% CI: 9.0 to 36.9%) more than 90 days after the second dose. There was a high degree of uncertainty in the estimates of vaccine effectiveness against infectiousness given infection during the Delta period (Table 1). Allowing for differences in vaccine effectiveness for the post-Delta period improved the model fit, on the basis of a comparison with the Akaike information criteria (figs. S9 and S10).

We further analyzed the effect of vaccination on infectiousness given infection when restricting our data to the susceptible unvaccinated population (i.e., children <12 years of age). We observed a larger reduction in risk for children exposed to a vaccinated versus unvaccinated infectious household member, with VE$_I$ = 41.0% (95% CI: −13.7 to 69.4%)

between 10 and 90 days from receiving the second dose (table S8). The corresponding vaccine effect during the Delta period was not significantly different from zero.

The probability of transmission per day from an infected household member to a susceptible adult during the pre-Delta period was 0.021 (95% CI: 0.020 to 0.021), leading to a SAR of 0.10 (95% CI: 0.09 to 0.10) (table S9; materials and methods). The risk of transmission from an infectious household member was ~100 times as high as that of the average risk of infection from the community. During the period when the Delta variant was dominant, there was no meaningful increase in household transmission probability, whereas there was an increase in the risk of infection from the community [relative risk (RR) = 1.13; 95% CI: 1.09 to 1.16] (table S9). Children <12 years old had a lower risk of infection from both the community and an infectious household member, whereas adults 40 to 64 and ≥65 years of age had a lower risk of infection from the community but a higher risk of infection within the household compared with individuals aged 12 to 39 years (table S9). In a sensitivity analysis, we found that children were slightly less infectious than adults (see SM).

To date, there is limited evidence with which to compare our estimates of vaccine effectiveness against infectiousness and transmission. A study of more than 550,000 households in England showed that vaccination with both the ChAdOx1 nCoV-19 and BNT162b2 vaccines reduced the odds of transmission from a vaccinated and infected household member by 40 to 50% compared with unvaccinated index cases (*1*, *3*). A similar study in Denmark estimated the reduction in transmission to be 42% during the Delta period (*16*). In previous studies, the index case in each household

was defined as the earliest case of laboratory-confirmed COVID-19 by diagnosis date, and all secondary infections in the household were attributed to the index case (3). By contrast, by inferring the date of infection we do not assume that the index case in the household was necessarily the first individual to be diagnosed, and we account for the risk of transmission from other infected household members and from the community. With our approach, we show a lower and uncertain reduction in infectiousness given infection, compared with simpler methods (3, 16, 17). A comparable statistical approach was used in another study in Israel, where members of households with confirmed cases were actively followed and tested. A notably higher reduction in infectivity was observed, though with large uncertainty; however, the study was limited to healthcare workers, who normally represent a younger and healthier population thereby potentially leading to a stronger vaccine effect (6). Other studies investigating the reduction in infection risk among household members of vaccinated versus unvaccinated healthcare workers were conducted in Scotland and Finland, providing indirect evidence of a lower risk of infection among household contacts of vaccinated individuals (1, 5, 18).

Our analyses suggest that before emergence of the Delta variant, breakthrough cases among vaccinated individuals had slightly reduced infectiousness compared with unvaccinated cases. However, both waning of vaccine-induced immunity and the emergence of the Delta variant were associated with a reduction in the VE_I. These results are in agreement with recent findings in a UK study, where the SAR was similar for vaccinated and unvaccinated index cases infected with the Delta variant (12). However, vaccination still reduces the risk of transmission by providing protection against susceptibility to infection, even if this effect is reduced over time because of both waning immunity and the Delta variant, as highlighted in real-world settings (12, 19, 20).

This study has several important limitations: First, we did not have information on the true infection times (and duration of infectiousness) of infected household members. To overcome this limitation, we sampled from three delay distributions parameterized from the literature to determine the potential infection status of each individual through time. Our approach is suboptimal, however, because it was not computationally feasible to estimate the parameters of the delay distributions conditional on the observed data, e.g., by means of an expectation-maximization or Markov chain Monte Carlo approach. As a result, parameter estimates do not reflect uncertainty in the delay distribution parameters. This could lead to artificially narrow confidence intervals for some parameters. In addition, the VE_I esti-

mates were dependent upon the specification of the time from onset of infectiousness to testing (fig. S5). Second, individuals who were infected but did not receive a SARS-CoV-2 test would be misclassified in our dataset. This is likely to have only a minor effect on our estimates, though the VE_I could be underestimated if the probability of detection per day is low (see SM, tables S6 and S7). We estimated a negative VE_I in partially vaccinated cases, suggesting possible sources of bias in our analysis (e.g., partially vaccinated individuals may be less likely to isolate at the first sign of symptoms). This effect is mitigated during the post-Delta period (Table 1). Controlling for the age of infectious individuals did not resolve the potential bias (table S10). Finally, our results do not include the period when the Omicron variant has become dominant, although recent findings suggest that SARs among unvaccinated household members are comparable to the Delta variant (21).

Vaccination can prevent transmission by both providing protection against infection (including asymptomatic infections) and reducing the infectiousness of vaccinated individuals who do become infected. Neither of these are typically directly measured in vaccine trials. By analyzing data on confirmed SARS-CoV-2 infections among household members in Israel, we provide measures of effectiveness of BNT162b2 against susceptibility to infection and against infectiousness given infection. Our results show evidence of a slight reduction in the infectiousness of vaccinated individuals who become infected in addition to protection against susceptibility to infection, leading to an overall reduction in the risk of transmission. However, the ability of vaccination to prevent transmission is reduced over time because of waning of vaccine-induced immunity and lower effectiveness against the Delta variant. It is highly unlikely that population-level transmission of SARS-CoV-2 can be eliminated through vaccination alone.

REFERENCES AND NOTES

1. A. Richterman, E. A. Meyerowitz, M. Cevik, *Open Forum Infect. Dis.* **9**, ofab259 (2021).
2. N. Dagan et al., *N. Engl. J. Med.* **384**, 1412–1423 (2021).
3. R. J. Harris et al., Effect of Vaccination on Household Transmission of SARS-CoV-2 in England. *N. Engl. J. Med.* NEJMc2107717 (2021).
4. Z. J. Madewell, Y. Yang, I. M. Longini Jr., M. E. Halloran, N. E. Dean, *JAMA Netw. Open* **3**, e2031756 (2020).
5. A. S. V. Shah et al., *NEJM* **385**, 1718–1720 (2021).
6. M. Layan et al., medRxiv 2021.07.12.21260377 [Preprint] (2021); doi:10.1101/2021.07.12.21260377.
7. M. Levine-Tiefenbrun et al., *Nat. Med.* **27**, 790–792 (2021).
8. X. Qiu, A. I. Nergiz, A. E. Maraolo, I. I. Bogoch, N. Low, M. Cevik, Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission–a living systematic review. *Clin. Microbiol. Infect.* (2021).
9. M. Marks et al., *Lancet Infect. Dis.* **21**, 629–636 (2021).
10. E. Petter et al., medRxiv 2021.02.08.21251329 [Preprint] (2021); doi:10.1101/2021.02.08.21251329.
11. F. P. Lyngse et al., medRxiv 2021.02.28.21252608 [Preprint] (2021); doi:10.1101/2021.02.28.21252608.
12. A. Singanayagam et al., *Lancet Infect. Dis.* **22**, 183–195 (2022).
13. V. E. Pitzer, T. Cohen, *Lancet Infect. Dis.* **20**, 1103–1104 (2020).
14. A. H. Rampey Jr., I. M. Longini Jr., M. Haber, A. S. Monto, *Biometrics* **48**, 117–128 (1992).
15. Y. Yang, I. M. Longini Jr., M. E. Halloran, *Appl. Stat.* **55**, 317–330 (2006).
16. F. P. Lyngse et al., medRxiv 2022.01.06.22268841 [Preprint] (2022); doi:10.1111/j.1467-9876.2006.00539.x.
17. D. W. Eyre et al., Effect of Covid-19 Vaccination on Transmission of Alpha and Delta Variants. *N. Engl. J. Med.* NEJMoa2116597 (2022).
18. J. Salo et al., medRxiv 2021.05.27.21257896 [Preprint] (2021); doi:10.1101/2021.05.27.21257896.
19. T. Patalon et al., *JAMA Intern. Med.* **182**, 179–184 (2022).
20. B. Mizrahi et al., medRxiv 2021.07.29.21261317 [Preprint] (2021); doi:10.1101/2021.07.29.21261317.
21. F. P. Lyngse et al., medRxiv 2021.12.27.21268278v1 [Preprint] (2021); doi:10.1101/2021.12.27.21268278v1.
22. O. W. Prunas et al., *Zenodo* (2021); http://dx.doi.org/10.5281/zenodo.5670699.

CORONAVIRUS

Indirect protection of children from SARS-CoV-2 infection through parental vaccination

Samah Hayek[1], Galit Shaham[1], Yatir Ben-Shlomo[1], Eldad Kepten[1], Noa Dagan[1,2,3,4], Daniel Nevo[5], Marc Lipsitch[6], Ben Y. Reis[3,4,7], Ran D. Balicer[1,8], Noam Barda[2,1,3,4]*

Children not vaccinated against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) may still benefit from vaccines through protection from vaccinated contacts. We estimated the protection provided to children through parental vaccination with the BNT162b2 vaccine. We studied households without prior infection consisting of two parents and unvaccinated children, estimating the effect of parental vaccination on the risk of infection for unvaccinated children. We studied two periods separately—an early period (17 January 2021 to 28 March 2021; Alpha variant, two doses versus no vaccination) and a late period (11 July 2021 to 30 September 2021; Delta variant, booster dose versus two vaccine doses). We found that having a single vaccinated parent was associated with a 26.0 and a 20.8% decreased risk in the early and late periods, respectively, and having two vaccinated parents was associated with a 71.7 and a 58.1% decreased risk, respectively. Thus, parental vaccination confers substantial protection on unvaccinated children in the household.

Since December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread globally (1), resulting in more than 200 million confirmed infections and more than 4 million deaths (2). COVID-19 vaccines serve a critical role in combating the spread of the pandemic. Vaccination exerts its effects through both direct protection of vaccinated individuals and indirect protection of individuals living in vaccinated environments (3).

Households have specific importance in the context of infectious disease dynamics. Several epidemiological studies have reported that a substantial amount of COVID-19 transmission occurs in settings that include close and prolonged contact, such as households (4–6). The importance of households in SARS-CoV-2 transmission was highlighted in a recent meta-analysis, in which the secondary attack rate was found to be 19.0% [95% confidence interval (CI): 16.2%, 22.0%] (7). The central role of households in SARS-CoV-2 transmission allows them to be used as alternatives to larger clusters for estimating the direct and indirect effects of vaccines (3).

Unlike the direct effect of the BNT162b2 mRNA COVID-19 vaccine, which has been ex-tensively explored in clinical trials (8) and observational studies (9, 10), the indirect effect of the vaccine has not received as much attention. Previous studies have shown that a single vaccinated household member confers modest protection [42.9% (95% CI: 22.3%, 58.1%), 10 weeks after the first dose] against SARS-CoV-2 infection on other unvaccinated adult household members (11). A study from Israel has shown that vaccination reduces the risk of infection and of transmission once an infection is introduced into the household and that unvaccinated spouses of health care workers are protected by their spouse's vaccination (12). A different study evaluated the indirect effect at a different level, using 177 geographical communities in Israel, and showed that higher rates of vaccination in each community were associated with a substantial decline in infections among a cohort of unvaccinated individuals aged 16 years or younger (13). In general, previous studies concerning indirect effects of vaccination had small sample sizes, included only specific populations (e.g., health care workers), did not adjust for certain important confounders, only covered a single period and disease variant, and did not explore the mechanism of the indirect effect.

In Israel, the BNT162b2 mRNA COVID-19 vaccine was authorized in December 2020 for individuals aged 16 years and older. In May 2021, this authorization was extended to children and adolescents aged 12 years and older and, in November 2021, to children aged 5 years and older. Third-dose booster shots were initiated in Israel on 11 July 2021 and were gradually extended to cover the entire population—who had received the second dose at least 5 months prior—over the month of August. In parallel, from December 2020 to March 2021, Israel underwent a third wave of the COVID-19 pandemic, in which the Alpha variant was dominant. This wave was accompanied by a nationwide lockdown that included closure of schools and limitation of social activities. A fourth wave occurred in Israel from June to October 2021, this time dominated by the Delta variant. No lockdowns were in effect during this wave; however, during the months of July and August, the schools were closed for summer vacation. Throughout 2021, COVID-19 polymerase chain reaction (PCR) tests were freely available nationwide and targeted sampling was performed in schools in which a teacher or a child were found to be infected. In Israel (14), as in Europe (15) and the US (16), the younger age groups remain the least vaccinated.

In this study, we use the integrated data repositories of Israel's largest health care organization to estimate the indirect vaccine effectiveness (VE) of the BNT162b2 mRNA COVID-19 vaccine on unvaccinated children within households. We perform this analysis over two time periods: an early period (17 January 2021 through 28 March 2021) in children <16 years old when the Alpha variant was dominant, in which we compare households with parents who were vaccinated with the primary vaccine series with households with unvaccinated parents, and a late period (11 July 2021 through 30 September 2021) in children <11 years old when the Delta variant was dominant, in which we compare households with parents who were vaccinated with a booster dose with households in which parents were previously vaccinated with two vaccine doses but have not received the booster dose. In each period, we assess the change in the risk of SARS-CoV-2 infection among susceptible children in the household (who are not eligible for vaccination) associated with the vaccination of one or both parents. Furthermore, in each period, we explore two of the mechanisms mediating this effect by estimating the decrease in risk that a vaccinated parent would be infected (direct VE) and the decrease in risk that a vaccinated infected parent would then proceed to infect a susceptible child [household infectiousness, or secondary attack rate (SAR)].

The early period of the study included 400,733 unvaccinated subjects (children and adolescents) from 155,305 distinct households who contributed 2,116,306 person-weeks (defined as one week of follow-up for one subject) of follow-up (fig. S1A). The median age of the children was 6 years old [interquartile range (IQR): 3, 9], and 52% of subjects were male. The late period of the study included 181,307 unvaccinated children from 76,621 distinct households who contributed 1,089,191 person-weeks of follow-up (fig. S1B). The median age of the children was 5 years old (IQR: 2, 7), and 52% of subjects were male.

[1]Clalit Research Institute, Clalit Health Services, Ramat Gan, Israel. [2]Software and Information Systems Engineering, Ben Gurion University, Be'er Sheva, Israel. [3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [4]The Ivan and Francesca Berkowitz Family Living Laboratory Collaboration at Harvard Medical School and Clalit Research Institute, Boston, MA, USA. [5]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel. [6]Center for Communicable Disease Dynamics, Department of Epidemiology, and Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [7]Predictive Medicine Group, Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA. [8]School of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheba, Israel.
*Corresponding author. Email: noambard@bgu.ac.il

Baseline demographic and clinical characteristics of the subjects in each time period are shown in Table 1. A more detailed description, including all potential confounders stratified by parental vaccination status, is presented in table S1. A time series of the cases observed in our study (the epidemic curve) during both periods, stratified by age group, are presented in fig. S2. Table S2 describes the differences between the infected and uninfected subjects in both study periods.

During the early period, focusing on the Alpha variant and comparing parents vaccinated with the primary vaccine series with unvaccinated parents, a single vaccinated parent was associated with a 26.0% (95% CI: 14.0%, 36.2%) decreased risk of infection for children living in the same household, and two vaccinated parents were associated with a 71.7% (68.6%, 74.6%) decreased risk of infection. This effect was fairly uniform across subject age groups and household sizes. For example, the adjusted VE was 67.1% (52.4%, 77.3%) for a household of size 3 in which both parents were vaccinated and 62.9% (44.2%, 75.4%) for a household of size 7 in which both parents were vaccinated (Fig. 1 and table S3).

During the late period, focusing on the Delta variant and comparing parents vaccinated with a third (booster) dose with parents who received only two doses at least 5 months prior, a single boosted parent was associated with a 20.8% (95% CI: 11.4%, 29.1%) decreased risk for infection, whereas two boosted parents were associated with a 58.1% (53.1%, 62.6%) decreased risk for infection. Some heterogeneity of the effect was observed between age groups and household sizes. For example, the adjusted VE was 65.9% (56.7%, 73.2%) for a subject aged 0 to 2 years living with two boosted parents, and the adjusted VE was 55.5% (48.6%, 61.6%) for a subject aged 7 to 11 years living with two boosted parents (Fig. 1 and table S3). In both periods, plots of the predicted versus observed incidence rates indicate a good model fit (fig. S3).

Analysis of the direct effect of the BNT162b2 mRNA COVID-19 vaccine on the risk of parental infection estimated a reduction of 94.4% (95% CI: 93.2%, 95.4%) in the risk of documented infection during the early period (Alpha variant) and 86.3% (83.4%, 88.6%) in the risk of documented infection during the late period (Delta variant) among fully vaccinated adults (Table 2). Full vaccination of an infected parent was associated with a 72.1% (36.6%, 89.3%) decreased odds of infection of one or more susceptible children in the household from that parent during the early period and a 79.6% (55.9%, 91.8%) decreased odds of transmission from a boosted, infected parent to one or more susceptible children during the late period (Table 3), in both cases adjusting for the vaccination status of the other parent.

Figure 2 shows a schematic representation of the mechanism by which the direct protection of the parent and the reduction in the secondary attack rate make up the indirect protection observed for the children. In the sensitivity analysis using bacterial diarrhea as a negative control outcome, the association (VE) was −14% (95% CI: −49%, 13%) for one vaccinated parent and −16% (−37%, 1.8%) for two vaccinated parents (table S4).

In this study, we estimated the indirect protective effect of vaccinating parents with the BNT162b2 mRNA COVID-19 vaccine on their children's risk of SARS-CoV-2 infection in households without prior infection. This estimation was performed for both the primary vaccine series during a period in which the Alpha variant was dominant and the vaccine booster dose during a period in which the Delta variant was dominant. In both periods, we found that parental vaccination substantially reduced the risk of children being infected with SARS-CoV-2, though the effect was somewhat smaller during the late period. Although this smaller effect could result from heterogeneity, as the populations are different in composition, it more likely stems from unboosted parents still being somewhat protected from the first two vaccine doses, which makes the relative effect of the additional booster vaccination dose smaller. Notably, we found the

Table 1. Descriptive statistics of the study population. The study population includes susceptible children under the age of vaccination eligibility and residing in the households included in the study. The early period was 17 January 2021 to 28 March 2021. The late period was 11 July 2021 to 30 September 2021. NA, not applicable.

Characteristic	Early period (N = 400,733)	Late period (N = 181,307)
Median age (IQR)	6 (3, 9)	5 (2, 7)
Median household size (IQR)	5 (4, 6)	5 (4, 5)
Age (N, %)		
0-2	81,672 (20%)	47,710 (26%)
3-6	135,230 (34%)	75,278 (42%)
7-12*	149,917 (37%)	58,319 (32%)
13-15	33,914 (8.5%)	NA
Sex (N, %)		
Female	194,272 (48%)	87,913 (48%)
Male	206,461 (52%)	93,394 (52%)
Population group (N, %)		
Arabs	101,557 (25%)	32,484 (18%)
General	277,444 (69%)	140,222 (77%)
Ultra-Orthodox Jewish	21,732 (5.4%)	8,601 (4.7%)
Socioeconomic status (N, %)		
Low	223,108 (56%)	88,023 (49%)
Medium	162,833 (41%)	87,380 (48%)
High	14,792 (3.7%)	5,904 (3.3%)
Household size (N, %)		
3	20,127 (5.0%)	11,936 (6.6%)
4	107,549 (27%)	63,866 (35%)
5	157,379 (39%)	73,150 (40%)
6	80,423 (20%)	24,878 (14%)
7	35,255 (8.8%)	7,477 (4.1%)
Residence type (N, %)		
Large city	127,887 (32%)	62,991 (35%)
Small city	149,260 (37%)	67,702 (37%)
Town	76,764 (19%)	28,817 (16%)
Rural	31,555 (7.9%)	13,888 (7.7%)
Kibbutz (communal residence)	15,267 (3.8%)	7,909 (4.4%)
Chronic conditions (N, %)		
Obesity	24,780 (6.2%)	9,524 (5.3%)
Cardiovascular conditions	1,833 (0.5%)	460 (0.3%)
Pulmonary disease	47,823 (12%)	18,779 (10%)
Type 2 diabetes	1,833 (0.5%)	636 (0.4%)
Hypertension	619 (0.2%)	238 (0.1%)
Active malignancy	240 (<0.1%)	135 (<0.1%)

*The late period includes children up to age 11.

effect for two vaccinated parents to be substantially larger than that for a single vaccinated parent in both periods (26.0% versus 71.7% in the early period; 20.8% versus 58.1% in the late period). This emphasizes that even a single unvaccinated parent remains an important vector for introducing infections into the household.

Previous findings have also shown a substantial indirect effect of SARS-CoV-2 vaccines. A study focusing on unvaccinated spouses of health care workers found the indirect effect to be 43% (95% CI: 23%, 58%) 10 weeks after receipt of the first vaccine dose (11). A study from Israel found that once COVID-19 is introduced into a household, vaccination reduces infectivity by 78% (30%, 94%) (12). Another study from Israel using geographical areas to estimate the community-level protection resulting from vaccinated individuals found that, on average, for every absolute 20% increase in the number of vaccinated individuals, the positive test fraction of the unvaccinated population decreased by a factor of ~2 (13). In general, it is difficult to directly compare the findings of these studies with the current study because of the different designs, adjustments, and exposure definitions.

The present study focused on the indirect benefits of vaccinated parents for unvaccinated children. Indirect vaccine effects are mediated by two main mechanisms: (i) by protecting potential contacts, vaccination reduces the likelihood that subjects will encounter an infectious individual and (ii) vaccination may reduce the infectiousness of vaccinated individuals who do acquire the infection (17, 18). We explored these two mechanisms by estimating the direct effect of parental vaccination on parental infection as well as the vaccination-related change in the risk of infection from an infected parent to a susceptible child. We found the direct effect of parental vaccination with two vaccine doses to be 94.4% (95% CI: 93.2%, 95.4%) for acquiring a documented infection with the Alpha variant and the direct effect of a booster dose to be 86.3% (83.4%, 88.6%) for acquiring a documented infection with the Delta variant. This high effectiveness when comparing parents who have received the booster dose with those who have not also hints at waning immunity after the second dose. Furthermore, we found that infectiousness to the children in the household from an infected parent vaccinated with two doses is reduced by 72.1% (36.6%, 89.3%) compared with the infectiousness of an unvaccinated parent, and infectiousness from a booster-vaccinated parent is reduced by 79.6% (55.9%, 91.8%) compared with that of a parent who did not receive the booster vaccination dose, in each case adjusting for the vaccination status of the other parent. It should be emphasized that we should not expect the in-
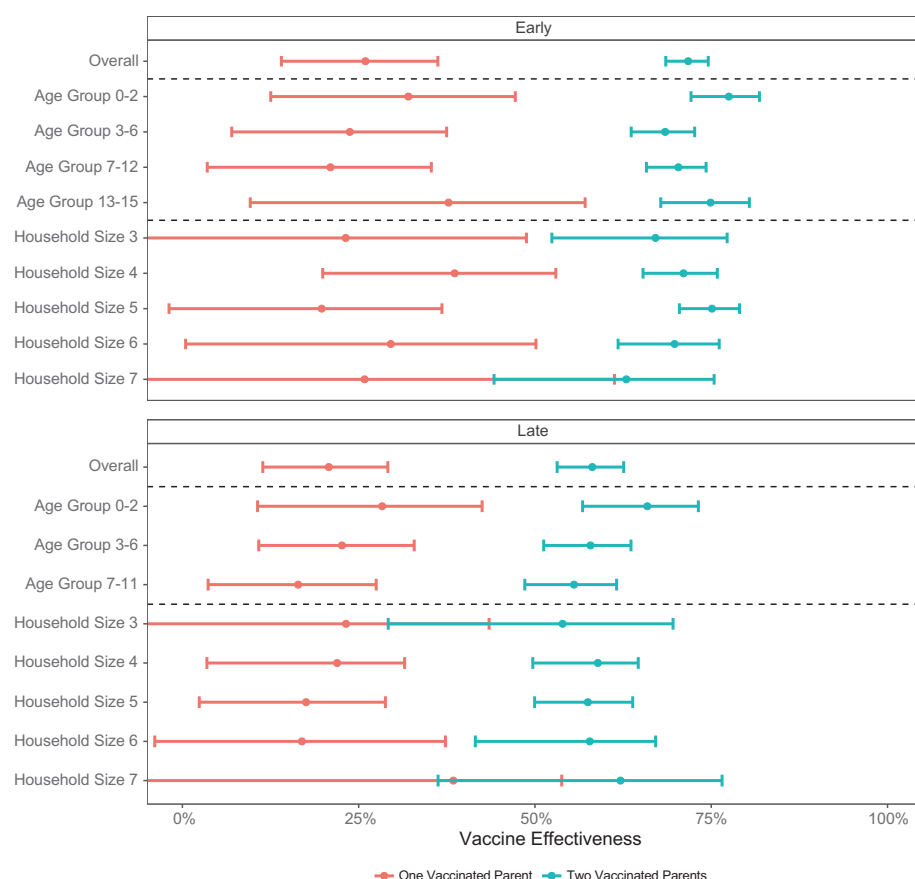
Fig. 1. Indirect effect of the BNT162b2 mRNA COVID-19 vaccine by age group and household size. Indirect VE (one minus the incidence rate ratio) of one vaccinated parent and two vaccinated parents on the probability of infection of a susceptible child within the household, overall and within age group and household size categories. Points represent the point estimates, and error bars represent the 95% CIs. The top part shows the early study period (vaccination with two doses at least 7 days prior versus no vaccination; Alpha variant), and the bottom part shows the late study period (receipt of the booster dose versus no receipt of the booster dose; Delta variant). The numeric results included in this figure are presented in table S3.

Table 2. Direct effect of BNT162b2 mRNA COVID-19 vaccine. Direct VE is the reduction in the probability of infection of a fully vaccinated parent compared with an unvaccinated or unboosted parent, defined as one minus the incidence rate ratio. During the early period, full vaccination was defined as the receipt of two doses at least 7 days prior (compared with no vaccination), and the dominant variant was Alpha. During the late period, full vaccination was defined as receipt of a third dose at least 7 days prior (compared with receipt of only two doses at least 5 months prior), and the dominant variant was Delta. Analysis was performed as per the main analysis, this time using parental infection as the outcome. The model was adjusted for individual- and household-level characteristics. See table S1 for the full list.

Characteristic	Early period	Late period
Direct VE (95% CI)	94.4% (93.2%, 95.4%)	86.3% (83.4%, 88.6%)

direct risk to be equal to the product of the direct risk and the infectiousness because children may also be infected outside of the household or, potentially, through the other parent. The estimated direct VE of the parents

is consistent with previous literature (9, 19) as are the results concerning the reduced SAR (12).

To detect possible bias originating from uncontrolled confounding, we performed an analysis using a negative control outcome (NCO)

(20, 21), bacterial diarrhea. Bacterial diarrhea was chosen because it plausibly shares confounders (e.g., health-related behavior and hygiene) with the outcome of interest but should not be affected by the exposure of interest (SARS-CoV-2 vaccine). This analysis did not detect substantial effects, further strengthening our findings and reducing the possibility of meaningful unmeasured confounding.

The protective effect of parental vaccination on children's risk described in this study has particular importance for several reasons: First, although children often experience asymptomatic or mild disease when infected with SARS-CoV-2, some do experience severe disease (22, 23) and enduring postinfection symptoms (known as Long Covid) (24), particularly when suffering from some degree of immunosuppression (25). Second, because of the important role of households in propagating COVID-19 transmission, reducing the number of infected children may help decrease the overall spread of the pandemic throughout the population.

This study is subject to several limitations. First, we did not determine the proportion of infections arising from a source outside the household. Changing the level of external exposure of the children, for example through school attendance, would alter the indirect effectiveness of the vaccine (26) because parental vaccination would not reduce children's exposure to infectious nonhousehold members. Second, determination of household membership was based on demographic records in our database. It is possible that some individuals reside at a different location than the address listed or that additional persons (e.g., grandparents or nonparent caregivers) reside in the same household. Third, infections were dated on the basis of the date of sampling, which is invariably several days later than the date of infection. This could result in errors when attributing infections to specific weeks or, in cases where both parent and child became infected, may misclassify the sequence of infections (27). Fourth, it is possible that we did not capture important confounders, particularly those related to behavior, which would lead to residual confounding. Fifth, it is possible that outcomes were differentially misclassified between the two study groups—e.g., because a positive diagnosis of an unvaccinated parent would prompt further tests of members of the household. This would result in elevated VE estimates. Lastly, the analysis for secondary attack rate is conditioned on a parent having been infected and on no further infection on days 0 to 2 after the index infection, which are both posttreatment variables. This could result in collider stratification bias.

The results of this study show that parental vaccination confers substantial protection on children residing in the same household. They

Table 3. Secondary transmission risk. The secondary attack rate (SAR) from an infected parent to susceptible children in the household by parent vaccination status. The unit of observation for this analysis consisted of households in which a parent (the index parent) was infected with SARS-CoV-2. The exposure was the vaccination status of the index parent. The outcome was infection of at least one child in the household at days 3 to 8 after diagnosis of the index parent. To maintain a well-defined point of entry of the infection, we excluded households in which the parent who is not the index parent or a child was diagnosed on days 0 to 2 after diagnosis of the index parent. During the early period, full vaccination was defined as the receipt of two doses at least 7 days prior (compared with no vaccination), and the dominant variant was Alpha. During the late period, full vaccination was defined as receipt of a third dose at least 7 days prior (compared with receipt of only two doses at least 5 months prior), and the dominant variant was Delta. The adjusted estimate was derived from a logistic regression model adjusted for all the household-level characteristics and the vaccination status of the nonindex parent.

Characteristic	Early period	Late period
SAR from a vaccinated or boosted parent (%)	9.0%	9.3%
SAR from an unvaccinated or unboosted parent (%)	24.7%	31.1%
1 – adjusted odds ratio (95% CI)	72.1% (36.6%, 89.3%)	79.6% (55.9%, 91.8%)

Direct Effect
Protection of parents from having been vaccinated

94.4%
(93.2%–95.4%)

Household Infectiousness Effect
Reduction in Risk of transmission from infected parent to unvaccinated child

72.1%
(36.6%–89.3%)

Indirect Effect
Protection of child afforded by parental vaccination

26.0%
(14.0%–36.2%)
One Vaccinated Parent

71.7%
(68.6%–74.6%)
Two Vaccinated Parents
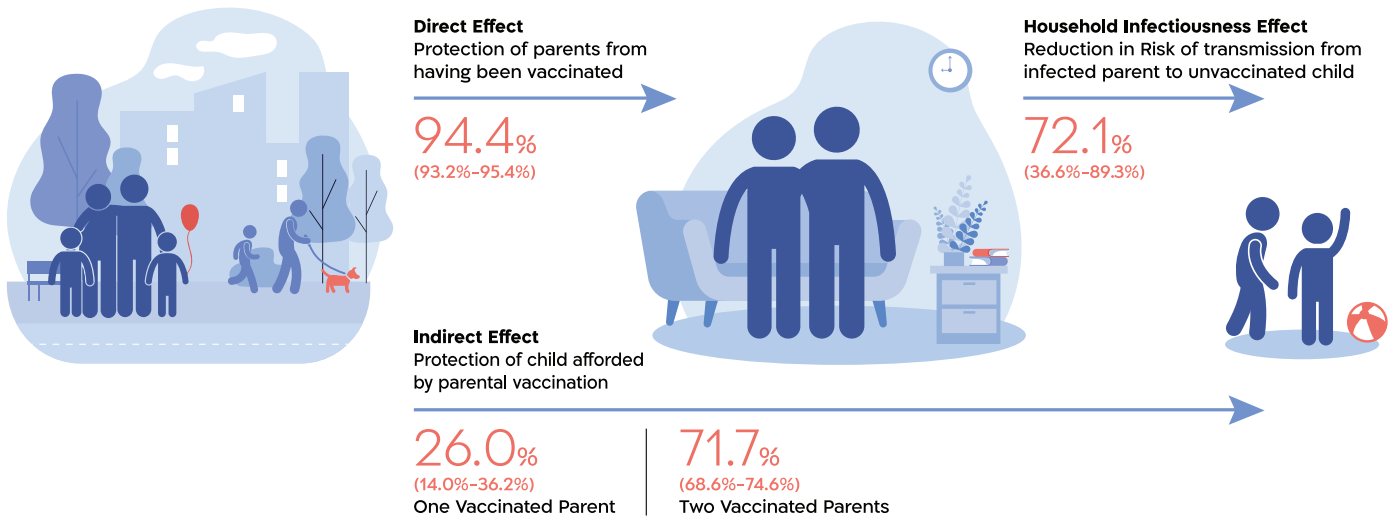
Fig. 2. Mechanism of disease transmission. An illustration showing the indirect effect of parental vaccination on children's risk of SARS-CoV-2 infection and two of its composite parts: the direct effect of vaccination on the parents (estimated as the incidence rate ratio of parental infection between vaccinated and unvaccinated parents) and the risk of transmission from an infected parent to his or her children (estimated as the odds ratio of an infected parent infecting at least one child in the household). We do not expect the indirect risk to equal the product of the direct risk and infectiousness because children may also be infected outside of the household or, potentially, through the other parent. Estimates shown are from the early study period, in which parents vaccinated with two vaccine doses at least 7 days prior were compared with unvaccinated parents and the dominant variant was Alpha.

also shed light on the mechanism through which this protection occurs. These results reinforce the importance of increasing vaccine uptake among the vaccine-eligible population to curb the spread of the SARS-CoV-2 pandemic and protect those who cannot be vaccinated.

REFERENCES AND NOTES

1. N. Zhu et al., N. Engl. J. Med. **382**, 727–733 (2020).
2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), "COVID-19 Dashboard" (2022); https://coronavirus.jhu.edu/map.html.
3. M. E. Halloran, M. Haber, I. M. Longini Jr., C. J. Struchiner, Am. J. Epidemiol. **133**, 323–331 (1991).
4. Z. Wang, W. Ma, X. Zheng, G. Wu, R. Zhang, J. Infect. **81**, 179–182 (2020).
5. K. Danis et al., Clin. Infect. Dis. **71**, 825–832 (2020).
6. I. Dattner et al., PLOS Comput. Biol. **17**, e1008559 (2021).
7. Z. J. Madewell, Y. Yang, I. M. Longini Jr., M. E. Halloran, N. E. Dean, JAMA Netw. Open **4**, e2122240 (2021).
8. F. P. Polack et al., N. Engl. J. Med. **383**, 2603–2615 (2020).
9. N. Dagan et al., N. Engl. J. Med. **384**, 1412–1423 (2021).
10. E. J. Haas et al., Lancet **397**, 1819–1829 (2021).
11. J. Salo et al., medRxiv 2021.05.27.21257896 [Preprint] (2021). https://doi.org/10.1101/2021.05.27.21257896.
12. M. Layan et al., medRxiv 2021.07.12.21260377 [Preprint] (2021). https://doi.org/10.1101/2021.07.12.21260377.
13. O. Milman et al., Nat. Med. **27**, 1367–1369 (2021).
14. Israel Ministry of Health, "Israel COVID-19 Data Tracker" (2022); https://datadashboard.health.gov.il/COVID-19/general.
15. European Centre for Disease Prevention and Control, "COVID-19 Vaccine Tracker" (2022); https://vaccinetracker.ecdc.europa.eu/public/extensions/Covid-19/vaccine-tracker.html.
16. Mayo Clinic, "U.S. COVID-19 Vaccine Tracker" (2022); www.mayoclinic.org/coronavirus-covid-19/vaccine-tracker.
17. T. J. Vanderweele, E. J. Tchetgen Tchetgen, M. E. Halloran, Epidemiology **23**, 751–761 (2012).
18. A. Richterman, E. A. Meyerowitz, M. Cevik, Open Forum Infect. Dis. **9**, ofab259 (2022).
19. N. Barda et al., Lancet **398**, 2093–2100 (2021).
20. M. Lipsitch, E. Tchetgen Tchetgen, T. Cohen, Epidemiology **21**, 383–388 (2010).
21. X. Shi, W. Miao, E. T. Tchetgen, Curr. Epidemiol. Rep. **7**, 190–202 (2020).
22. L. S. Shekerdemian et al., JAMA Pediatr. **174**, 868–873 (2020).
23. J. Toubiana et al., BMJ **369**, m2094 (2020).
24. J. F. Ludvigsson, Acta Paediatr. **110**, 914–921 (2021).
25. A. Mastrangelo et al., Clin. J. Am. Soc. Nephrol. **16**, 449–451 (2021).
26. E. Goldstein et al., Math. Biosci. **221**, 11–25 (2009).
27. E. K. Accorsi et al., Eur. J. Epidemiol. **36**, 179–196 (2021).
28. S. Hayek et al., Indirect protection of children from SARS-CoV-2 infection through parental vaccination, Zenodo (2022); https://doi.org/10.5281/zenodo.5883892.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

MOLECULAR TRANSPORT

Programmable molecular transport achieved by engineering protein motors to move on DNA nanotubes

Ryota Ibusuki[1], Tatsuya Morishita[1], Akane Furuta[2,3], Shintaro Nakayama[1], Maki Yoshio[3], Hiroaki Kojima[3], Kazuhiro Oiwa[1,3], Ken'ya Furuta[3]*

Intracellular transport is the basis of microscale logistics within cells and is powered by biomolecular motors. Mimicking transport for in vitro applications has been widely studied; however, the inflexibility in track design and control has hindered practical applications. Here, we developed protein-based motors that move on DNA nanotubes by combining a biomolecular motor dynein and DNA binding proteins. The new motors and DNA-based nanoarchitectures enabled us to arrange the binding sites on the track, locally control the direction of movement, and achieve multiplexed cargo transport by different motors. The integration of these technologies realized microscale cargo sorters and integrators that automatically transport molecules as programmed in DNA sequences on a branched DNA nanotube. Our system should provide a versatile, controllable platform for future applications.

A living cell can be described as a miniature factory in which diverse molecular machines are continuously working to perform essential life processes. In particular, linear biomolecular motors such as myosin, kinesin, and dynein move along specific cytoskeletal tracks to perform sorting, actuation, and assembly of materials inside eukaryotic cells (1, 2). These protein-based motors exhibit autonomous, long-range transport in cells and thus can potentially be used to manipulate molecules at the micrometer scale in future medicine, material science, and computational applications (3–6). However, the inflexibility in track design and control has hindered practical applications. Actin and tubulin, which compose the cytoskeletal tracks, hydrolyze ATP and GTP, respectively (7, 8), and are extensively posttranslationally modified (9, 10). These properties render cytoskeletal tracks highly dynamic and heterogeneous, which is crucial for diverse cellular processes. Nevertheless, these properties do not favor most applications in which reproducibility and precise placement of tracks are important. By contrast, DNA tracks have several advantages over cytoskeletal tracks. For example, DNA tracks are highly stable and can be accurately designed to form desired three-dimensional structures with control at the single-base level (11, 12). The specificity of DNA sequence recognition is useful for conveying information and programming reactions in a microscale environment, as observed in cells. One option for using a DNA track is to use a "DNA walker" as a motor. Diverse synthetic DNA walkers have been developed to move directionally or diffusively on DNA tracks. Although such DNA walkers have potential, they are currently very slow or adopt a "burnt-bridges" strategy, i.e., their track can only be used once (13–15). In the present study, we developed protein-based motors that move on DNA >100 times faster than corresponding DNA walkers that can reuse their tracks.

We used a modular protein–engineering approach to create hybrid motors that move on DNA nanostructures (16–18). We chose dynein as an engine because it consists of several modules with distinct roles, including hydrolysis of ATP and binding to the track (18–20). We replaced the original track-binding domain of monomeric human cytoplasmic dynein with a DNA binding domain (Fig. 1A). Directional binding to the track was considered to be critical

[1]Graduate School of Life Science, University of Hyogo, Harima Science Park City, Hyogo 678-1297, Japan. [2]Japan Society for the Promotion of Science, Chiyoda-ku, Tokyo 102-0083, Japan. [3]Advanced ICT Research Institute, National Institute of Information and Communications Technology, Kobe, Hyogo 651-2492, Japan.
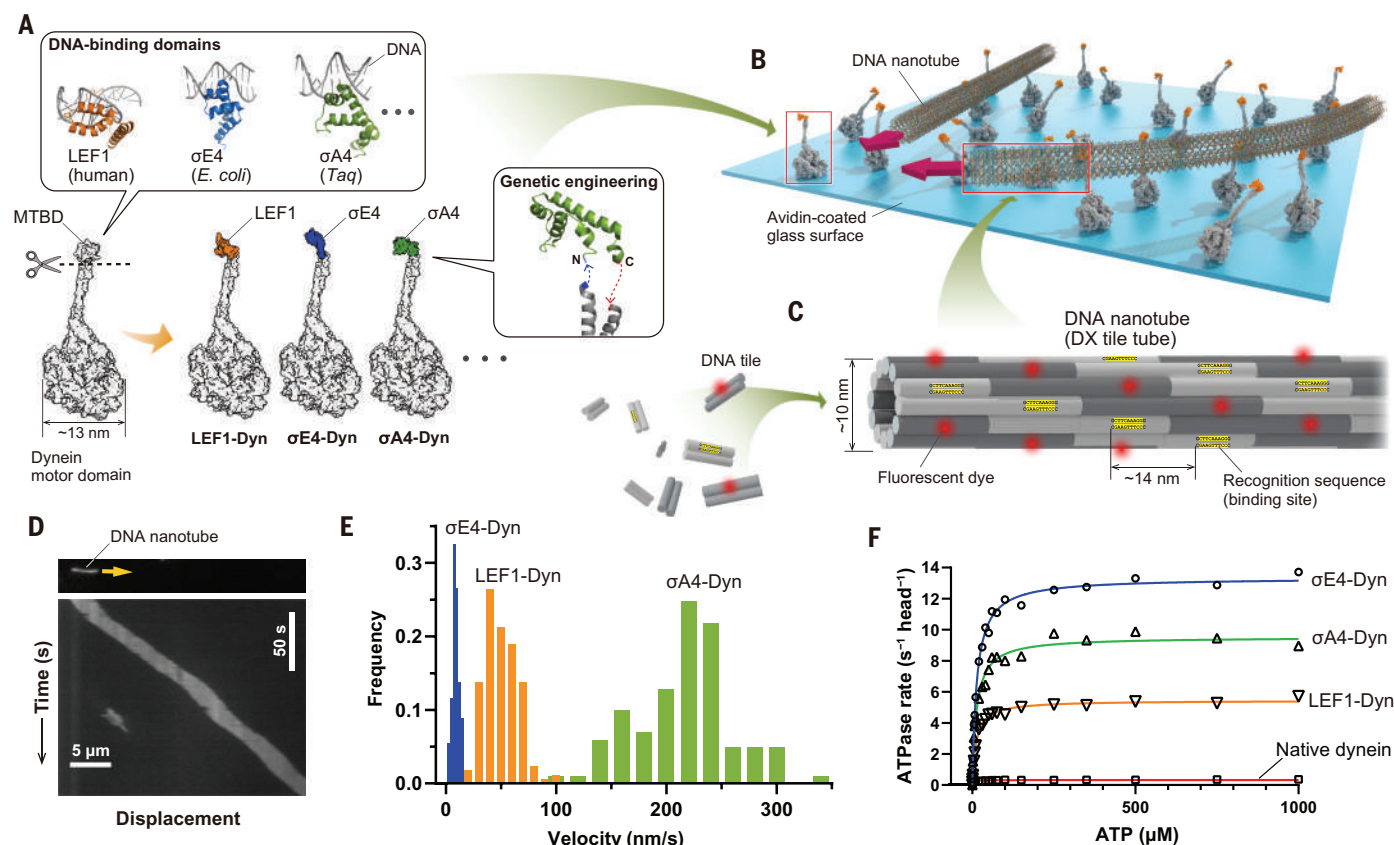*Corresponding author. Email: furutak@nict.go.jp

Fig. 1. DNA-nanotube gliding assay and ATPase measurement of the hybrid motors. (**A**) Construction of the hybrid motors. The illustration is based on the structures of human cytoplasmic dynein (PDB ID: 3VKH), human LEF1 (PDB ID: 2LEF), *E. coli* σE4 (PDB ID: 2H27), and *Taq* σA4 (PDB ID: 1KU7). (**B**) Illustration of the motility assay. (**C**) Schematic representation of the DNA nanotube (DX tile tube). The recognition sequences of the DNA binding domain and fluorescent dye were incorporated along the DNA nanotube. (**D**) Kymograph depicting a representative DNA nanotube (σA4-DX tile v2) movement shown in movie S1. (**E**) Velocity histograms of the DNA nanotube movement driven by the hybrid motors in the presence of 5 mM ATP and 350 mM potassium chloride (KCl). Average velocities (mean ± SEM) were 50 ± 6 nm/s (n = 181 molecules), 9.3 ± 1.7 nm/s (n = 173), 220 ± 16 nm/s (n = 101) for the LEF1-Dyn, σE4-Dyn, and σA4-Dyn motors, respectively. The velocities were calculated from three independent experiments. (**F**) Steady-state ATPase activities of hybrid and native cytoplasmic dynein in the absence of DNA tracks and microtubules, respectively. The plots show a hyperbolic dependence with k_{cat} values of 5.4, 13, 9.5, and 0.32 s^{-1} for the LEF1-Dyn, σE4-Dyn, σA4-Dyn, and native dynein motors, respectively.

in creating a motor that moves unidirectionally along a track. Thus, only DNA binding proteins that recognize nonpalindromic DNA sequences were selected (fig. S1), such as transcription factors (table S1 and figs. S2 to S7). To assess motor activity, we first performed conventional gliding assays in which motors were immobilized on a glass surface and DNA nanotubes glided on them. Eleven out of 19 types of DNA binding domains were determined to mediate the directional movement of DNA nanotube tracks (Fig. 1, B to D; tables S1 and S2; and movies S1 and S2) in a sequence-specific manner (fig. S8, A and B, and movies S3 and S4). The average velocity of DNA nanotubes was highly dependent on the type of hybrid motors, ranging from 5.3 to 220 nm/s (Fig. 1E and fig. S9). ATP hydrolysis by the dynein motor domain is essential for DNA nanotube movement because ATPase-defective mutants of the hybrid motor did not drive gliding movements (figs. S8C and S10 and movie S5).

Moreover, a linear relationship between the DNA-gliding velocity and the ATPase rate was not found (Fig. 1, E and F, and fig. S11), suggesting that other factors such as binding kinetics (figs. S12 and S13) define the gliding velocity.

The overall structure of the DNA tracks can be designed, an attractive feature not currently feasible with cytoskeletal filaments. Two types of DNA tubes were prepared to choose a suitable filament for a motor track: single-stranded tile (SST) tubes (*21*) (figs. S14 to S20) and double-crossover (DX) tile tubes (*22*) (figs. S21 to S34). SST tubes have the advantage of controlling the tube stiffness, whereas DX tile tubes can be used to control the number of motor-binding sites. The stiffness of the DNA nanotubes affected the gliding movement. All types of DNA nanotubes moved directionally on the hybrid motor–coated surface; however, the four-helix SST tubes showed an alternating stretching and shrinking motion,

which was not observed for 10- and 20-helix SST tubes (Fig. 2, A to C; fig. S35; and movies S6 to S8).

Unlike the SST tube, which has only one binding site per cross section, the DX tile tube has multiple binding sites that can be placed in a helical or checkered pattern (Fig. 2D). These tracks with different architectures showed different movements with slightly different optimal ionic strength conditions. The DX tile tube containing three binding sites per section was translocated four times faster compared with the SST tube (Fig. 2, E and F). Presumably, the larger number of binding sites on the DX tile tube facilitated faster recognition of the binding sites by the motor. However, the velocity slowed by a factor of approximately two when the number of binding sites per section was increased to six. As observed for natural biomolecular motors (*23*, *24*), the binding of too many motors can act as a drag on each other (fig. S36). Therefore, DX tile tubes
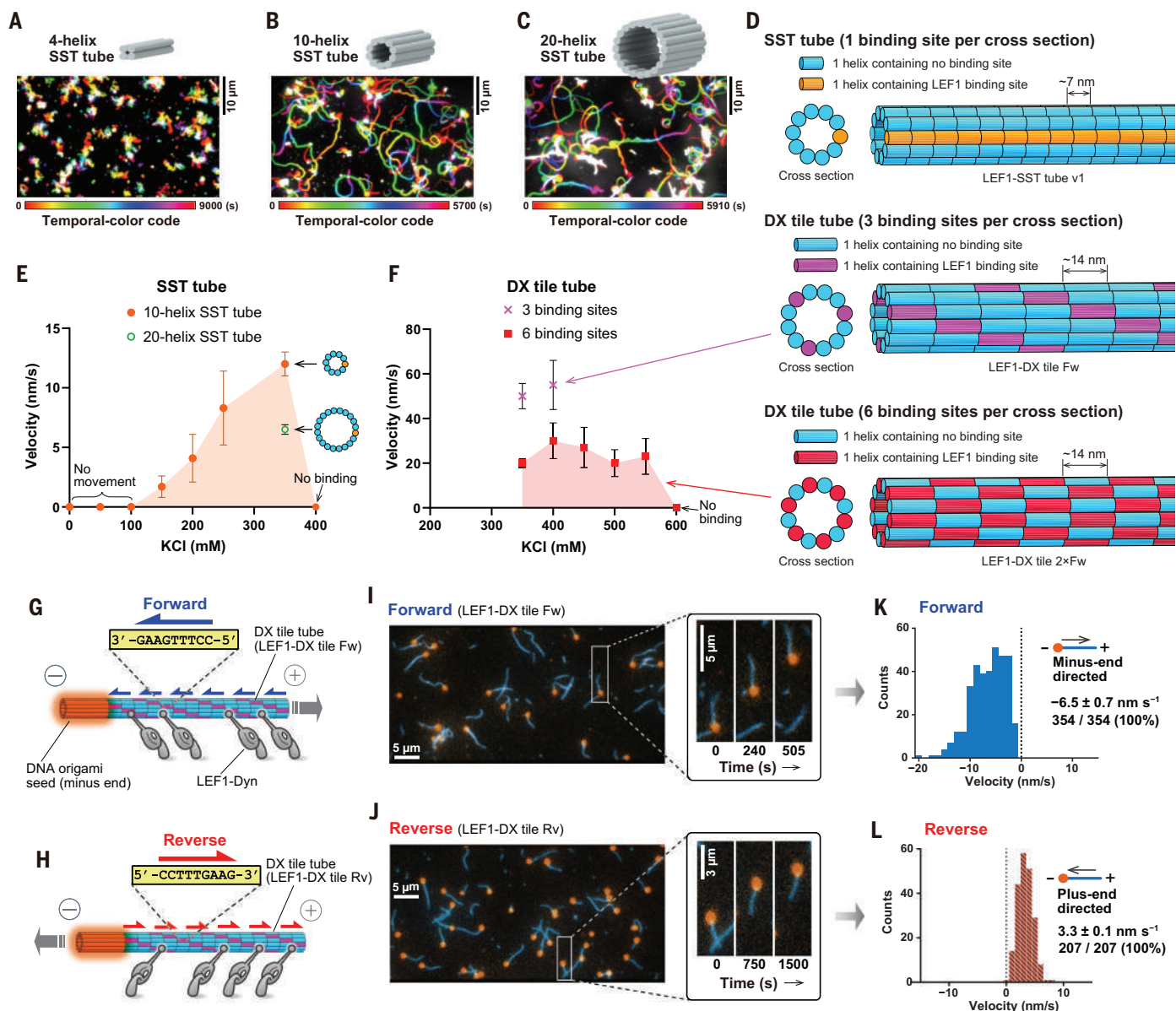
Fig. 2. Effect of the DNA nanotube architecture and control of directionality. (**A** to **C**) Temporal color-coded images depicting the sliding movement of LEF1-SST tube v1 (four helices) (A), LEF1-SST tube v1 (10 helices) (B), and LEF1-SST tube v1 (20 helices) (C) on LEF1-Dyn–coated surfaces. To visualize the gliding movement of DNA nanotubes, the time series images were color coded from red to magenta and superimposed. The color scale is shown at the bottom of each image. (**D**) Diagrams depicting different DNA nanotube architectures. (**E** and **F**) Ionic strength–dependent velocities of SST tubes (E) and DX tile tubes (F). The error bars at 350 mM KCl were calculated from SEM for comparison, and the other error bars represent SD. (**G** and **H**) Schematic representation of DNA origami-seeded DX tile tubes carrying the recognition sequence of LEF1-Dyn. For convenience, the end of the nanotube labeled with the ATTO 647N dye is defined as the minus end and the other end as the plus end. (**I** and **J**) Typical gliding images of DNA nanotubes. (**K** and **L**) Velocity histograms of the gliding DNA nanotubes, LEF1-DX tile Fw (K), and LEF1-DX tile Rv (L). For the forward orientation, for example, 354 out of 354 gliding nanotubes exhibited minus-end directed movement (K). The average velocities (mean ± SEM) were calculated from three independent experiments. Note that the velocity range was different from those obtained in Fig. 1E, probably because of the sticky interactions between the DNA origami seed and the glass surface.

containing three binding sites per section were mainly used in subsequent experiments.

The direction of transport by hybrid motors is programmable in a simple and versatile way, namely by flipping only the orientation of the recognition sequence in a DNA nanotube. Polarity-marked DNA nanotubes (25) marked at one end of the filament with the ATTO 647N dye were transported with the marked end leading on an LEF1-Dyn–coated surface (Fig. 2, G, I, and K, and movie S9), whereas the same polarity-marked DNA nanotubes with flipped recognition sequences moved in the opposite direction, i.e., with the marked end trailing (Fig. 2, H, J, and L; fig. S37; and movie S10). The results showed that the di-

rectionality of the motor is determined by the orientation of the DNA binding domain on the track, enabling control of directionality for molecular transport. Nevertheless, it remains unclear how the motor directionality, velocity, and affinity to DNA tracks are determined. These will be interesting targets for future research because DNA tracks can be systematically
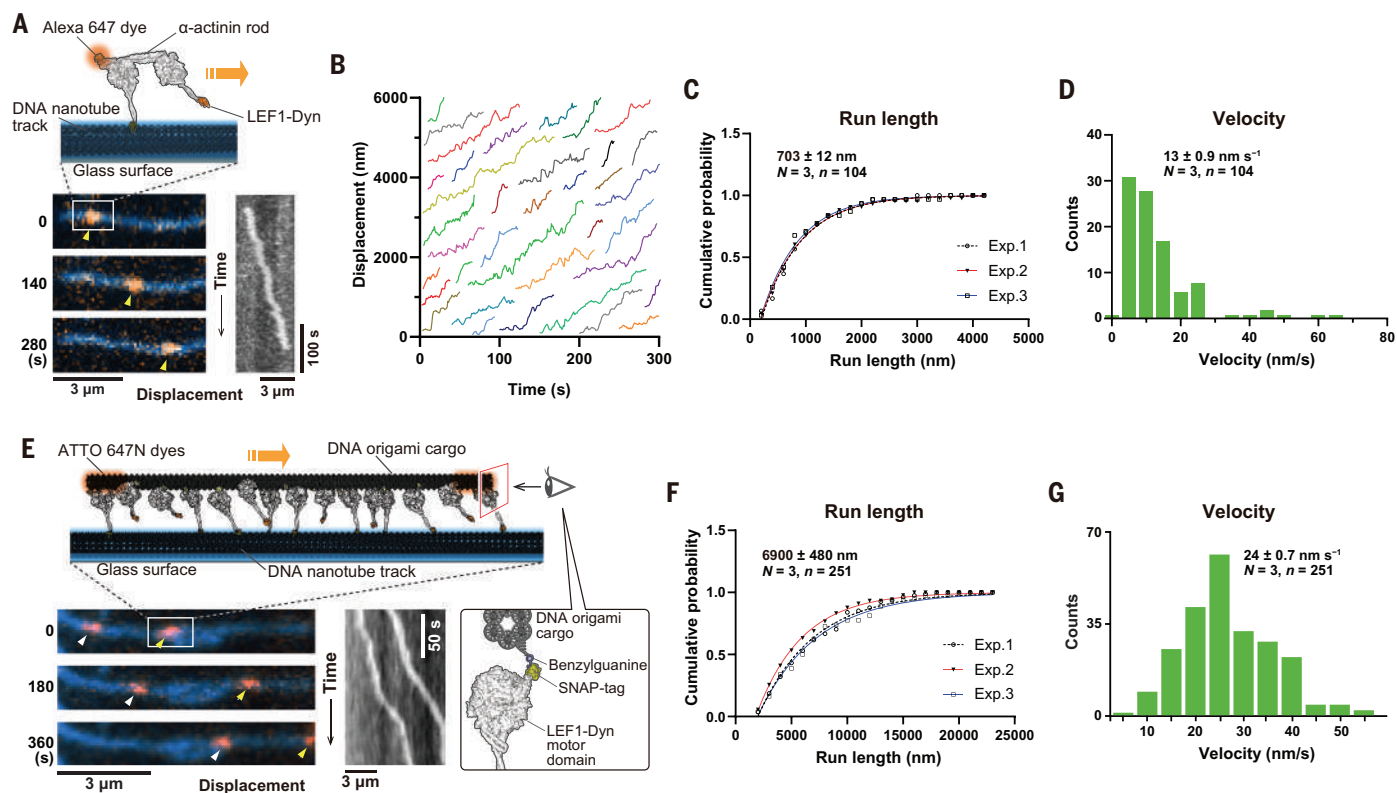
Fig. 3. Single- and multiple-molecule motility assays on DNA nanotube tracks. (**A**) Schematic representation of the ACTNrod-LEF1-Dyn dimer moving on a DNA nanotube track (LEF1-DX tile 2×Fw, top), with sequential frames showing representative single-dimer movement from movie S11 (bottom left), and the kymograph of the same movement (bottom right). The yellow arrowheads indicate the positions of the ACTNrod-LEF1-Dyn dimer shown in orange. (**B**) Example trajectories of ACTNrod-LEF1-Dyn motors along DNA nanotubes. (**C**) Cumulative frequency plots of run lengths of moving spots (5 mM ATP, 125 mM KCl; N = 3 experiments). One-phase exponential decay fits to the individual datasets are shown as lines. (**D**) Velocity histogram of processively moving spots (mean ± SEM, N = 3). (**E**) Schematic representation of 15×BG DNA origami-templated assembly of multiple LEF1-Dyn motors (top), with sequential frames showing representative cargo transport from movie S12 (bottom left), and the kymograph of the same movement (bottom right). DNA origami cargos (orange) were transported along the DNA nanotube track (cyan, LEF1-DX tile Fw). The orientation of the long axis of the DNA origami cargos was aligned with the DNA nanotube track, suggesting that multiple motors on the cargo simultaneously interacted with the track. (**F**) Cumulative frequency plots of run lengths of moving cargos (5 mM ATP, 350 mM KCl; N = 3). One-phase exponential decay fits to the individual datasets are shown as lines. (**G**) Velocity histogram of processively moving cargos (mean ± SEM, N = 3).

modified at single-base resolution to modulate the binding kinetics of hybrid motors.

We then investigated whether the hybrid motor LEF1-Dyn can transport cargo on DNA nanotube tracks. In cells, cytoplasmic dynein is autoinhibited by self-dimerization of the motor domains when free of cargo (*26*, *27*). Therefore, we constructed an artificial dimer that separated the motor domains by the rigid rod domain of α-actinin (ACTNrod-LEF1-Dyn) to release the motor domain from the auto-inhibition. Although the resultant single artificial dimers exhibited directional movement on DNA nanotubes, the movement contained a diffusive component and was not highly processive (~700 nm on average), which is insufficient for micrometer-scale systems (Fig. 3, A to D; fig. S38; and movie S11). To increase the run length, we constructed teams of hybrid motors using a DNA origami scaffold with 15 motor-binding sites (*28*) as a template (*29–32*) (Fig. 3E and figs. S39 to S41). The movement be-

came longer when the hybrid motors were linked together (run length, 6.9 ± 0.5 µm; velocity, 24 ± 0.7 nm/s; Fig. 3, F and G, and movie S12).

Functions such as sorting and integration of molecules on DNA tracks require a junction connecting different DNA tracks. This was achieved by building Y-shaped DNA nanotube tracks that are grown from a Y-shaped DNA origami junction (Y-junction) (*33*).

Initially, we constructed a "disperser" and an "aggregator" (Fig. 4A and fig. S42). In the disperser, the LEF1 recognition sequence was placed so that LEF1-Dyn moved from the center to the periphery of the Y-shaped track. In the aggregator, the direction was set to the opposite. The fluorescently labeled cargos were successfully dispersed or aggregated, as programmed in the DNA tracks (Fig. 4, B and C, and movies S13 and S14).

Next, we set out to achieve a more elaborate system using two different transporters, each recognizing a different DNA sequence on a

single track without cross-talk. We chose the combination of LEF-Dyn and σE4-Dyn because they had virtually no cross-talk and were in the same velocity range. Nevertheless, to work simultaneously with LEF-Dyn, the DNA-binding interface of σE4-Dyn needed to be modified (σE4-Dyn+KR; fig. S5) to improve the affinity for the track, and the number of motors on individual DNA origami cargos was increased to >30 (figs. S41B, S43, and S44).

To build a "sorter" and an "integrator" system (Fig. 4D), we designed new Y-junctions from which three types of DNA nanotubes grow: (i) a nanotube containing the LEF1 recognition sequence (referred to as the "LEF1 track"), (ii) a nanotube containing the σE4 recognition sequence ("σE4 track"), and (iii) a nanotube containing both the LEF1 and σE4 recognition sequences ("LEF1+σE4 track"). The sorter was designed as a branched, Y-shaped track in which the LEF1 and σE4 tracks branched from an LEF1+σE4 track (Fig. 4E and fig. S45). A
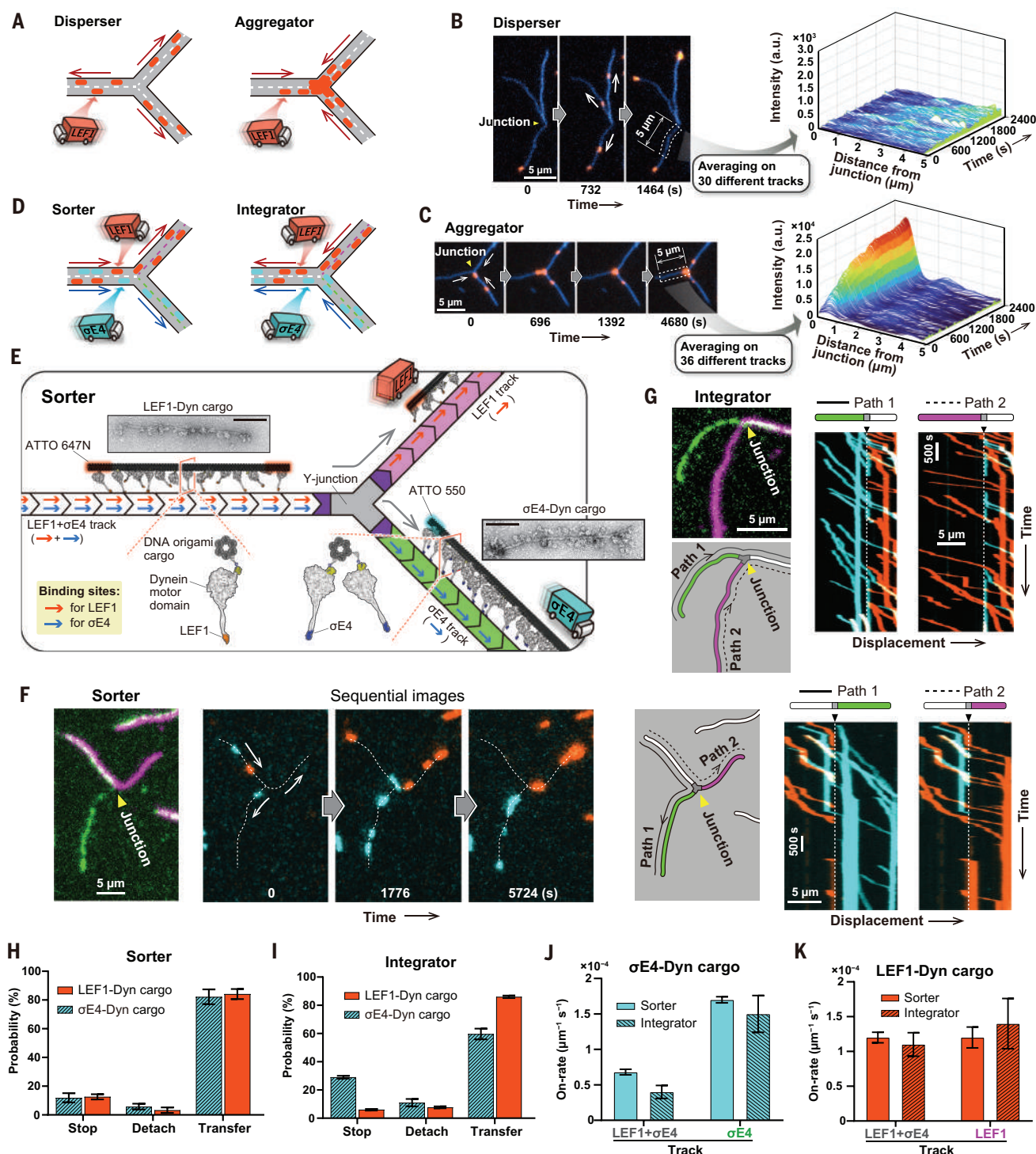
Fig. 4. Molecular transport systems on Y-shaped DNA nanotube tracks.
(**A**) Schematic representations of the disperser and aggregator. (**B** and
C) Sequential frames showing representative behavior of the disperser (B) and
the aggregator (C). The waterfall plot on the right shows how the fluorescence
intensity on the DNA tracks changed over time. (**D**) Schematic representations of
the sorter and integrator. (**E**) Design of the sorter system. The images in the panel
show negative-stain electron micrographs of motor cargo complexes. Scale bars,
100 nm. (**F**) Typical time course of the sorter. Sequential images at the left show
the successful delivery of two different cargos. Multicolor kymographs on the
right depict the movement of LEF1-Dyn cargos (orange) or σE4-Dyn cargos
(cyan) on the LEF1 track (magenta) or σE4 track (green), respectively. Note that

the LEF1+σE4 track appears white because of the overlay of magenta and green.
(**G**) Typical kymograph of the integrator. The image at the left represents the
Y-shaped track. The images at the right represent multicolor kymographs depicting
the movement of LEF1-Dyn cargos (orange) and σE4-Dyn cargos (cyan) on the
LEF1 track (magenta) and σE4 track (green), respectively. (**H** and **I**) The probability
of cargos that stopped or detached from the track at the junction for the sorter
[(H); LEF1, $n = 661$ cargos; σE4, $n = 409$; $N = 3$ experiments] or integrator [(I);
LEF1, $n = 818$; σE4, $n = 979$; $N = 3$]. In the analysis, only cargos that reached
the junction were counted. (**J** and **K**) On-rate at which each motor binds to each
track measured for σE4-Dyn cargos [(J); sorter, $n = 1159$; integrator, $n = 1015$;
$N = 3$] and LEF1-Dyn cargos [(K); sorter, $n = 1055$; integrator, $n = 1023$; $N = 3$].

large number of Y-shaped tracks were observed after immobilization on a glass surface in a flow cell (figs. S46 and S47). Then, the σE4-Dyn and LEF1-Dyn cargo complexes were simultaneously introduced into the flow cell to observe transport on the Y-shaped tracks in the presence of ATP. Both complexes moved on LEF1+σE4 tracks. The σE4-Dyn cargo complexes were generally found only on σE4 tracks, and the LEF1-Dyn cargo complexes were found only on LEF1 tracks (Fig. 4, F and G; fig. S48; and movies S15 to S17). Cross-talk occurred only when the two different DNA origami cargos were nonspecifically bound to each other; i.e., no cross-talk was observed for DNA origami cargos that were transported alone. The rate of cross-talk was $3.9 \pm 0.9\%$ (mean ± SEM, $N = 3$). In the sorter, DNA origami cargos carried by the LEF1-Dyn or σE4-Dyn motors were sorted into LEF1 or σE4 tracks, respectively. The sorter had high specificity; the rate of wrongly sorted cargos was only $3.0 \pm 0.3\%$ (mean ± SEM, $N = 3$). Again, this cross-talk occurred only when two different DNA origami cargos were nonspecifically bound to each other. The integrator is defined as the opposite of the sorter and has a junction where an LEF1 and σE4 track merges into an LEF1+σE4 track (fig. S45). The integrators successfully integrated two types of cargos into a single track (Fig. 4G and movies S18 to S20). In some cases, a motor–origami complex repeatedly changed its orientation at the junction, implying that the motor was searching for its track by Brownian motion (fig. S47E and movie S21). For both the sorter and the integrator, the subsequent behavior of the motor–origami complexes that reached the junction was quantified. The percentage of cargos that stopped or detached from the track at the junction was <20%, except for the σE4-Dyn–origami complexes in the integrator (Fig. 4, H and I). To identify the cause of the inefficiency of σE4-Dyn cargo transfer in the integrator, we measured the binding on-rate of the motor to each track (Fig. 4, J and K). For LEF1-Dyn, there was no difference in the on-rate for the LEF1 and LEF1+σE4 tracks, whereas for σE4-Dyn, the on-rate for the σE4 track was considerably higher compared with that of the LEF1+σE4 track. This difference suggests that it took longer to transfer from a track with a high on-rate to a track with a low on-rate, so the probability of stopping or detaching increased. Modulating the on-rate difference by modifying the sequence of the DNA tracks should enable control of the efficiency of the system (see the supplementary materials, note S49).

The hybrid motors created in this study are engineered protein-based motors that directly attach and move on DNA nanostructures. There are two types of synthetic DNA walkers so far: burnt-bridges motors, which are faster but destroy their tracks as they move (0.5 to 80 nm/s) (14, 34–36), and non–burnt-bridges motors, which are slower but can reuse their tracks (0.01 to 0.1 nm/s) (13). Moreover, there are directional and nondirectional (diffusive) walkers for each type. Our protein-based motor achieves fast walking on DNA (220 nm/s on average) and moves directionally without destroying its track, which should be advantageous for many applications. Biomolecular motors that transport vesicles or operate in large arrays are very fast, ranging from 200 to 60,000 nm/s, whereas a variety of mitotic motors and other transporters move at speeds ranging from 5 to 200 nm/s (2) (see the supplementary materials, note S50; fig. S51). Our motors fall into the low end of the speed range but should suffice to perform various tasks in cells. At the same time, unlike cytoskeletal tracks, our motors allow the flexible design of their track, enabling a highly controllable transport and actuation system.

A large part of DNA nanotechnology research has focused on building three-dimensional structures. Using these structures as scaffolds, researchers have recently organized biomolecular motors on them, controlling the number and geometry of motors (30, 32). Our motors exploit DNA nanostructures as tracks, which provide a fourth dimension, i.e., a dynamic property that changes over time, to static DNA nanostructures. For example, components made of DNA may be rapidly moved to change their shape or propel themselves like biological muscles and flagella do. In previous studies, DNA nanostructures were driven by slow DNA strand displacement reactions (typically on the order of minutes) (37). Much faster movement should be possible if the nanostructures are driven directly by DNA-walking motors.

Further development of sorters and integrators may enable the high-throughput production of peptides or other functional polymers (38), in which motors polymerize the monomers as they walk on a DNA track. DNA can also be used as an information processor by using a hybridization or transcription reaction cascade (39, 40). DNA computing has relied on slow DNA strand displacement but has begun to accelerate by spatially confining reaction paths into a tiny space (40). Combining a fast DNA motor with such reaction circuits may facilitate even faster DNA computing and enable operations in response to various inputs from the environment, creating a "sensor-motor coupling." Because DNA circuits would connect seamlessly with a DNA track–based system and other nucleic acid–based controls (31, 41), their combination could eventually realize microscopic robots or microswimmers that accept external signals, harvest energy, process information, and affect their environments.

REFERENCES AND NOTES

1. R. D. Vale, *Cell* 112, 467–480 (2003).
2. J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, 2001).
3. T. Korten, A. Månsson, S. Diez, *Curr. Opin. Biotechnol.* 21, 477–488 (2010).
4. H. Hess, *Annu. Rev. Biomed. Eng.* 13, 429–450 (2011).
5. A. Goel, V. Vogel, *Nat. Nanotechnol.* 3, 465–475 (2008).
6. D. V. Nicolau Jr. et al., *Proc. Natl. Acad. Sci. U.S.A.* 113, 2591–2596 (2016).
7. J. R. McIntosh, V. Volkov, F. I. Ataullakhanov, E. L. Grishchuk, *J. Cell Sci.* 123, 3425–3434 (2010).
8. T. D. Pollard, G. G. Borisy, *Cell* 112, 453–465 (2003).
9. S. S. Prassanawar, D. Panda, *Biochem. J.* 476, 1359–1376 (2019).
10. P. Vedula, A. Kashina, *J. Cell Sci.* 131, jcs215509 (2018).
11. N. C. Seeman, *Annu. Rev. Biochem.* 79, 65–87 (2010).
12. S. M. Douglas et al., *Nature* 459, 414–418 (2009).
13. J. Pan, F. Li, T. G. Cha, H. Chen, J. H. Choi, *Curr. Opin. Biotechnol.* 34, 56–64 (2015).
14. T. G. Cha et al., *J. Am. Chem. Soc.* 137, 9429–9437 (2015).
15. M. Liu et al., *ACS Nano* 10, 5882–5890 (2016).
16. G. Tsiavaliaris, S. Fujita-Becker, D. J. Manstein, *Nature* 427, 558–561 (2004).
17. L. Chen, M. Nakamura, T. D. Schindler, D. Parker, Z. Bryant, *Nat. Nanotechnol.* 7, 252–256 (2012).
18. A. Furuta et al., *Nat. Nanotechnol.* 12, 233–237 (2017).
19. A. P. Carter et al., *Science* 322, 1691–1695 (2008).
20. S. Can, S. Lacey, M. Gur, A. P. Carter, A. Yildiz, *Nature* 566, 407–410 (2019).
21. P. Yin et al., *Science* 321, 824–826 (2008).
22. P. W. Rothemund et al., *J. Am. Chem. Soc.* 126, 16344–16352 (2004).
23. R. S. Rock, M. Rief, A. D. Mehta, J. A. Spudich, *Methods* 22, 373–381 (2000).
24. T. Shima, T. Kon, K. Imamula, R. Ohkura, K. Sutoh, *Proc. Natl. Acad. Sci. U.S.A.* 103, 17736–17740 (2006).
25. A. M. Mohammed, R. Schulman, *Nano Lett.* 13, 4006–4013 (2013).
26. T. Torisawa et al., *Nat. Cell Biol.* 16, 1118–1124 (2014).
27. K. Zhang et al., *Cell* 169, 1303–1314.e18 (2017).
28. H. Bui et al., *Nano Lett.* 10, 3367–3372 (2010).
29. K. Furuta et al., *Proc. Natl. Acad. Sci. U.S.A.* 110, 501–506 (2013).
30. N. D. Derr et al., *Science* 338, 662–665 (2012).
31. T. Omabegho et al., *Nat. Nanotechnol.* 13, 34–40 (2018).
32. R. F. Hariadi et al., *Nat. Nanotechnol.* 10, 696–700 (2015).
33. T. D. Jorgenson, A. M. Mohammed, D. K. Agrawal, R. Schulman, *ACS Nano* 11, 1927–1936 (2017).
34. A. Bazrafshan et al., *Angew. Chem. Int. Ed.* 59, 9514–9521 (2020).
35. K. Yehl et al., *Nat. Nanotechnol.* 11, 184–190 (2016).
36. J. Li et al., *Nat. Nanotechnol.* 13, 723–729 (2018).
37. J. Song et al., *Science* 357, eaan3377 (2017).
38. B. Lewandowski et al., *Science* 339, 189–193 (2013).
39. G. Gines et al., *Nat. Nanotechnol.* 12, 351–359 (2017).
40. G. Chatterjee, N. Dalchau, R. A. Muscat, A. Phillips, G. Seelig, *Nat. Nanotechnol.* 12, 920–927 (2017).
41. A. J. Wollman, C. Sanchez-Cano, H. M. Carstairs, R. A. Cross, A. J. Turberfield, *Nat. Nanotechnol.* 9, 44–47 (2014).

MAGNETISM

Frequency multiplication by collective nanoscale spin-wave dynamics

Chris Koerner[1], Rouven Dreyer[1], Martin Wagener[1,2], Niklas Liebing[1], Hans G. Bauer[3], Georg Woltersdorf[1]*

Frequency multiplication is a process in modern electronics in which harmonics of the input frequency are generated in nonlinear electronic circuits. Devices based on the propagation and interaction of spin waves are a promising alternative to conventional electronics. The characteristic frequency of these excitations is in the gigahertz (GHz) range and devices are not readily interfaced with conventional electronics. Here, we locally probe the magnetic excitations in a soft magnetic material by optical methods and show that megahertz-range excitation frequencies cause switching effects on the micrometer scale, leading to phase-locked spin-wave emission in the GHz range. Indeed, the frequency multiplication process inside the magnetic medium covers six octaves and opens exciting perspectives for spintronic applications, such as all-magnetic mixers or on-chip GHz sources.

Frequency multiplication in modern electronics is usually achieved in nonlinear electronic circuits or transmission lines. In recent years, spin-based electronics has emerged as a promising extension to conventional electronics, providing new functionalities such as nonvolatile memory (1, 2). In such spin-based devices, the spin transport and spin dynamics in the gigahertz (GHz) frequency range are inherently coupled, as demonstrated by the spin torque effect. For many applications, the possibility of frequency up-conversion from megahertz (MHz) frequencies using the nonlinearity of the magnetization dynamics would be desirable but has been restricted typically to a few harmonics only (3–5). Nonperturbative frequency multiplication processes over several octaves were only observed at optical frequencies in response to extremely intense laser pulses (6). In most cases, nonlinear behavior either emerges when an external driving force exceeds a certain threshold or is caused by the intrinsic nonlinearity of the equation of motion. In the case of spin waves, major attention has been directed to the threshold behavior with respect to the radio frequency (rf) pumping field, showing that large excitation amplitudes drive nonlinear spin-wave modes (7, 8). These nonlinear excitations are utilized in spintronic devices such as memory applications (9, 10) and spin transfer torque-driven nano-oscillators (11–13).

We explore the nonlinear response of a soft ferromagnet at very small magnetic fields. Spin waves are probed by using diamond nitrogen vacancy (NV) center nanoscale magnetometry (14–18) and time-resolved magneto-optic Kerr microscopy (19). In addition to spin waves predicted by a generalized theory of spin-wave turbulence oscillating at 3/2 of the driving frequency (20), we observe a series of spin-wave excitations precessing at up to the 60th harmonic of the pumping frequency. This phenomenon can be understood as the result of dynamic, periodic, and synchronized switching of magnetic texture.

In our experiments, a ferromagnetic (FM) $Ni_{80}Fe_{20}$ layer is deposited on top of an Au

[1]Department of Physics, Martin Luther University Halle-Wittenberg, Von-Danckelmann-Platz 3, 06120 Halle, Germany. [2]Institute for Quantum Electronics, ETH Zürich, Otto-Stern-Weg 1, 8093 Zürich, Switzerland. [3]Jahnstrasse 23, 96050 Bamberg, Germany.
*Corresponding author. Email: georg.woltersdorf@physik.uni-halle.de
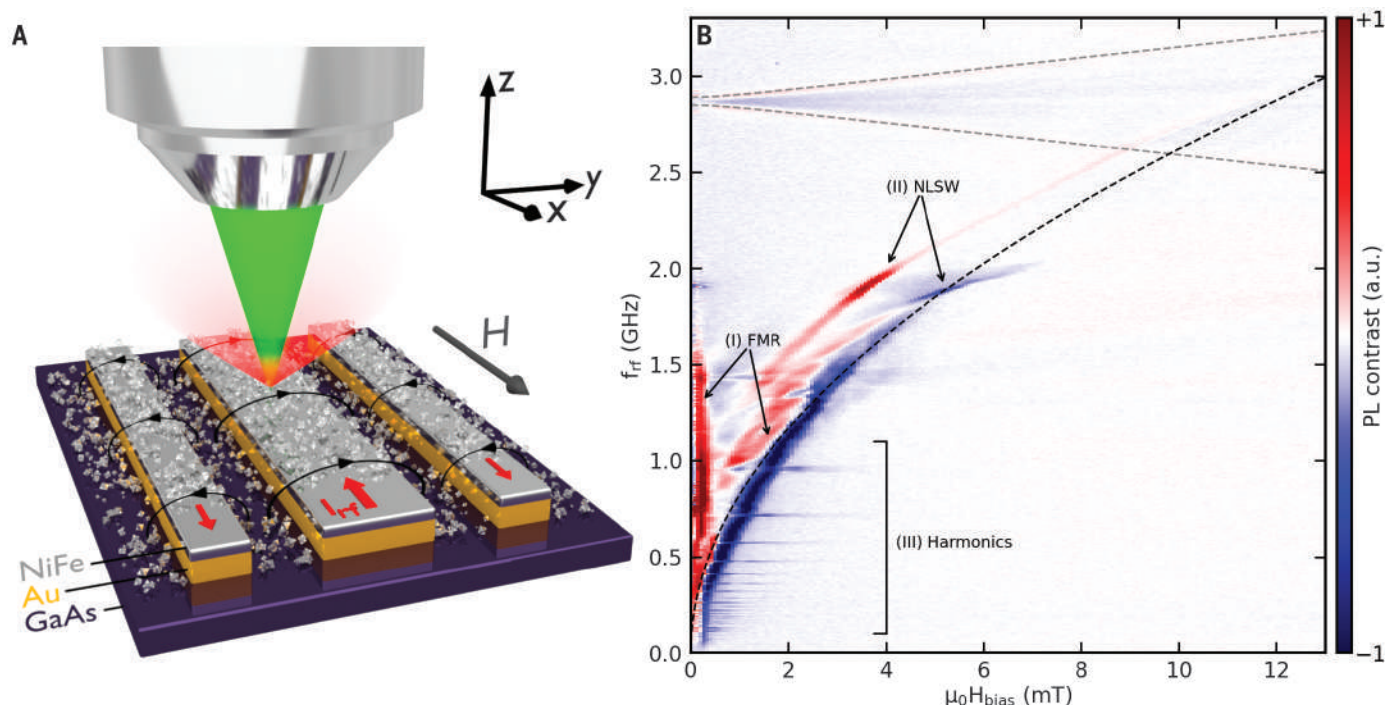
Fig. 1. NV center detection of spin waves. (**A**) Measurement geometry. We investigate a 20-nm-thick ferromagnetic $Ni_{80}Fe_{20}$ layer on top of a coplanar waveguide (CPW) by means of nanodiamonds containing NV centers dispersed on the sample, which act as local stray field sensors. (**B**) Photoluminescence (PL) contrast as a function of bias field and excitation frequency. The dashed gray lines indicate the calculated Zeeman-split electron spin resonance (ESR) cone of the NV ensemble. The adjacent ferromagnet introduces additional PL signals: (I) FMR, (II) nonlinear spin waves at 2/3 of f_{ESR}, and (III) a series of replicas of the ESR at low magnetic bias fields and frequencies. The dashed black line traces the calculated $Ni_{80}Fe_{20}$ FMR frequency.

coplanar waveguide (CPW) and covered with nanodiamonds containing NV centers (14, 17, 21) (Fig. 1A). The rf magnetic fields generated by the CPW excite the electron spin resonance (ESR) in diamond NV centers as well as magnetization dynamics in the FM layer. In NV center magnetometry, the ESR of the NV centers is optically detected by monitoring their photoluminescence (PL), i.e., optical detection of magnetic resonance (ODMR) (22, 23), as explained in section M2 of materials and methods. The measured PL intensity modulation due to microwave excitation recorded from the NV center ensemble is shown as a function of static bias field and the rf excitation frequency f_{rf} in Fig. 1B. Whenever applied magnetic field and rf frequency coincide with the ESR condition, a change in the PL signal is observed. In addition to the expected ESR signal of the individual NV centers (dotted lines in Fig. 1B), we find a number of additional features in the ODMR signal caused by the proximity of the FM layer. These may be characterized as follows: (I) a continuous resonance at and below the ferromagnetic resonance (FMR) field, (II) a strong enhancement at 2/3 of the NV centers resonance frequency, and (III) a series of resonance lines at low bias fields. The occurrence of the FMR in the ODMR signal (I) has been observed before (15–17, 24), and the effect is attributed to dipolar stray fields generated by spin waves

acting on the NV centers (16). Feature (II) occurs at a frequency of 2/3 f_{ESR}, i.e., at 1.9 GHz. This feature is attributed to nonlinear spinwave excitations as predicted for small bias fields by Bauer et al. (20) (for further details, see fig. S1). For magnetic fields below 4 mT, a series of narrow resonance lines whose spacing scales down with decreasing frequency (Fig. 2) (feature III) is observed. All of these resonances are very sharp replicas of the ESR cone of the NV centers (around 2.87 GHz) occurring at the nth fraction of the ESR frequency, where n is an integer. The observed Zeeman splitting of these replicas is proportional to the bias field and scales inversely with n. In the inset of Fig. 2, more than 60 replicas of the ESR are observed with a signal amplitude comparable to that of the main ESR signal of the NV ensemble and only weakly decaying with increasing n. The replicated ESR lines in the PL signal are a consequence of the presence of the FM layer (fig. S2), and their origin must be attributed to its dynamic response. Because the NV centers are sensitive only to a narrow frequency range around f_{ESR}, the observed $1/n$-spaced sharp lines can be considered a fingerprint of a frequency comb, which induces a signal whenever a comb line coincides with f_{ESR}. For example, at an excitation frequency of f_{rf} = 287 MHz, the 10th comb line of a frequency comb matches f_{ESR}, leading to the observed signal (Fig. 2; n = 10).

The observation of such a frequency comb is not at all expected and deserves close attention.

To investigate the spin-wave modes responsible for the frequency comb, we directly map the dynamic response at a specific harmonic of the excitation using super-Nyquist-sampling MOKE (SNS-MOKE) (25). By addressing the individual harmonics of the frequency comb (expected from the NV center results), we observe phase-stable spin-wave patterns. In Fig. 3, spatially resolved measurements of magnitude and phase of the dynamic magnetization are shown that represent the 1st, 10th, and 20th harmonic. Here, an almost homogeneous response at the fundamental f_{rf} and an increasingly fragmented spatial modulation in the micrometer and submicrometer range is found for higher harmonics (all harmonics are shown in fig. S5). The frequency comb generation is most pronounced at magnetic bias fields below 2 mT.

To understand the physical mechanism responsible for this peculiar behavior, we perform micromagnetic simulations. The most simple case, a periodically excited homogeneous magnetic layer, does not reproduce the experiments, as only a few odd harmonics of the excitation frequency with rapidly decaying amplitude are observed for the z-component of the magnetization, as shown for f_{rf} = 191 MHz in Fig. 4A (black curve). This result is not surprising because, based on the equation of motion of the magnetization and its intrinsic nonlinearity, one indeed expects just a few odd harmonics with rapidly decreasing amplitude. However, in the experiments, both even and odd multiples of the excitation frequency occur, extending up to the 60th harmonic. Therefore, a crucial aspect must be missing in the discussion.

Because the frequency comb is only observed at very small fields, where the magnetization may not be fully saturated, we map the magnetic texture of the sample close to zero magnetic field using static MOKE (fig. S8). At very small fields (below 1 mT), we find that the magnetization of the $Ni_{80}Fe_{20}$ stripe shows a magnetic ripple structure. This is a well-known effect that occurs in soft magnetic materials such as $Ni_{80}Fe_{20}$, and it is caused by the interplay between internal magnetic fields—e.g., due to strain or polycrystalline grain structure—and stray field energy, leading to a buckling structure of the magnetization known as a ripple pattern in extended layers or concertina patterns in stripes (26). When the simulation includes the magnetic stray fields from the edges and a small randomly oriented anisotropy field (reflecting the polycrystalline sample structure), the micromagnetic simulation well reproduces the observed static ripple pattern; fig. S9A. When this texture is used, the simulated frequency spectrum (Fig. 4A) fully reproduces the experimental result and shows odd
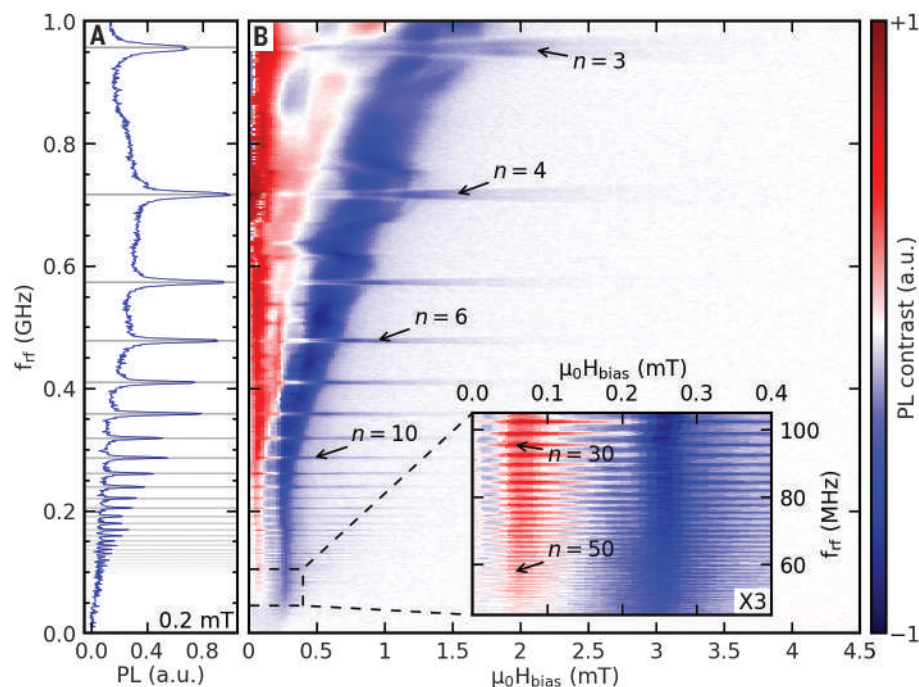

Fig. 2. ODMR detection of frequency comb generation at low–magnetic bias field. (**A**) scan of the excitation frequency at constant bias field of 0.5 mT, showing multiple signals at integer $1/n$ fractions of the ESR frequency (2.87 GHz). The gray lines indicate peak positions. (**B**) Frequency- and field- resolved map of the PL signal. More than 50 harmonics of the driving frequency are detected (inset). The contrast in the inset is scaled by a factor of 3. The driving rf power is 16 dBm. Both panels are obtained with different modulation methods, as discussed in materials and methods M2.
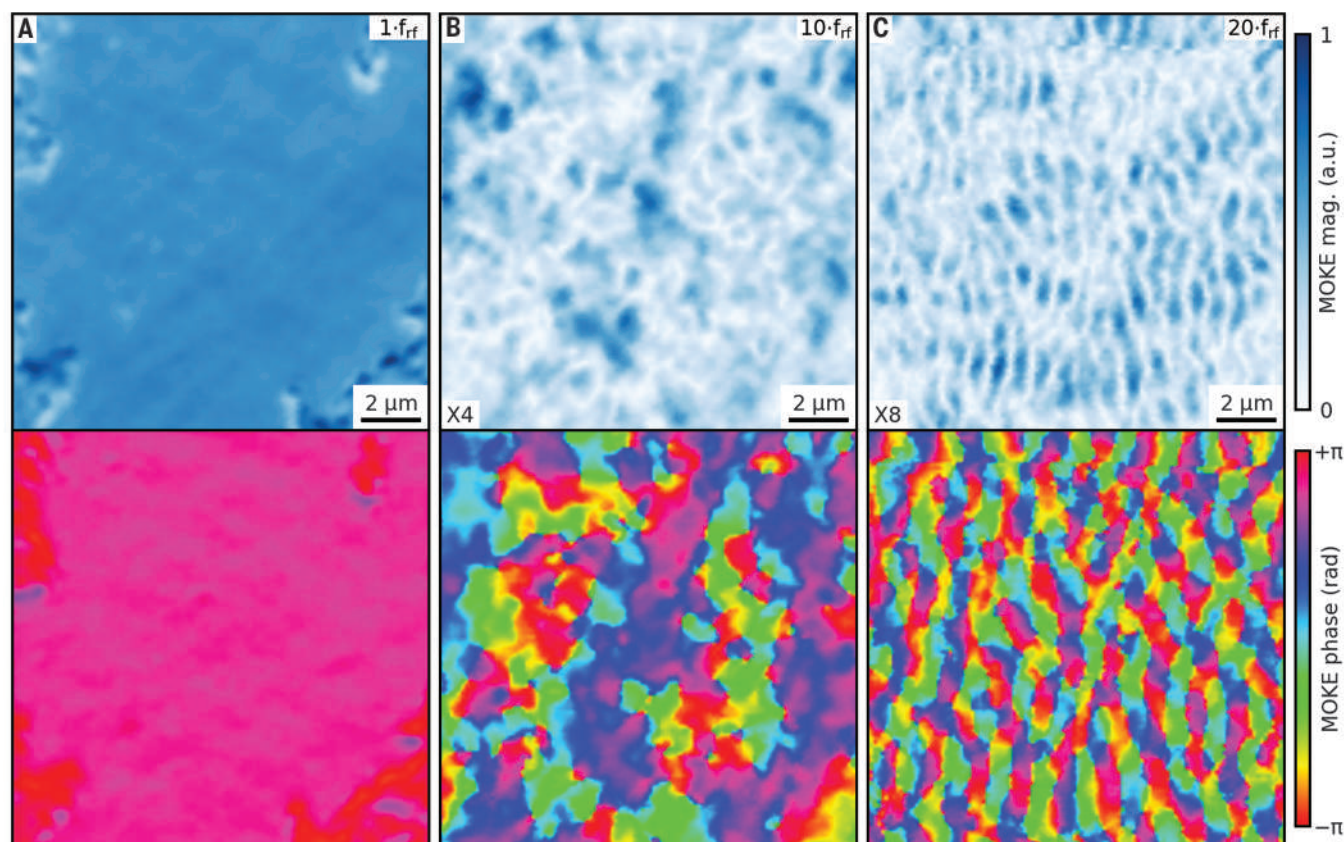
Fig. 3. SNS-MOKE measurements. Spatially resolved SNS-MOKE magnitude (upper panels) and phase (lower panels). All measurements were obtained at 0.5-mT bias field with a fixed excitation frequency of 191.4 MHz at 19 dBm while the detection frequency was set to the (**A**) 1st, (**B**) 10th, and (**C**) 20th harmonic of the excitation, respectively. For better visibility, the contrast is scaled by a factor of 4 or 8, as denoted in the lower left corner.

and even harmonics with orders-of-magnitude larger amplitude. Analyzing this simulation in detail, we find that local switching events of the transverse magnetization components of the ripple pattern cause the emission of spin waves (fig. S9C and movie S1). Next, to simplify the discussion, we only consider a single step in magnetization density perpendicular to the applied field (Fig. 4C). Here, the stray field energy due the magnetic surface charges is minimized by a spontaneous tilt of the magnetization at the boundary. Two energetically equivalent states (transverse component up or down) are possible, as illustrated in Fig. 4, D and E. In this situation, a low-frequency (MHz range) rf excitation can cause switching of the magnetization between the two states at the excitation frequency. To demonstrate that this simple model indeed fully captures the essential physical processes, we simulate its dynamics. A corresponding snapshot of the magnetization dynamics during this process is shown in Fig. 4B.

Because of the open boundary conditions (materials and methods M4), the switching process starts at one edge and propagates along the boundary (Fig. 4B, and movie S3). As shown in fig. S11, the spin waves are emitted from the fast-propagating switching front. Here,

the transverse magnetization component rapidly changes from $+0.4\,M_S$ to $-0.4\,M_S$ on a time scale of less than 100 ps, where M_S is the saturation magnetization. Because the switching front moves at a speed of 4 km/s, and thus three to four times faster than the spin-wave phase velocity (in the GHz range), the spin waves pile up and form a shock front analogous to a supersonic cone moving across the sample. This phenomenon is known as the spin-wave Cherenkov effect (27). We conclude that in the actual sample, with a spatial distribution of switching regions, the emitted spin-wave shock fronts from one region can directly trigger the switching in the neighboring regions and thereby synchronize the emitted spin waves. This effect is known as spin wave–assisted switching (28).

Overall, this process leads to the observed coherent standing spin-wave amplitude pattern obtained by SNS-MOKE measurements at all harmonics phase locked to the MHz-range driving field. This behavior is observed in simulations and experiments. An example of such behavior is shown in a phase-resolved fashion for the 18th harmonic in movie S2. The result explains both the observation of ESR replicas in ODMR measurements and the higher harmonics directly observed by

SNS-MOKE. However, in contrast to ODMR, SNS-MOKE is only sensitive to fully repeatable coherent dynamics. We emphasize that the observed behavior is extremely robust. ODMR and SNS-MOKE measurements show that the frequency comb is generated at static magnetic fields of up to 2 mT and with an exceptionally broad range of driving frequencies, from 50 MHz up to 1 GHz (Fig. 2). The range of generated spin waves is potentially even wider, but owing to limitations of our experimental setup, is not accessible. The simulations show that the frequency comb generation is largely independent of the details of the magnetic texture (compare figs. S14 and S13) and easily extends to above 25 GHz (see fig. S12).

Our results demonstrate that a simple layer of ferromagnetic material (in our case $Ni_{80}Fe_{20}$) can be used to generate a frequency comb spanning six octaves. Using time-resolved Kerr microscopy, we correlate this frequency-multiplying behavior to the emergence of a collective dynamic phase in which the magnetization of the ferromagnet partially switches in a coherent and synchronized fashion due to spin-wave coupling effects. As we demonstrate by magneto-optic imaging (Fig. 3), a standing spin-wave pattern establishes itself at all
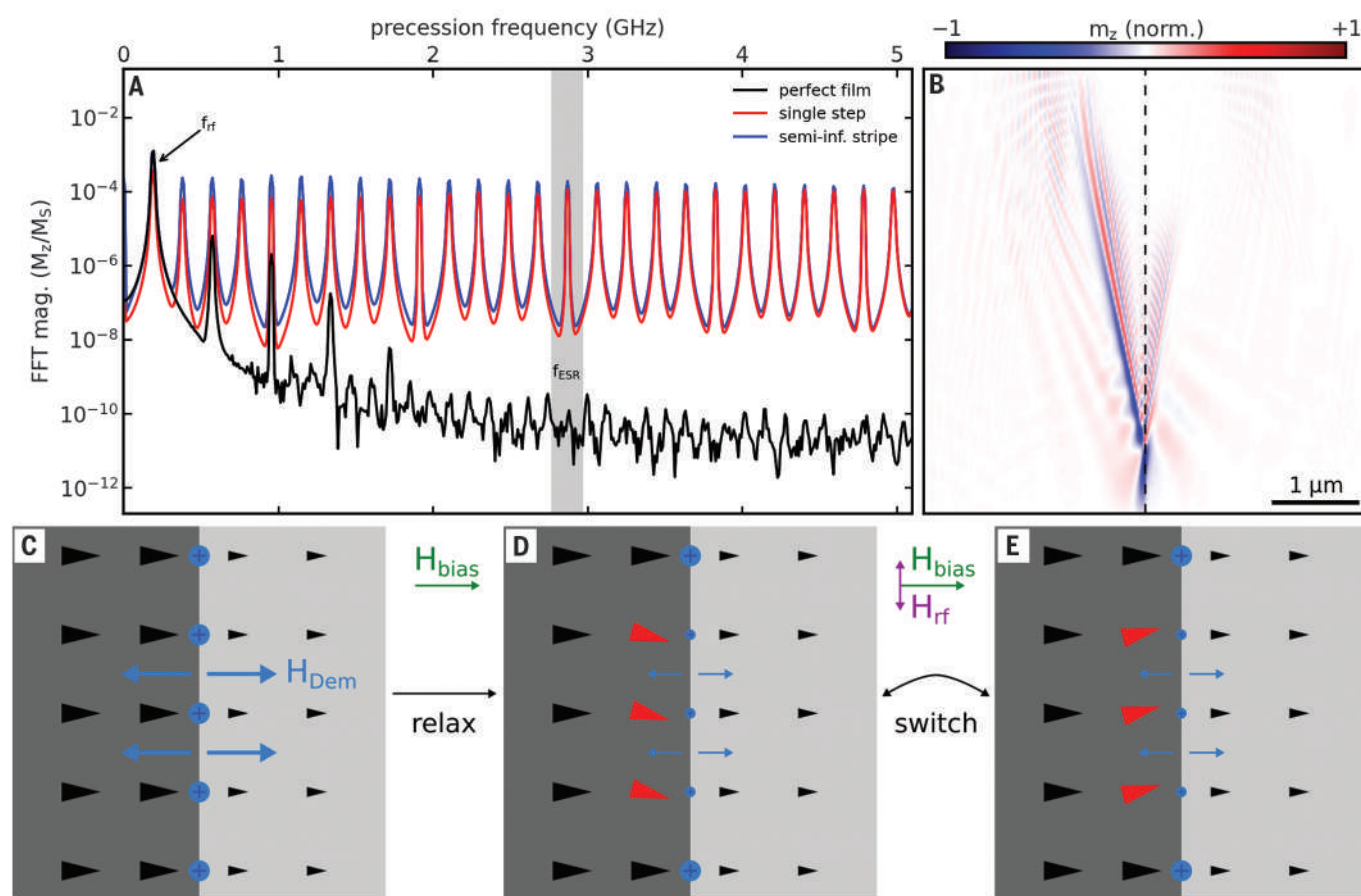
Fig. 4. Micromagnetic simulations and switching model. (**A**) Fourier spectrum obtained from micromagnetic simulations of different FM film textures excited at f_{rf} = 191.4 MHz, a homogeneous film (black), a stripe with two regions where M_S differs by 20% (red), and a 25-μm-wide semi-infinite stripe with 1-μm-sized grains with randomly varying anisotropy direction (blue). The gray area indicates the frequency f_{ESR} to which NV centers are sensitive. (**B**) Snapshot of m_z of the two-region simulation, revealing a sharp spin shockwave emitted during switching of the magnetization upon rf excitation. The location of the step is indicated by the dashed line. All simulations are carried out at 0.5-mT static bias field and a perpendicular 700-mT peak rf excitation field. (**C** to **E**) Simplified model of the dynamic behavior at a boundary with rf excitation.

harmonics of the excitation frequency within the spin-wave band. The excited spin waves preferentially follow a linear dispersion (fig. S6). This behavior is expected as it allows for phase coherence of all spin waves between the localized emission centers, resulting in a standing spin-wave pattern. The coherent nature of the high harmonic response also implies that the line width of the individual comb lines is determined by the signal source used for pumping and is not related to spin-wave damping.

From a fundamental point of view, the effect of phase-stable pattern formation at many harmonics as discussed in this Report represents a self-synchronization phenomenon. Such effects occur in a large variety of nonlinear physical systems. In the present case, the rf switching events lead to spin-wave emission that mutually synchronizes the switching events and leads to the formation of a phase-stable time-periodic spin-wave pattern.

The presented possibility of frequency multiplication under continuous-wave conditions opens exciting perspectives for spin-based applications operating in the GHz range while conveniently being controlled with MHz input signals. In doing so, specific harmonic frequencies may be selected either by introducing a periodicity in the magnetic structure, resonantly enhancing the desired magnon mode, or by the initial excitation of all harmonics with subsequent separation. Here, the different k-vectors may be exploited, e.g., by coupling the spin waves into a tailored waveguide (*29*) or by using nanoscale narrow-band Fabry-Pérot spin-wave filters, which we recently demonstrated (*19*).

REFERENCES AND NOTES
1. J. Åkerman, *Science* **308**, 508–510 (2005).
2. S. S. Parkin, M. Hayashi, L. Thomas, *Science* **320**, 190–194 (2008).
3. S. J. Hermsdoerfer et al., *Appl. Phys. Lett.* **94**, 223510 (2009).
4. V. E. Demidov et al., *Phys. Rev. B Condens. Matter Mater. Phys.* **83**, 054408 (2011).
5. D. R. Rodrigues et al., *Phys. Rev. Appl.* **16**, 014020 (2021).
6. O. Schubert et al., *Nat. Photonics* **8**, 119–123 (2014).
7. H. Suhl, *J. Phys. Chem. Solids* **1**, 209–227 (1957).
8. H. Ulrichs, V. E. Demidov, S. O. Demokritov, S. Urazhdin, *Phys. Rev. B Condens. Matter Mater. Phys.* **84**, 094401 (2011).
9. J. M. Daughton, *J. Appl. Phys.* **81**, 3758–3763 (1997).
10. J.-M. Hu, Z. Li, L.-Q. Chen, C.-W. Nan, *Nat. Commun.* **2**, 553 (2011).
11. S. Kaka et al., *Nature* **437**, 389–392 (2005).
12. V. E. Demidov et al., *Nat. Commun.* **5**, 3179 (2014).
13. M. Haidar et al., *Nat. Commun.* **10**, 2362 (2019).
14. C. S. Wolfe et al., *Phys. Rev. B Condens. Matter Mater. Phys.* **89**, 180406 (2014).
15. C. Du et al., *Science* **357**, 195–198 (2017).
16. C. M. Purser et al., *Appl. Phys. Lett.* **116**, 202401 (2020).
17. B. A. McCullian et al., *Nat. Commun.* **11**, 5229 (2020).
18. I. Bertelli et al., *Sci. Adv.* **6**, eabd3556 (2020).
19. H. Qin et al., *Nat. Commun.* **12**, 2293 (2021).
20. H. G. Bauer, P. Majchrak, T. Kachel, C. H. Back, G. Woltersdorf, *Nat. Commun.* **6**, 8274 (2015).
21. A. M. Schrand, S. A. C. Hens, O. A. Shenderova, *Crit. Rev. Solid State Mater. Sci.* **34**, 18–74 (2009).
22. G. Balasubramanian et al., *Nature* **455**, 648–651 (2008).
23. R. Schirhagl, K. Chang, M. Loretz, C. L. Degen, *Annu. Rev. Phys. Chem.* **65**, 83–105 (2014).
24. T. van der Sar, F. Casola, R. Walsworth, A. Yacoby, *Nat. Commun.* **6**, 7886 (2015).
25. R. Dreyer, N. Liebing, E. R. J. Edwards, A. Müller, G. Woltersdorf, *Phys. Rev. Mater.* **5**, 064411 (2021).
26. J. Steiner, R. Schäfer, H. Wieczoreck, J. McCord, F. Otto, *Phys. Rev. B Condens. Matter Mater. Phys.* **85**, 104407 (2012).
27. M. Yan, A. Kákay, C. Andreas, R. Hertel, *Phys. Rev. B Condens. Matter Mater. Phys.* **88**, 220412 (2013).

28. T. Seki, K. Utsumiya, Y. Nozaki, H. Imamura, K. Takanashi, *Nat. Commun.* **4**, 1726 (2013).
29. S.-K. Kim, K.-S. Lee, D.-S. Han, *Appl. Phys. Lett.* **95**, 082507 (2009).
30. C. Koerner *et al.*, Frequency multiplication by collective nanoscale spin wave dynamics, Version 1, Zenodo (2022); https://doi.org/10.5281/zenodo.5873799.

VOLCANOLOGY

Magmatic water content controls the pre-eruptive depth of arc magmas

Daniel J. Rasmussen[1,2]*, Terry A. Plank[2], Diana C. Roman[3], Mindy M. Zimmer[4]

Vanguard efforts in forecasting volcanic eruptions are turning to physics-based models, which require quantitative estimates of magma conditions during pre-eruptive storage. Below active arc volcanoes, observed magma storage depths vary widely (~0 to 20 kilometers) and are commonly assumed to represent levels of neutral buoyancy. Here we show that geophysically observed magma depths (6 ± 3 kilometers) are greater than depths of neutral buoyancy, ruling out this commonly assumed control. Observed depths are instead consistent with predicted depths of water degassing. Intrinsically wetter magmas degas water and crystallize deeper than dry magmas, resulting in viscosity increases that lead to deeper stalling of ascending magma. The water–depth relationship provides a critical constraint for forecasting models by connecting depth of eruption initiation to its volatile fuel.

The way we view magmatic plumbing systems beneath active volcanoes has shifted from the notion of singular melt-rich pools (i.e., "magma chambers") toward models that describe complex, multitiered networks of crystal-rich mushes (*1, 2*). These systems may span the vertical depth of the crust (*3*) and may undergo reorganizations over time scales of volcanic unrest (*4*). In this paradigm, the degree of order in these complex systems is unclear.

Observations show that some sense of order exists in magmatic systems. Despite the resolution limitations, geophysical studies commonly find evidence for discrete regions of magma accumulation that are vertically restricted to a few kilometers (*5*). In contrast to the often-proposed dynamic nature of plumbing systems (*3, 4*), some geophysical observations are consistent with regions of storage that may persist for years at a particular depth throughout multiple phases of unrest and eruption (*6, 7*). These findings demonstrate our ability to identify favorable regions for magma storage, but we lack a general under-standing of the physical controls over depth. This variable is of central importance for historically active arc volcanoes. These systems supply Earth's daily volcanic eruptions and will be a focal point for eruption-forecasting models that use physics-based criteria to predict eruption onset, style, and duration (*8*).

Little is known about the primary controls over the depth of magma storage before eruption. A prevailing notion is that arc magmas are stored at their levels of neutral buoyancy, where the density of the bulk magma is equivalent to that of the host rock (*9*). Although this idea has been rigorously tested at mid-ocean ridges (*10*), the assumption remains largely untested at arcs. Alternatively, magma storage may be controlled by regional stress states or preexisting structures in the crust (*5, 11–13*). An associated idea is that magma storage occurs at crustal rheology boundaries (*14*), the locations of which depend not only on crustal properties but also on strain rate (*15*). Though crustal rheology is an important factor in the long-term survivability of reservoirs (*15*), we lack evidence that differences in crustal rheology alone are responsible for the observed variation in magma storage depths at arcs globally.

A different hypothesis—one that is not consistent with magma storage depth being controlled by extrinsic factors—highlights the potential importance of intrinsic factors such as magmatic water concentration (*15–17*).

During magma ascent and decompression, water solubility decreases until saturation is reached, at which point water degases progressively (*18*). Melt density, which relates to the buoyant force for magma ascent (*19*), may increase or decrease depending on whether exsolved volatiles remain entrained with the ascending magma (*20*). In all cases, a decrease in magmatic water content leads to an increase in melt viscosity (*20*), which can inhibit magma ascent and dike propagation (*21, 22*). Magmatic water content is also a strong control on liquidus temperature (*20*). Water degassing is thought to result in undercooling and crystallization (*23*), further increasing magma viscosity. The occurrence of such processes during magma ascent may result in a positive feedback loop in which degassing leads to increased magma viscosity and crystallinity, inhibiting magma ascent and causing magmas to become increasingly subjected to the effects of conductive cooling, thus leading to further melt crystallization and increases in viscosity. An analogous process has been suggested to occur at mid-ocean ridges, where magma ascent is halted as magma reaches a "freezing horizon" (*24, 25*). At arcs, the process has been described as "viscous death" (*16*). Viscous death may be the ultimate cause of reservoir formation in arcs, and some have found regional evidence for it (*17*). However, its global reach has not been tested.

To elucidate the role of water in the formation depth of magma reservoirs, we compare new and existing observations of magmatic water contents with corresponding geophysical observations of magma storage depth (*26*) (tables S1 to S3). Arcs are the best setting for our study because initial magmatic water contents are variable (*27*), and arc volcanoes are common targets for geophysical studies of magmatic plumbing systems (*5*). Our focus is on improving understanding of the observed reservoir depths rather than their long-term evolution. Therefore, our results are less relevant for long-lived silicic systems, which have been well studied elsewhere (*15, 28*).

Estimates of magmatic water content come from melt inclusion data, which include additional data for the central-eastern Aleutian arc and compiled data from arcs globally (table S2).

[1]National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. [2]Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA. [3]Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA. [4]Pacific Northwest National Laboratory, Richland, WA, USA.
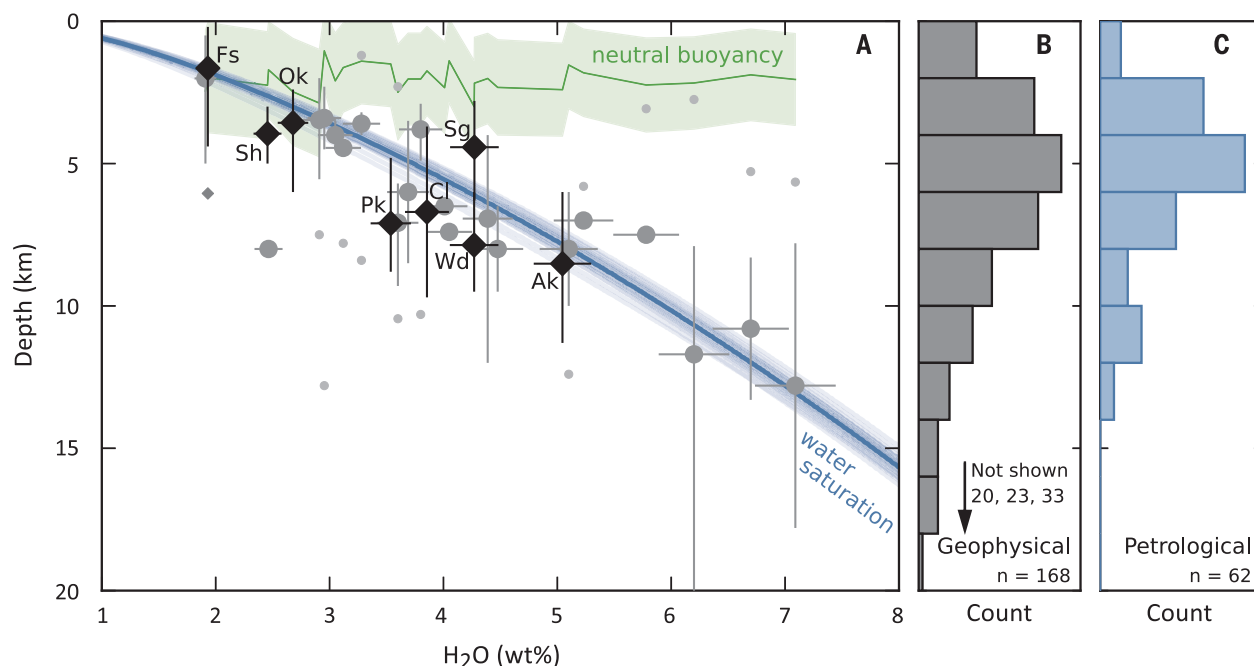*Corresponding author. Email: rasmussend@si.edu

Fig. 1. Link between magmatic water content and storage depth. (A) Magma storage occurs primarily where water degassing becomes important, not at the levels of neutral buoyancy. Data are for central-eastern Aleutian volcanoes (diamonds) and other locations (circles). Large symbols denote the reservoirs closest to the water saturation curve at each volcano, which fall on or near the curve in 27 of 28 cases (RMSE: 1.6 km). About half of the volcanoes have one or more other reservoirs (small markers) that are plotted off of the curve. Markers show the average centroid depth. Vertical error bars indicate the range of observed depths, and horizontal error bars denote assumed 5% measurement errors. Water degassing is approximated to start at the water saturation curve (blue lines), which was calculated with MagmaSat (*29*) and 27 density models. The same density models were used to calculate average (green line) and 1σ SD (green shaded area) levels of neutral buoyancy for open-system ascent. Closed-system models (not shown) indicate that magma is buoyant throughout the crust. Central-eastern Aleutian volcanoes are labeled (Ak, Akutan; Cl, Cleveland; Fs, Fisher; Pk, Pakushin; Ok, Okmok; Sg, Seguam; Sh, Shishaldin; Wd, Westdahl). **(B)** Histogram of 168 magma storage regions at 112 volcanoes estimated using geophysical methods. Three depth estimates (20, 23, and 33 km) are not shown. **(C)** Histogram of the water-saturation depth for 62 volcanoes. The similarity in the distributions in (B) and (C) supports the observation from (A) that storage depth relates to magmatic water content.

Our dataset includes 3856 naturally glassy melt inclusions from which we derive estimates of magmatic water content for 62 volcanoes. We restrict our dataset to mafic-intermediate melt compositions (<63 wt % SiO_2, <7 wt % total alkalis), which are less likely to have experienced water enrichment during melt evolution in the middle to upper crust. We assume that these melt compositions represent the melt feeding the reservoir. We use the maximum observed water concentration as the estimate of water content of the magma in the middle to upper crust before degassing. We take this approach to minimize the influence of processes that decrease water, such as degassing and post-entrapment diffusive loss of water (*26*).

We compiled magma storage depth estimates from geophysical studies, particularly those based on geodetic or seismic observations (table S3). Our compilation includes 331 depth estimates for 168 distinct magma reservoirs. The depths we consider are referenced to the land surface (not sea level, as commonly reported). The depth below the surface relates closely to the overburden pressure, which can then be linked to the depth of magma degassing through volatile solubility relationships. Many geochemical approaches

for determining magma storage depth are highly dependent on magmatic water content. Therefore, to avoid circular logic in comparing magmatic water content to storage depth, we restrict our estimates of magma storage depth to those determined using geophysical data.

Water contents of mafic-intermediate magmas from arc volcanoes vary from ~1 to 7 wt % (average: 4.0 ± 1.3 wt %) on the basis of our compilation, consistent with earlier work (*27*). Geophysical observations of magma storage are mostly constrained to the upper ~20 km of crust. A distinctive mode in observed storage depths occurs at 4 to 6 km below the surface with a long tail to greater depths (6.4 ± 2.8 km, 10%-trimmed average, 6.0-km median; Fig. 1B). These observations are likely biased to shallow depths because many of the geophysical techniques employed have better resolution at shallower depths. However, our results show substantial shallow (~3 to 9 km depth) storage of magmas at arcs.

We can investigate the control of magmatic water content on magma storage depth by comparing the geochemical and geophysical datasets (Fig. 1A). Water and depth estimates increase concurrently in 27 of the 28 volcanic systems (the exception being Semisopochnoi).

Not only do pre-eruptive storage regions form a strong water–depth trend, the trend also coincides with the water-saturation curve [1.6-km root mean square error (RMSE) for large symbols in Fig. 1A]. This provides strong evidence for a link between observed water content and depth. For half of the studied systems, this is the only known region of magma storage. For the other half, two or more discrete regions of magma storage have been identified (table S1). The other regions of magma storage (small symbols in Fig. 1A) are plotted off of the water-saturation curve. Therefore, although we observe magma storage along the water-saturation curve in nearly all of the volcanic systems we analyzed, this is not the only storage depth. Degassing of water will initiate at depths greater than that of water saturation if a vapor phase containing CO_2 is present, perhaps explaining the general tendency for storage to occur slightly below the water-saturation curve (Fig. 1A), but the effect of this is generally minor (*29*).

We can also compare all of the geophysically identified magma storage regions at arcs (*n* = 168) with the predicted depths at which arc magmas begin degassing water (water-saturation depth) at all arc volcanoes with water constraints (*n* = 62). Water-saturation

depths form a distinct mode at 4 to 6 km (5.5 ± 1.8 km, 10%-trimmed average, 4.9-km median; Fig. 1C), which coincides closely with geophysical estimates of magma depth (Fig. 1B). The water-saturation depth is determined by calculating the pressure at which melt would saturate with a pure-water vapor (29) and converting pressure to depth by using a summary density model based on data from 27 volcanoes (table S4). Additionally, our calculated water-saturation depths are influenced by our choice of solubility model, but different models yield similar results (26) except for one (30), which yields lower pressure (depth) estimates.

Although the empirical link between magmatic water content and storage depth is clear (Fig. 1), the interpretation requires further consideration. Two plausible explanations exist: (i) Magmatic water contents control magma stalling [mantle control (27)], or (ii) the storage depth of magmas dictates the water contents of the observed melt inclusions [crust control (27)]. Therefore, we need to know whether melt inclusions accurately record magmatic water contents. If the mantle control is dominant, then the water content of melt inclusions should correlate with other nonvolatile tracers of slab- and mantle-melting processes (e.g., Nb/Ce and Ba/La). If instead the water contents of melt inclusions are the result of a crustal control, melt should degas or melt inclusions should diffusively equilibrate to reflect their stalling depth. Any correlations that existed in the parental magma between water and nonvolatile tracers would be destroyed during these processes. Such correlations can be difficult to identify because they require a sufficiently large sampling of an arc segment where a relationship is expected. However, the central-eastern Aleutian volcanoes show systematic relationships between water and trace elements (Fig. 2) (27), and other melt inclusion studies have found similar correlations at other arcs, such as the Cascade (31) and Central American (32) arcs. Such systematics support a mantle control in which intrinsically wetter magmas, with distinct trace element compositions, degas and crystallize deeper than dry magmas, resulting in deeper storage before eruption. This trend exists globally despite substantial differences in magmatic flux and crustal properties between volcanoes, demonstrating the global importance of water content to the depth of reservoir formation. Whether a reservoir formed in this way would grow upward or downward with subsequent magmatic additions is unclear, which may cause some of the observed scatter (Fig. 1A). The longevity of such reservoirs is a separate question, which is best approached by means of thermomechanical modeling (15). Additionally, though water saturation can explain the presence of the observed storage reservoirs, their formation is not required. For example,

Fig. 2. Melt inclusion data for the central-eastern Aleutians. (**A** and **B**) The correlations indicate that degassing and diffusive water exchange have not substantially affected the water contents of the melt inclusions. Trace element data do not exist for most (five of eight) melt inclusions with maximum water content. For these melt inclusions, we use average trace element compositions of bulk rock samples with similar (i.e., within 0.25 wt %) K_2O concentrations. Bulk rock data at Cleveland do not extend to a sufficiently low K_2O concentration, so we extrapolated the bulk rock trend (table S8). Markers represent the mean ratios, and error bars denote SDs.
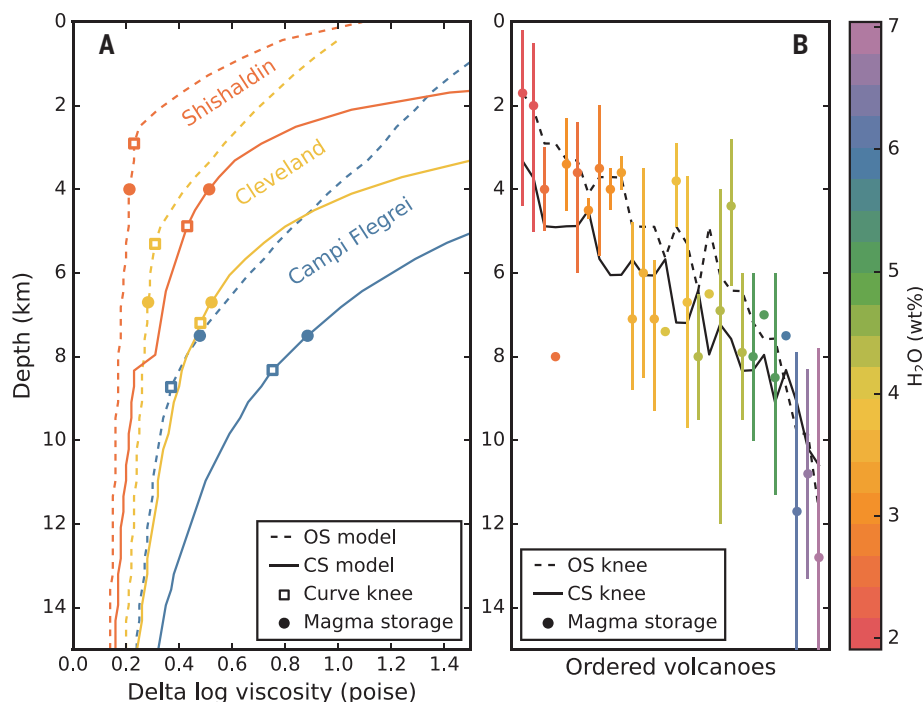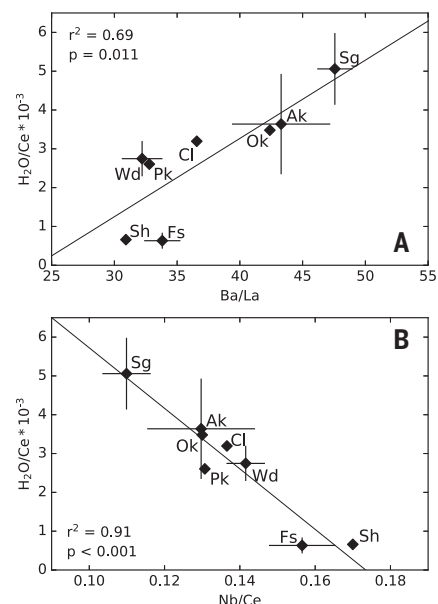
Fig. 3. Relationship between magma storage depth and magma viscosity. Pressure was converted to depth using our summary density model. (**A**) Change in magma viscosity during adiabatic magma ascent for three example volcanoes calculated using rhyolite-MELTS (29, 33) for cases of open-system (OS) and closed-system (CS) crystallization and volatile exsolution. For each volcanic system, the model (curve), the calculated viscosity knee (i.e., point of maximum convexity; square marker), and the average centroid depth of magma reservoirs determined geophysically (circle marker) are shown. Information on the MELTS modeling and calculation of the viscosity knee can be found in (26). (**B**) Comparison of the depth of the viscosity knee for OS and CS models (dashed and solid lines, respectively) with the geophysically observed depths of magma storage for volcanic systems in Fig. 1A. The volcanic systems are ordered along the x axis by the average depth of the OS and CS knees. In most cases, magma storage occurs at either the OS or CS knee, supporting viscosity change as a control on magma storage depth. Error bars indicate the range of observed depths.

cinder cones can erupt water-saturated magma with primitive compositions that are inconsistent with storage in the middle to upper crust.

Neutral buoyancy has long been invoked as a control over magma storage. We estimated the level of neutral buoyancy (LNB) for volcanoes in Fig. 1A by comparing our 27 compiled crustal-density models (table S4) with models of ascent, degassing, and crystallization conducted using rhyolite-MELTS (29, 33) (tables S5 to S7). For each volcanic system, we performed two models of adiabatic magma ascent: (i) an open-system model in which crystals and vapor were fractionated from the melt upon formation and (ii) a closed-system model in which crystals and vapor remain in equilibrium with the ascending magma. The results of our open-system models have consistent LNBs at ~2 km depth. Closed-system models universally show that magmas are buoyant throughout the crust, owing to the strong influence of a vapor phase on magma buoyancy (table S5). Although there are a small number of reservoirs at depths equivalent to the LNB assuming open-system melt evolution (Fig. 1A), most reservoirs occur at greater depth. Therefore, we reject neutral buoyancy as a primary control over magma storage depths. This result is consistent with findings at mid-ocean ridges (10); however, perhaps owing to differences in crustal density and stress state between arcs and ridges, neutral buoyancy has remained a fixture in arc-volcano literature for decades (15).

Degassing of water upon magma ascent causes increases in magma viscosity and liquidus temperatures (20), which can explain the observed depths of magma storage (Fig. 3 and table S6). We demonstrate this in the results of the MELTS models. We consider the change of the viscosity of the magma during ascent that we calculated with MELTS rather than the absolute viscosity. We make this our focus because our model assumes that magma ascent is occurring, which means that any barrier to initial ascent imposed by the absolute melt viscosity has been overcome. This must be true of the natural systems we studied because their magmas escaped the lower crust to form reservoirs in the middle to upper crust. Alternatively, the change of viscosity during magma ascent may influence dike propagation (21, 22). Our results show that magma viscosity increases during magma ascent (Fig. 3A). Open-system ascent typically leads to smaller increases in melt viscosity during ascent. The "knee" in the depth-viscosity curve (26) indicates the point of maximum convexity. Physically, it represents the depth at which viscosity increases during ascent become substantial. We find that most observed magma storage depths occur at either the open- or closed-system viscosity knees (Fig. 3B), supporting the idea that changes in magma viscosity during magma ascent control magma storage depth.

The magmatic systems we studied can generally be considered eruptible. Most of the systems in this study have been historically active. Geophysical data used to constrain magma depth commonly include coeruptive data (7), and we used erupted melt inclusions. Our results demonstrate that magmas are buoyant at their storage depth, providing a driving force for ascent and eruption. These systems have clearly experienced viscous stalling, not viscous death, consistent with noneruptive episodes of stalled intrusions that are commonly observed years to decades before many eruptions (34). Whether a magma body in the crust ultimately erupts, cools to form a pluton, or grows depends critically on crustal rheology and magma injection rate (15). Exsolved volatiles also play a role, as they would exist in a compressible vapor phase that dampens pressure increases due to injection of new magma, enabling reservoir growth (35). We investigated initial reservoir formation in the middle to upper crust and show that intrinsically wetter magmas degas water and crystallize deeper, leading to increases in magma viscosity and the formation of reservoirs at greater depths. These results will contribute to models of eruption triggering that depend critically on the conditions of the magma storage region (8).

REFERENCES AND NOTES

1. B. D. Marsh, *Mineral. Mag.* **60**, 5–40 (1996).
2. O. Bachmann, G. W. Bergantz, *J. Petrol.* **45**, 1565–1582 (2004).
3. K. V. Cashman, R. S. J. Sparks, J. D. Blundy, *Science* **355**, eaag3055 (2017).
4. R. S. J. Sparks, K. V. Cashman, *Elements* **13**, 35–40 (2017).
5. E. Chaussard, F. Amelung, *Geochem. Geophys. Geosyst.* **15**, 1407–1418 (2014).
6. D. Elsworth, G. Mattioli, J. Taron, B. Voight, R. Herd, *Science* **322**, 246–248 (2008).
7. Z. Lu, D. Dzurisin, in *InSAR Imaging of Aleutian Volcanoes* (Springer, 2014), pp. 87–345.
8. M. P. Poland, K. R. Anderson, *J. Geophys. Res. Solid Earth* **125**, e2018JB016974 (2020).
9. M. P. Ryan, in *Magmatmatic Processes: Physicochemical Principles*, B. O. Mysen, Ed. (Spec. Pub. no. 1, The Geochemical Society, 1987), pp. 259–287.
10. E. E. Hooft, R. S. Detrick, *Geophys. Res. Lett.* **20**, 423–426 (1993).
11. A. Gudmundsson, *J. Geodyn.* **43**, 6–29 (2007).
12. M. J. Stock, M. C. S. Humphreys, V. C. Smith, R. Isaia, D. M. Pyle, *Nat. Geosci.* **9**, 249–254 (2016).
13. A. H. Lerner *et al.*, *Geophys. Res. Lett.* **47**, e2020GL087856 (2020).
14. F. Mazzarini, G. Musumeci, D. Montanari, G. Corti, *Tectonophysics* **484**, 139–146 (2010).
15. C. Huber, M. Townsend, W. Degruyter, O. Bachmann, *Nat. Geosci.* **12**, 762–768 (2019).
16. C. Annen, J. D. Blundy, R. S. J. Sparks, *J. Petrol.* **47**, 505–539 (2006).
17. G. F. Zellmer *et al.*, *Am. Mineral.* **101**, 2405–2422 (2016).
18. G. Moore, *Rev. Mineral. Geochem.* **69**, 333–362 (2008).
19. R. F. Weinberg, Y. Podladchikov, *J. Geophys. Res.* **99**, 9543–9559 (1994).
20. C. E. Lesher, F. J. Spera, in *The Encyclopedia of Volcanoes* (Elsevier, 2015), pp. 113–141.
21. D. A. Spence, D. L. Turcotte, *J. Geophys. Res.* **90**, 575–580 (1985).
22. J. R. Lister, R. C. Kerr, *J. Geophys. Res.* **96**, 10049–10077 (1991).
23. J. Blundy, K. Cashman, *Contrib. Mineral. Petrol.* **140**, 631–650 (2001).
24. D. W. Sparks, E. M. Parmentier, *Earth Planet. Sci. Lett.* **105**, 368–377 (1991).
25. J. P. Morgan, Y. J. Chen, *J. Geophys. Res.* **98**, 6283–6297 (1993).
26. Materials and methods are available as supplementary materials.
27. T. Plank, K. A. Kelley, M. M. Zimmer, E. H. Hauri, P. J. Wallace, *Earth Planet. Sci. Lett.* **364**, 168–179 (2013).
28. L. Karlstrom, S. R. Paterson, A. M. Jellinek, *Nat. Geosci.* **10**, 604–608 (2017).
29. M. S. Ghiorso, G. A. R. Gualda, *Contrib. Mineral. Petrol.* **169**, 53 (2015).
30. P. Papale, R. Moretti, D. Barbato, *Chem. Geol.* **229**, 78–95 (2006).
31. K. J. Walowski, P. J. Wallace, M. A. Clynne, D. J. Rasmussen, D. Weis, *Earth Planet. Sci. Lett.* **446**, 100–112 (2016).
32. S. J. Sadofsky, M. Portnyagin, K. Hoernle, P. van den Bogaard, *Contrib. Mineral. Petrol.* **155**, 433–456 (2008).
33. G. A. R. Gualda, M. S. Ghiorso, R. V. Lemons, T. L. Carley, *J. Petrol.* **53**, 875–890 (2012).
34. S. C. Moran, C. Newhall, D. C. Roman, *Bull. Volcanol.* **73**, 115–122 (2011).
35. W. Degruyter, C. Huber, O. Bachmann, K. M. Cooper, A. J. R. Kent, *Geochem. Geophys. Geosyst.* **18**, 4123–4135 (2017).
36. D. J. Rasmussen, T. A. Plank, Melt inclusion data for the central-eastern Aleutian volcanoes, version 1.0, EarthChem, Interdisciplinary Earth Data Alliance (2021); https://doi.org/10.26022/IEDA/111873.
37. D. J. Rasmussen, T. A. Plank, Melt inclusion olivine-host data for the central-eastern Aleutian volcanoes, version 1.0, EarthChem, Interdisciplinary Earth Data Alliance (2021); https://doi.org/10.26022/IEDA/111871.
38. D. Rasmussen, T. Plank, Melt inclusion host data for the Cleveland volcano, version 1.0, EarthChem, Interdisciplinary Earth Data Alliance (2021); https://doi.org/10.26022/IEDA/111872.
39. D. Rasmussen, T. Plank, Bulk rock data for the central-eastern Aleutian volcanoes, version 1.0, EarthChem, Interdisciplinary Earth Data Alliance (2021); https://doi.org/10.26022/IEDA/111870.

CHROMOSOMES

Structural basis of human telomerase recruitment by TPP1-POT1

Zala Sekne†, George E. Ghanim*†, Anne-Marie M. van Roon, Thi Hoang Duong Nguyen*

Telomerase maintains genome stability by extending the 3' telomeric repeats at eukaryotic chromosome ends, thereby counterbalancing progressive loss caused by incomplete genome replication. In mammals, telomerase recruitment to telomeres is mediated by TPP1, which assembles as a heterodimer with POT1. We report structures of DNA-bound telomerase in complex with TPP1 and with TPP1-POT1 at 3.2- and 3.9-angstrom resolution, respectively. Our structures define interactions between telomerase and TPP1-POT1 that are crucial for telomerase recruitment to telomeres. The presence of TPP1-POT1 stabilizes the DNA, revealing an unexpected path by which DNA exits the telomerase active site and a DNA anchor site on telomerase that is important for telomerase processivity. Our findings rationalize extensive prior genetic and biochemical findings and provide a framework for future mechanistic work on telomerase regulation.

Telomerase restores telomeric repeats [(TTAGGG)$_n$ in humans] by de novo DNA synthesis using a telomerase reverse transcriptase (TERT) subunit and an RNA template embedded within telomerase RNA (hTR in humans) (1). Telomerase activity is critical for the long-term proliferation of most cancers and germline and stem cells, whereas telomerase deficiency results in premature aging diseases (2). Human telomerase consists of two flexibly tethered functional lobes (3, 4): a catalytic core, in which TERT associates with the pseudoknot/template (PK/t) domain and conserved regions 4 and 5 (CR4/5) of hTR (5), and an H/ACA ribonucleoprotein (RNP) lobe, which is essential for telomerase biogenesis (6) (Fig. 1, A to C). The catalytic core also associates with a histone H2A-H2B dimer (4) (Fig. 1, A, D, and E). Telomerase access to and activity at telomeres is tightly regulated (7). Mam-

malian telomeric DNA is bound and protected by a six-membered protein complex called shelterin (8). The shelterin component TPP1 has been implicated in telomerase recruitment to telomeres (9–13). Within shelterin, TPP1 binds TIN2 and POT1, a single-stranded telomeric DNA binding protein (14, 15). TPP1-POT1 and TPP1-POT1-TIN2 complexes have been demonstrated to stimulate telomerase repeat addition processivity (RAP), which is the ability of telomerase to add multiple telomeric repeats with each DNA binding event (16, 17). Both RAP and interactions with shelterin are essential for telomerase function in vivo (9–11, 18). Yet the structural basis of telomerase-shelterin interactions and shelterin-mediated telomerase processivity remains elusive.

We prepared the human TPP1-POT1-TIN2 (TPT) complex (fig. S1A), demonstrated that it stimulated telomerase processivity in vitro (fig. S1, B and C), reconstituted its complex with human telomerase and the telomeric DNA substrate (T$_2$AG$_3$)$_5$ (fig. S1, D and E), and

Fig. 1. Structures of the telomerase catalytic core with telomeric DNA and shelterin components. (**A**) Domain organization of protein subunits in the catalytic core and TPT complex. Regions not observed in the structures are displayed as semitransparent. The same domain colors are used throughout, unless indicated otherwise. (**B**) Secondary structure of hTR. (**C**) Schematic of human telomerase bound to TPP1 and POT1. (**D**) A 3.2-Å cryo-EM reconstruction of the telomerase catalytic core–DNA–TPP1 complex (fig. S6). (**E**) A 3.9-Å cryo-EM reconstruction of the telomerase catalytic core–DNA–TPP1–POT1 complex (fig. S7).
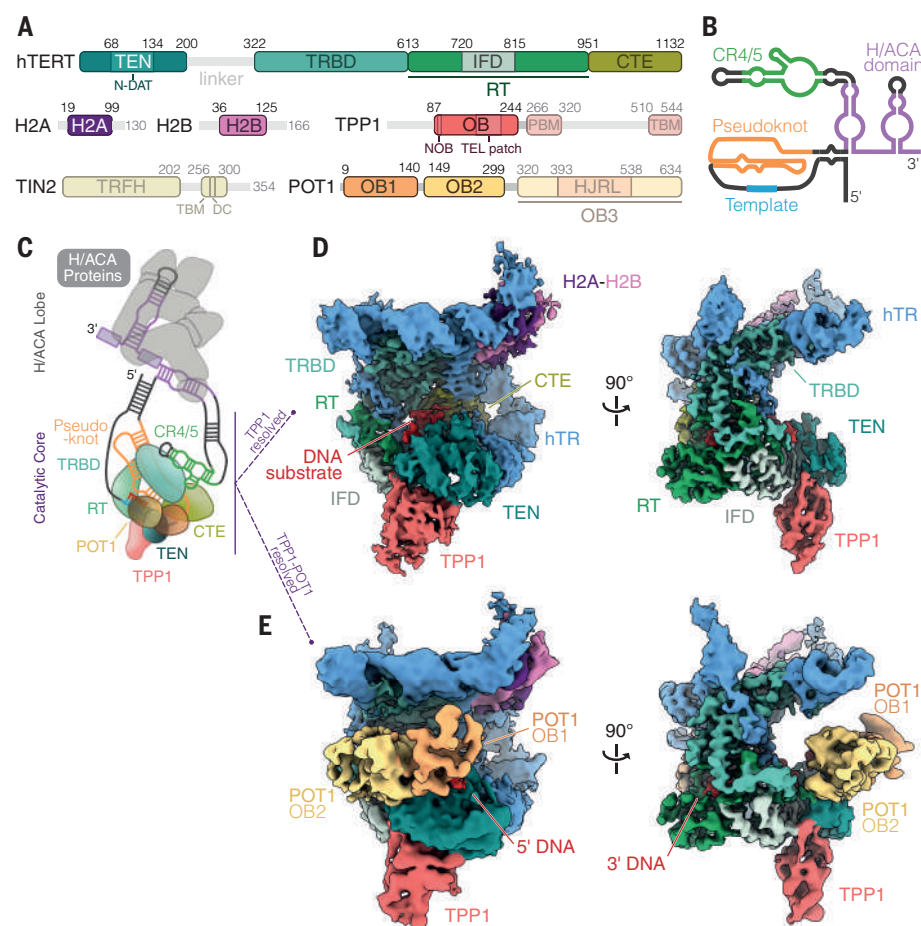
verified the retention of the DNA within this complex (fig. S1F). Negative-stain electron microscopy (EM) analyses showed that TPT binds the telomerase catalytic core (fig. S2, A to D). Cryo-EM analyses showed conformational flexibility between the catalytic core, H/ACA lobe, and TPT (figs. S2, E to G; S3; and S4, A to E). Focused classification and refinement resolved the DNA-bound telomerase catalytic core with either TPP1 or TPP1-POT1 at overall resolutions of 3.2 and 3.9 Å, respectively (Fig. 1, D and E; figs. S3 and S5 to S7; and tables S1 and S2). TIN2 could not be resolved by image processing.

In both structures, the TPP1 oligonucleotide/oligosaccharide-binding (OB)–fold domain binds the bottom face of the telomerase catalytic core (Fig. 1, D and E). TERT has four domains: the telomerase essential N-terminal (TEN) domain, the telomerase RNA-binding domain (TRBD), the reverse transcriptase (RT) domain, and the C-terminal extension (CTE) domain (Fig. 1A). TPP1 interposes between the TEN domain and a telomerase-specific insertion in fingers subdomain (IFD) within the RT domain (Fig. 1D, right). In previous structures (3, 4), the TEN domain showed the most conformational variability. TPP1 binding induces conformational changes in the TEN domain and reduces its flexibility, thereby improving its local resolution (figs. S5G and S8A). We also observe a slight compaction of the CTE domain and the P2 stem of hTR (fig. S8, B and D).

The 3.2-Å telomerase-DNA-TPP1 structure reveals the molecular basis of TPP1-telomerase interactions (Fig. 2 and fig. S6, I to K). Previous studies identified an N-terminal region of TPP1 OB-fold domain (NOB) and a TPP1 glutamate and leucine-rich (TEL) patch that are crucial for stimulating telomerase processivity in vitro and telomerase recruitment to telomeres in vivo (fig. S9B) (9–11, 13). The NOB and TEL patch form three areas of contact with TERT (Fig. 2A). The NOB seats into a hydrophobic cleft formed by the TEN domain and the IFD, in contrast to its extended conformation in the crystal structure of the isolated TPP1 OB-fold domain (16) (Fig. 2, B and C, and fig. S6I). The glutamate-rich region of the TEL patch (residues 168 to 172) folds into a short α helix that is accommodated by a basic surface on the TEN domain (Fig. 2D and fig. S6J). The other region of the TEL patch [residues 210 to 215, Arg[180] (R180), and Leu[183] (L183)] interacts extensively with the TEN domain and IFD (Fig. 2E and fig. S6K). Superimposition of the unbound versus telomerase-bound TPP1 OB-fold structures shows that the TEL-patch loop (residues 212 to 215) has some flexibility to enable interaction with TERT (Fig. 2B). Mutations in the NOB or the TEL patch severely disrupt TPP1-telomerase interactions and recruitment to telomeres (9–11). Thus, the interactions observed here are functionally important for telomerase recruitment by TPP1.

Fig. 2. Telomerase-TPP1 interactions. (**A**) Interactions between the TPP1 OB-fold domain and TERT. The dashed circles indicate the three contacts that the NOB (orange) and TEL patch (red) of TPP1 make with TERT. The β hairpin, which becomes ordered upon TPP1 binding, is highlighted. (**B**) Comparison of the unbound [gray, PDB ID 2I46 (16)] and telomerase-bound (colored) OB-fold domain of TPP1. (**C**) Close-up view of the interactions between the NOB of TPP1 and TERT (fig. S6I). (**D** and **E**) Close-up views of the interactions between the TEL patch and TERT (fig. S6, J and K). G, Gly; N, Asn; S, Ser; T, Thr. For (C) to (E), the orange dashed lines indicate hydrogen bond and salt bridge interactions.
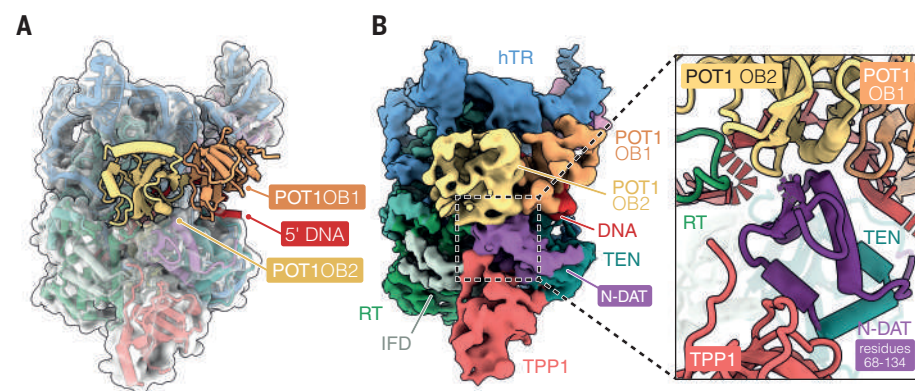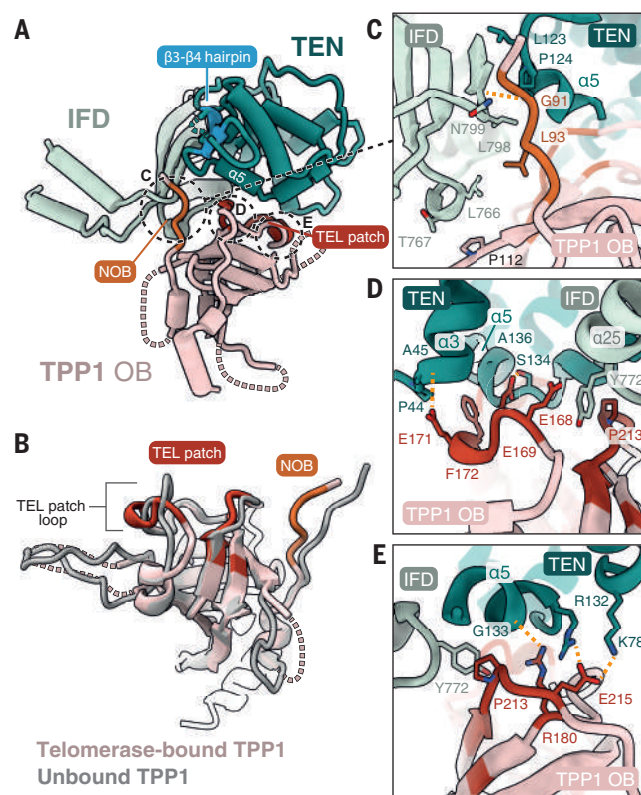

Fig. 3. Telomerase-POT1 interactions. (**A**) Structure of the telomerase-DNA-TPP1-POT1 complex. (**B**) A 3.9-Å cryo-EM reconstruction of the telomerase-DNA-TPP1-POT1 complex. The inset highlights interactions between POT1 and TERT. The N-DAT region (purple) interacts with both TPP1 and POT1.

Our structure reveals the key TERT regions necessary for TPP1 interaction. Helix α5 (residues 122 to 135) extensively interacts with TPP1 NOB and the TEL patch (Fig. 2, C to E, and fig. S6, I to K). Upon binding to TPP1, the β3-β4 hairpin and loop 60-76 of TERT become ordered (Fig. 2A and fig. S8A). Both helix α5 and the β3-β4 hairpin are located in the N-terminal dissociates-of-activities domain of telomerase (N-DAT) that is embedded in the TEN domain (19). Mutations in the N-DAT were shown to have small effects on telomerase catalytic activity but impaired telomere elongation in vivo, suggesting defects in telomerase recruitment to telomeres (19, 20). Additional contacts were observed between TEN domain Lys[78] (K78), helix α3 (residues 44 to 46), and the TPP1 TEL patch (Fig. 2, D and E, and fig. S6J). Our findings explain previous studies that showed that charge-reversal mutations at R132 and K78 of the TEN domain substantially reduced the RAP stimulation by TPP1-POT1 and disrupted telomerase localization to telomeres in vivo. These defects were rescued by a charge-swap Glu[215]→Lys (E215K) mutation in TPP1 (12, 21). The IFD has also been suggested to be involved in TPP1 binding and telomerase recruitment to telomeres (22).
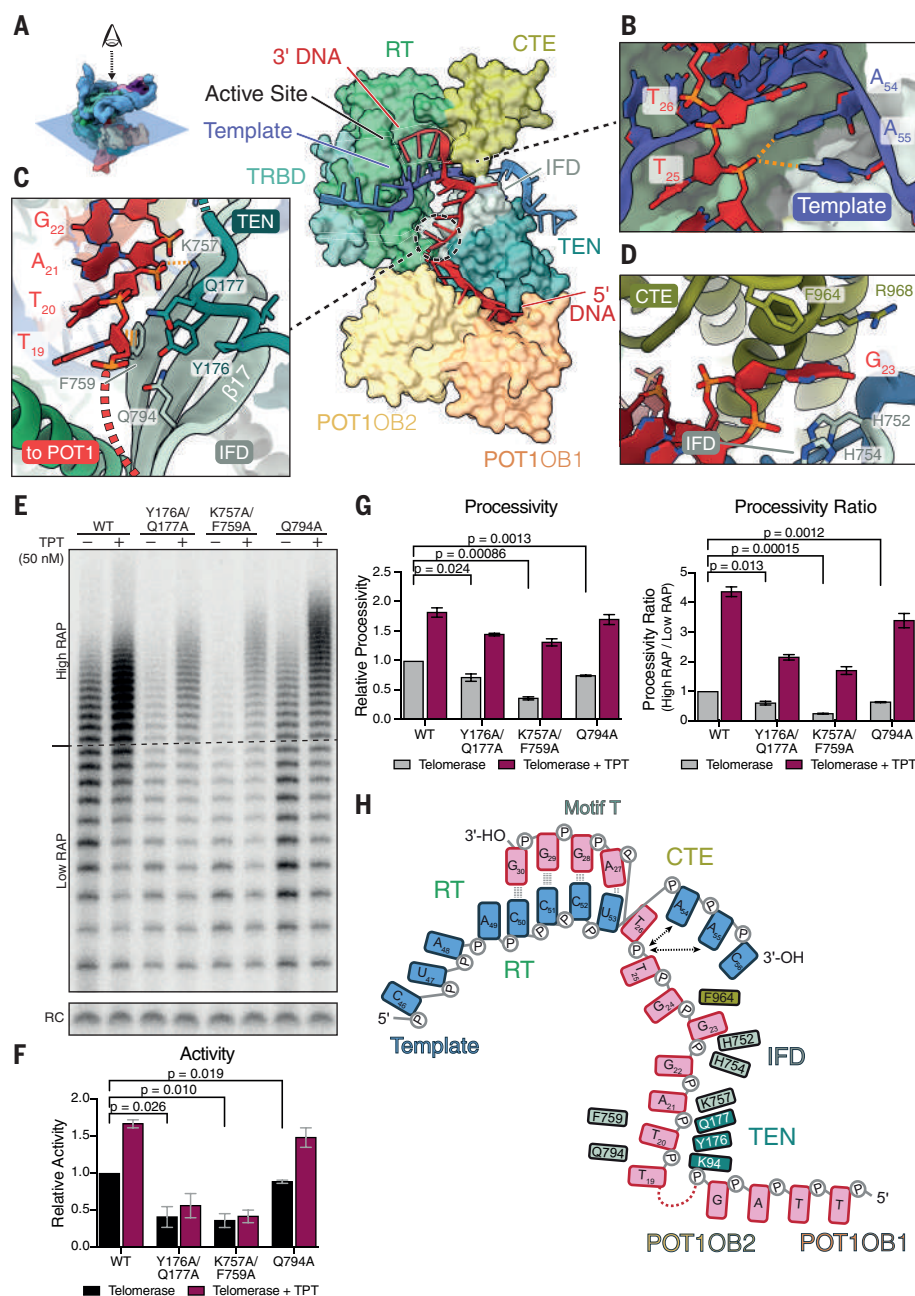
Fig. 4. Telomeric DNA substrate. (**A**) The DNA path guided by TERT, hTR, and POT1. The graphic at the top left shows how the view is related to the telomerase-DNA-TPP1-POT1 reconstruction. (**B**) Interactions between the template RNA and the DNA upstream of the RNA template–DNA duplex (fig. S6L). For (B) and (C), the orange dashed lines indicate hydrogen bonding and stacking interactions. (**C**) Close-up view of the interactions between the DNA substrate and the proposed anchor site on the TEN domain and the IFD (fig. S6N). The red dashed line indicates the unmodeled DNA connection. (**D**) Interactions between the flipped-out G23 DNA base and TERT (fig. S6M). H, His. (**E**) Telomerase activity assays in the absence and presence of the TPT complex for the wild-type (WT) enzyme and TERT mutants (Y176A/Q177A, K757A/F759A, and Q794A) at the proposed DNA anchor site (fig. S6, H and N). Assays were performed after enzyme enrichment on MagStrepXT resins using ZZ-TEV-twin-Strep–tagged TERT. Experiments were performed in triplicate. Detailed analyses of the input lysates are shown in fig. S12. RC, recovery control. (**F** and **G**) Bar graphs showing the quantifications of telomerase activity and processivity of the activity assays shown in (E), respectively. Values were normalized to telomerase without TPT. Error bars are the standard errors of the mean obtained from the replicates. (**H**) Schematic of the DNA substrate and its interactions with TERT, hTR, and POT1. Only interactions with the 5′ part of the DNA following the DNA-RNA duplex are highlighted because interactions of the DNA-RNA duplex with TERT have been discussed previously (4). P, phosphate.

In our structure, the base of an extended β sheet formed by the IFD (residues 766, 767, and 797 to 799) and the TEN domain hold the NOB in place (Fig. 2C and fig. S6I). Additionally, part of the IFD helix α25 (residues 771 to 775) is sandwiched between the two contact sites in the TEL patch (Fig. 2D).

POT1 is more flexibly engaged with telomerase than TPP1, which is evident from its local resolution range of 7 to 9 Å (figs. S4, C and D, and S5H). The 3.9-Å telomerase-DNA-TPP1-POT1 map allowed us to rigid-body fit a DNA-bound POT1 crystal structure (POT1 OB1 and OB2) (Figs. 1, A and E, and 3; and figs. S7F and S9C) (23). OB3 and the Holliday-junction resolvase-like (HJRL) domains of POT1 and the C-terminal region of TPP1, which interacts with TIN2 and POT1, are unresolved in our structures (Fig. 1A) (24, 25). With high affinity for single-stranded DNA (23), POT1 was thought to associate with telomerase via TPP1 and enhance telomerase RAP by increasing avidity to telomeric DNA (16, 26, 27). Surprisingly, POT1 binds not only the DNA substrate but also TERT and forms a gate in front of the telomerase active site (Fig. 3). The N-DAT within the TEN domain wedges between the two OB-fold domains of POT1 (Fig. 3B). Substitutions of POT1-contacting residues in the N-DAT severely affect telomere elongation in vivo with minimal effects on telomerase catalytic activity in vitro (19). Thus, the identified interactions between POT1 and telomerase in our structure are crucial for telomerase recruitment to telomeres in vivo.

We previously used a $T_{12}(T_2AG_3)$ DNA substrate and observed only the 3′ terminal TTAGGG repeat (4). Here, we used a longer $(T_2AG_3)_5$ DNA substrate to allow for TPP1-POT1 binding. Unlike its ciliate TEBPβ homolog (28), TPP1 does not contact the DNA (fig. S10). Yet its association with telomerase stabilizes the DNA, allowing us to resolve another TTAGGG repeat in our 3.2-Å telomerase-DNA-TPP1 map (fig. S6G). The telomerase-DNA-TPP1-POT1 map shows additional DNA density, forming a continuous path between the 12-nucleotide modeled DNA and the DNA bound to POT1 from the previous crystal structure [Protein Data Bank (PDB) ID 1XJV] (23) (fig. S7K). Although the resolution of this region is insufficient for de novo modeling, we use this density and the two DNA models to propose the DNA path through telomerase. The DNA 5′ end threads through a positively charged tunnel formed by the POT1 OB1 domain and the TEN domain and is guided along the TEN domain by the POT1 OB2 domain (Fig. 4, A and H, and fig. S11, A and B). The DNA then makes a sharp turn and traverses the IFD and CTE domains before reaching the telomerase active site (Fig. 4A and figs. S6G and S7G).

The telomeric DNA path reveals a previously undiscovered DNA threading surface on

the TEN domain and the IFD of TERT, with implications for telomerase processivity. The TEN domain is essential for RAP and telomerase recruitment to telomeres (12, 29, 30). Several studies have suggested that the TEN domain contains an anchor site that binds the single-stranded DNA substrate and prevents DNA dissociation, thereby promoting RAP (20, 30–34). However, the molecular determinants for the DNA binding activity of the TEN domain and the location of the anchor site remained unclear. We identified a Pro-Leu-Tyr-Gln (PLYQ) motif in the TEN domain (residues 174 to 177), which binds and turns the DNA substrate toward the TEN-POT1 interface (Fig. 4C and figs. S6N, S7I, and S11, A and B). This motif is highly conserved in vertebrates (fig. S9A). Binding of the TEN domain to the DNA is assisted by the IFD (Q794 and β17, residues 752 to 759) (Fig. 4C).

To test the role of the PLYQ motif and the IFD as contributors to the DNA anchor site, we reconstituted telomerase with TERT Y176A/Q177A, K757A/F759A, and Q794A mutations (where A is Ala and F is Phe), which significantly reduced telomerase activity and RAP (Fig. 4, E to G, and fig. S12). Addition of the TPT complex rescued the RAP defects in these mutants (Fig. 4, E to G). Similar observations were made previously with Y176A and Q177A single mutations (10, 26, 30). Substitutions at TERT residues 170 to 175 considerably increase the Michaelis constant (K_m) for the $(T_2AG_3)_3$ primer and affect telomere elongation in vivo (19, 30). This suggests that the RAP defect observed in the mutants is caused by DNA binding defects during repeat synthesis, and the TPT complex rescued the RAP defects by compensating for the reduced DNA affinity. Thus, we propose that the PLYQ motif of the TEN domain and strand β17 of the IFD form the anchor site that is crucial for DNA retention during RAP.

hTR also contributes to shaping the DNA path. Immediately downstream of the 4–base pair DNA-RNA template duplex, the RNA template base A54 stabilizes the flipped-out DNA base T26 by base-stacking and coordinates the phosphate backbone of the neighboring T25 together with the RNA template base A55 (Fig. 4, A, B, and H, and fig. S6L). Previous studies showed that the RNA template mutants A54U and A55U compromised telomerase RAP (18), suggesting that these unexpected DNA-RNA template interactions may be important for RAP. We also observe another flipped-out DNA base (G23), which is stabilized by the IFD and CTE domain (Fig. 4, D and H, and fig. S6M).

The *Tetrahymena* p50-TEB complex is functionally comparable to human TPP1-POT1. Whereas TPP1-POT1 is part of the shelterin complex that resides at mammalian telomeres, the p50-TEB complex is part of the *Tetrahymena*

telomerase holoenzyme (35, 36). In *Tetrahymena* telomerase (36), p50 occupies a similar position to TPP1 relative to the TEN domain (fig. S11E). However, POT1 arrangement is notably different from that of TEB (fig. S11E). The DNA substrate in *Tetrahymena* does not travel along the TEN domain as observed in our structure (fig. S11E). The positive charge of the DNA threading surface on the human TEN domain is not conserved in *Tetrahymena* and *Saccharomyces cerevisiae* (fig. S11, C and D). Therefore, the DNA path along the human TEN domain may be vertebrate specific.

Finally, our structures explain how TPP1 and POT1 facilitate telomerase association with telomeric DNA and stimulate telomerase RAP. With each round of telomeric repeat synthesis, the product DNA can either realign with the template for another round of repeat synthesis or dissociate. In the telomerase structure alone (4), the TEN domain is conformationally flexible, resulting in unstable DNA binding and increased DNA dissociation during repeat synthesis and thus lower RAP (fig. S13A). TPP1 binding stabilizes the TEN domain and allows the anchor site to engage the DNA substrate more stably. POT1 further stabilizes DNA binding by promoting the DNA-TEN domain interaction. Together, TPP1 and POT1 cooperatively increase RAP by reducing DNA dissociation during repeat synthesis, which agrees with previous kinetic studies (37) (fig. S13B). The observed flexibility of POT1 may be necessary for efficient translocation during RAP and/or for accommodating G-quadruplex structures formed by the telomeric DNA product (27) (fig. S4F). Recruitment by shelterin (via TPP1-POT1) would ensure high processivity of telomere-engaged telomerase. Telomerase recruited to a telomere without shelterin would dissociate more easily, thus impairing processive DNA synthesis.

REFERENCES AND NOTES

1. E. H. Blackburn, K. Collins, *Cold Spring Harb. Perspect. Biol.* **3**, a003558 (2011).
2. B. Bernardes de Jesus, M. A. Blasco, *Trends Genet.* **29**, 513–520 (2013).
3. T. H. D. Nguyen *et al.*, *Nature* **557**, 190–195 (2018).
4. G. E. Ghanim *et al.*, *Nature* **593**, 449–453 (2021).
5. T. L. Beattie, W. Zhou, M. O. Robinson, L. Harrington, *Curr. Biol.* **8**, 177–180 (1998).
6. E. D. Egan, K. Collins, *Mol. Cell. Biol.* **32**, 2428–2439 (2012).
7. J. C. Schmidt, T. R. Cech, *Genes Dev.* **29**, 1095–1105 (2015).
8. T. de Lange, *Annu. Rev. Genet.* **52**, 223–247 (2018).
9. J. Nandakumar *et al.*, *Nature* **492**, 285–289 (2012).
10. A. N. Sexton, D. T. Youmans, K. Collins, *J. Biol. Chem.* **287**, 34455–34464 (2012).
11. F. L. Zhong *et al.*, *Cell* **150**, 481–494 (2012).
12. J. C. Schmidt, A. B. Dalby, T. R. Cech, *eLife* **3**, e03563 (2014).
13. S. Grill, V. M. Tesmer, J. Nandakumar, *Cell Rep.* **22**, 1132–1140 (2018).
14. P. Baumann, T. R. Cech, *Science* **292**, 1171–1175 (2001).
15. J. Z. Ye *et al.*, *Genes Dev.* **18**, 1649–1654 (2004).
16. F. Wang *et al.*, *Nature* **445**, 506–510 (2007).
17. A. M. Pike, M. A. Strong, J. P. T. Ouyang, C. W. Greider, *Mol. Cell. Biol.* **39**, e00593–e00518 (2019).
18. R. A. Wu, J. Tam, K. Collins, *EMBO J.* **36**, 1908–1927 (2017).
19. B. N. Armbruster, S. S. R. Banik, C. Guo, A. C. Smith, C. M. Counter, *Mol. Cell. Biol.* **21**, 7775–7786 (2001).
20. D. C. Sealey *et al.*, *Nucleic Acids Res.* **38**, 2019–2035 (2010).
21. J. L. Stern, K. G. Zyner, H. A. Pickett, S. B. Cohen, T. M. Bryan, *Mol. Cell. Biol.* **32**, 2384–2395 (2012).
22. T. W. Chu, Y. D'Souza, C. Autexier, *Mol. Cell. Biol.* **36**, 210–222 (2015).
23. M. Lei, E. R. Podell, T. R. Cech, *Nat. Struct. Mol. Biol.* **11**, 1223–1229 (2004).
24. C. Chen *et al.*, *Nat. Commun.* **8**, 14929 (2017).
25. C. Rice *et al.*, *Nat. Commun.* **8**, 14928 (2017).
26. A. J. Zaug, E. R. Podell, J. Nandakumar, T. R. Cech, *Genes Dev.* **24**, 613–622 (2010).
27. L. I. Jansson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9350–9359 (2019).
28. M. P. Horvath, V. L. Schweiker, J. M. Bevilacqua, J. A. Ruggles, S. C. Schultz, *Cell* **95**, 963–974 (1998).
29. A. R. Robart, K. Collins, *Mol. Cell* **42**, 308–318 (2011).
30. J. Jurczyluk *et al.*, *Nucleic Acids Res.* **39**, 1774–1788 (2011).
31. H. D. Wyatt, D. A. Lobb, T. L. Beattie, *Mol. Cell. Biol.* **27**, 3226–3240 (2007).
32. N. F. Lue, *J. Biol. Chem.* **280**, 26586–26591 (2005).
33. S. A. Jacobs, E. R. Podell, T. R. Cech, *Nat. Struct. Mol. Biol.* **13**, 218–225 (2006).
34. B. M. Akiyama, J. W. Parks, M. D. Stone, *Nucleic Acids Res.* **43**, 5537–5549 (2015).
35. H. E. Upton, K. Hong, K. Collins, *Mol. Cell. Biol.* **34**, 4200–4212 (2014).
36. Y. He *et al.*, *Nature* **593**, 454–459 (2021).
37. C. M. Latrick, T. R. Cech, *EMBO J.* **29**, 924–933 (2010).

NEUROSCIENCE

Angiotensin-converting enzyme gates brain circuit–specific plasticity via an endogenous opioid

Brian H. Trieu[1,2], Bailey C. Remmers[3], Carlee Toddes[1], Dieter D. Brandner[1,2], Emilia M. Lefevre[3], Adrina Kocharian[1,2], Cassandra L. Retzlaff[3], Rachel M. Dick[1], Mohammed A. Mashal[3], Elysia A. Gauthier[3], Wei Xie[4], Ying Zhang[5], Swati S. More[4], Patrick E. Rothwell[3]*

Angiotensin-converting enzyme (ACE) regulates blood pressure by cleaving angiotensin I to produce angiotensin II. In the brain, ACE is especially abundant in striatal tissue, but the function of ACE in striatal circuits remains poorly understood. We found that ACE degrades an unconventional enkephalin heptapeptide, Met-enkephalin-Arg-Phe, in the nucleus accumbens of mice. ACE inhibition enhanced µ-opioid receptor activation by Met-enkephalin-Arg-Phe, causing a cell type–specific long-term depression of glutamate release onto medium spiny projection neurons expressing the Drd1 dopamine receptor. Systemic ACE inhibition was not intrinsically rewarding, but it led to a decrease in conditioned place preference caused by fentanyl administration and an enhancement of reciprocal social interaction. Our results raise the enticing prospect that central ACE inhibition can boost endogenous opioid signaling for clinical benefit while mitigating the risk of addiction.

As neural circuit dysfunction in brain disorders becomes increasingly well defined, there is a growing need for interventions that specifically target dysfunctional circuit elements (1). Multiple brain disorders (2–4) involve imbalanced output of nucleus accumbens (NAc) medium spiny projection neurons expressing dopamine receptor Drd1 (D1-MSNs) or Drd2 (D2-MSNs). This imbalance has proven difficult to correct with standard interventions because these two MSN subtypes are physically intermingled, receive synaptic inputs from common sources, and have similar molecular profiles. A rare exception is ACE (angiotensin-converting enzyme), which exhibits enriched expression by D1-MSNs in the dorsal striatum (5, 6) and the NAc (table S1 and fig. S1). Inhibitors of ACE and other peptidases can be combined to regulate striatal excitatory synaptic transmission in an opioid-dependent fashion (7); this finding suggests that in addition to angiotensin conversion, ACE cleaves and degrades a peptide ligand for opioid receptors (Fig. 1A).

To separately measure how ACE inhibition affects excitatory synaptic transmission onto D1-MSNs and D2-MSNs, we performed whole-cell recordings in acute NAc brain slices from double-transgenic Drd1-tdTomato/Drd2-eGFP reporter mice (Fig. 1C and fig. S2). Brief exposure to 10 µM captopril, a prototypical ACE inhibitor (8), caused long-term depression (captopril-LTD) of excitatory synaptic transmission onto D1-MSNs (Fig. 1D). Captopril did not alter excitatory synaptic transmission onto D2-MSNs, which express ACE at a lower level than D1-MSNs (table S1 and fig. S1). There was also no effect of captopril at excitatory synapses onto layer V pyramidal neurons in the anterior cingulate cortex (fig. S3), where ACE expression is low (5, 6).

Captopril and other ACE inhibitors canonically block conversion of angiotensin I to angiotensin II, preventing activation of the angiotensin II type 1 receptor (AT1R) and increasing levels of angiotensin I (Fig. 1B). However, LTD was not observed in D1-MSNs exposed to valsartan (2 to 20 µM), an AT1R antagonist, or exogenous angiotensin I peptide (1 µM) (Fig. 1E). In contrast, captopril-LTD in D1-MSNs was blocked in the continuous presence of 10 µM naloxone, an opioid receptor antagonist, but was not reversed by chasing captopril with naloxone (Fig. 1F). Captopril-LTD in D1-MSNs was associated with an increase in paired-pulse ratio and a decrease in $1/CV^2$ (Fig. 1, G to I), two changes that indicate a decreased presynaptic probability of glutamate release, likely due to activation of presynaptic opioid receptors (7).

Local release of enkephalin peptides by D2-MSNs can regulate excitatory synaptic input to D1-MSNs (9). ACE can cleave enkephalin peptides but is not principally responsible for degrading conventional Met-enkephalin or Leu-enkephalin in brain tissue (10). The pro-enkephalin gene (Penk) also encodes Met-enkephalin-Arg-Phe (MERF), a heptapeptide abundant in the NAc (11). MERF has high binding affinity for opioid receptors (12), is more potent than Met-enkephalin (13), and can be degraded by ACE (14). Using liquid chromatography–tandem mass spectrometry (LC-MS/MS), we simultaneously quantified extracellular levels of enkephalins and other neuropeptides (15) released from mouse brain slices (Fig. 2A and fig. S4). After stimulation with KCl (50 mM), we observed increased extracellular levels of MERF as well as Met-enkephalin and Leu-enkephalin, along with dynorphins and substance P (Fig. 2B). Concentrations of MERF released from isolated NAc tissue punches were higher than those from dorsal striatum tissue punches (fig. S5, A to D). We could not detect an appreciable release of angiotensin II (Fig. 2B) or bradykinin (fig. S5E). Enkephalin signals were absent in constitutive Penk knockout mice (fig. S5, F to K), which indicates that Penk is the primary source of enkephalin in this preparation.

Inhibition of ACE with 10 µM captopril robustly increased extracellular levels of MERF without affecting conventional enkephalins or other neuropeptides (Fig. 2C and table S2). We observed similar effects using trandolaprilat, a different ACE inhibitor that also caused LTD of excitatory synaptic input to D1-MSNs (fig. S6). In contrast, extracellular levels of MERF were not affected by pharmacological inhibition of aminopeptidase N and neprilysin, the enzymes responsible for degrading conventional enkephalins (10) (fig. S7A). A cocktail of inhibitors for all three enzymes blocked degradation of enkephalins as well as other neuropeptides (fig. S7B). D2-MSNs express high levels of Penk (6), making them a likely source of MERF. To evaluate this possibility, we bred mice with genetic expression of channelrhodopsin-2 in D2-MSNs (Fig. 2D and fig. S8). Optogenetic stimulation of acute brain slices from these mice increased extracellular levels of conventional enkephalins, but only MERF levels were elevated in the presence of captopril (Fig. 2, E and F, and table S3).

To investigate how MERF regulates NAc synaptic transmission, we measured miniature excitatory postsynaptic currents (mEPSCs; fig. S9). Increasing concentrations of MERF caused a dose-dependent decrease in mEPSC frequency without altering mEPSC amplitude (Fig. 3, A and B), consistent with a presynaptic reduction of glutamate release probability. Both MERF and Met-enkephalin (fig. S10, A to C) had similar effects on D1-MSNs and D2-MSNs, which suggests that presynaptic terminals onto both cell types are equally sensitive to endogenous opioids. We used these data to construct dose-response curves and found that MERF [half-maximal inhibitory concentration (IC_{50}) = 438 nM; Fig. 3C] was more potent than Met-enkephalin (IC_{50} = 993 nM; fig. S10D), as previously reported (13).

These experiments identified a threshold MERF concentration (100 nM) that did not reliably affect synaptic transmission. Captopril alone (10 µM) also had no effect on frequency or amplitude of mEPSCs, which are measured

[1]Graduate Program in Neuroscience, University of Minnesota Medical School, Minneapolis, MN, USA. [2]Medical Scientist Training Program, University of Minnesota Medical School, Minneapolis, MN, USA. [3]Department of Neuroscience, University of Minnesota Medical School, Minneapolis, MN, USA. [4]Center for Drug Design, College of Pharmacy, University of Minnesota, Minneapolis, MN, USA. [5]Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN, USA.
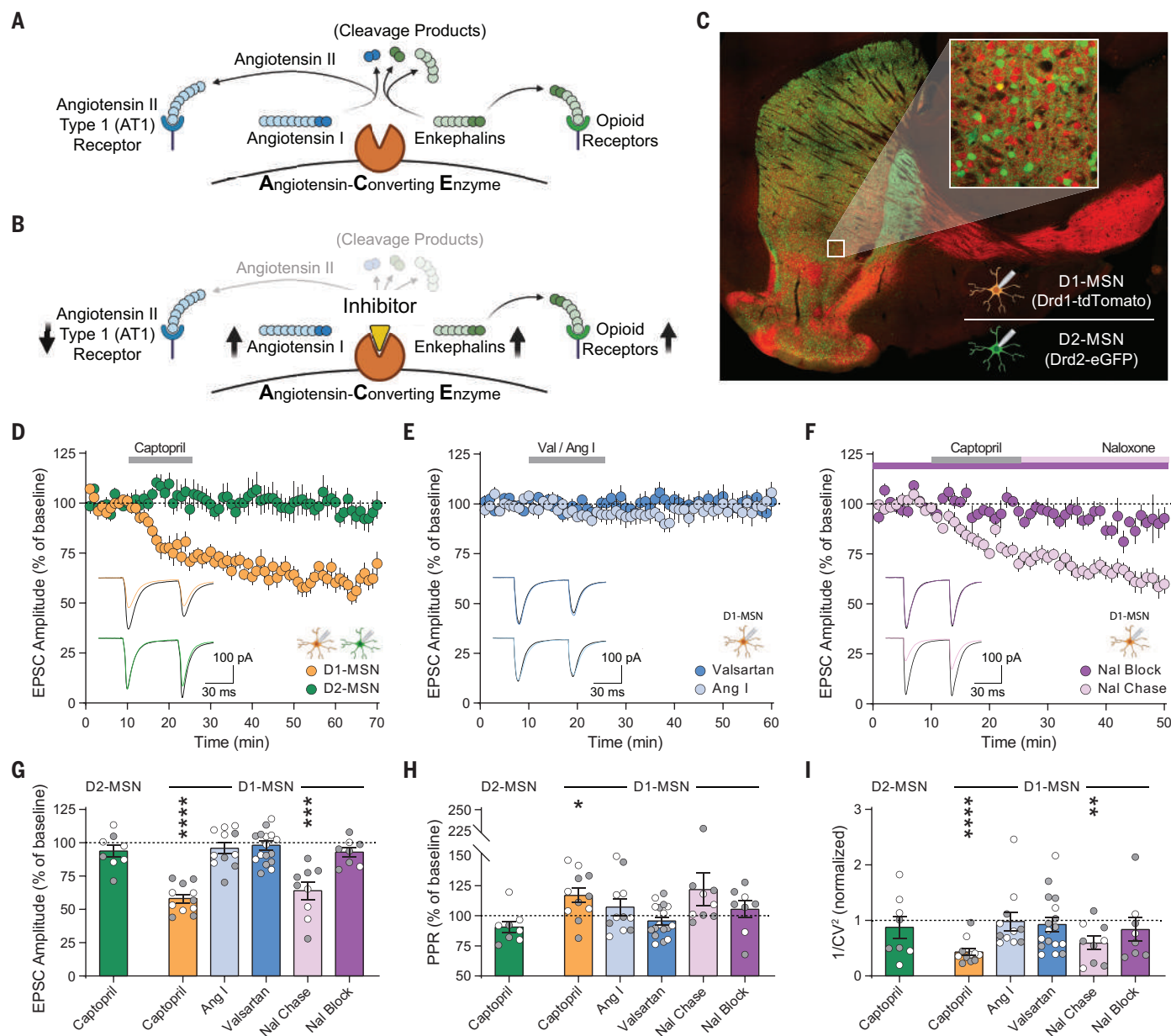*Corresponding author. Email: rothwell@umn.edu

Fig. 1. ACE inhibition reduces excitatory input to D1-MSNs via endogenous opioid signaling. (**A** and **B**) Schematic of angiotensin and enkephalin regulation by ACE, in the absence (A) and presence (B) of ACE inhibition. (**C**) Drd1-tdTomato expression (red) in D1-MSNs, and Drd2-eGFP expression (green) in D2-MSNs; box size is 250 μm × 250 μm. (**D** to **F**) EPSC amplitude before, during, and after 15-min bath perfusion (gray bar) of 10 μM captopril in D1-MSNs (orange, $n = 11$) or D2-MSNs (green, $n = 8$) (D); AT1R antagonist valsartan [dark blue, 2 μM ($n = 8$) and 20 μM ($n = 9$)] or 1 μM angiotensin I peptide (light blue, $n = 11$) in D1-MSNs (E); or 10 μM captopril in continual presence of opioid receptor antagonist naloxone (10 μM, dark purple, $n = 8$) or chased by naloxone (10 μM, light purple, $n = 9$) in D1-MSNs (F). Insets show traces before (black lines) and after (last 5 min of recording, colored lines). (**G** to **I**) EPSC parameters during the last 5 min of each recording, expressed as percentage of baseline prior to drug application: EPSC amplitude (G), paired-pulse ratio (H), and $1/CV^2$ (I). Data are means ± SEM for all panels; open and solid circles in (G) to (I) indicate recordings from female and male mice, respectively. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ [analysis of variance (ANOVA) followed by one-sample t test versus baseline]; see data S1 for complete statistics.

in the absence of stimulation required to release endogenous opioids (7, 9). However, the combination of captopril and a threshold MERF concentration caused synergistic depression of mEPSC frequency in D1-MSNs, but not D2-MSNs (Fig. 3, D and E, and fig. S11). This effect was absent after conditional genetic deletion of ACE from D1-MSNs (fig. S12). Inhib-itors of aminopeptidase N and neprilysin did not enhance the effects of MERF but did po-tentiate the effects of a threshold concentration of Met-enkephalin (100 nM) in both D1-MSNs and D2-MSNs (fig. S10, E to I).

To determine the opioid receptor subtype engaged by ACE inhibition, we recorded mEPSCs in D1-MSNs and applied captopril (10 μM) with threshold MERF (100 nM) in the presence of selective opioid receptor antagonists. Blocking δ-opioid receptors with SDM25N (500 nM) or κ-opioid receptors with NOR-BNI (100 nM) did not prevent the decrease in mEPSC frequency (Fig. 3F and fig. S13). However, this effect was completely blocked by the μ-opioid receptor (MOR) antagonist CTAP (1 μM), with no change
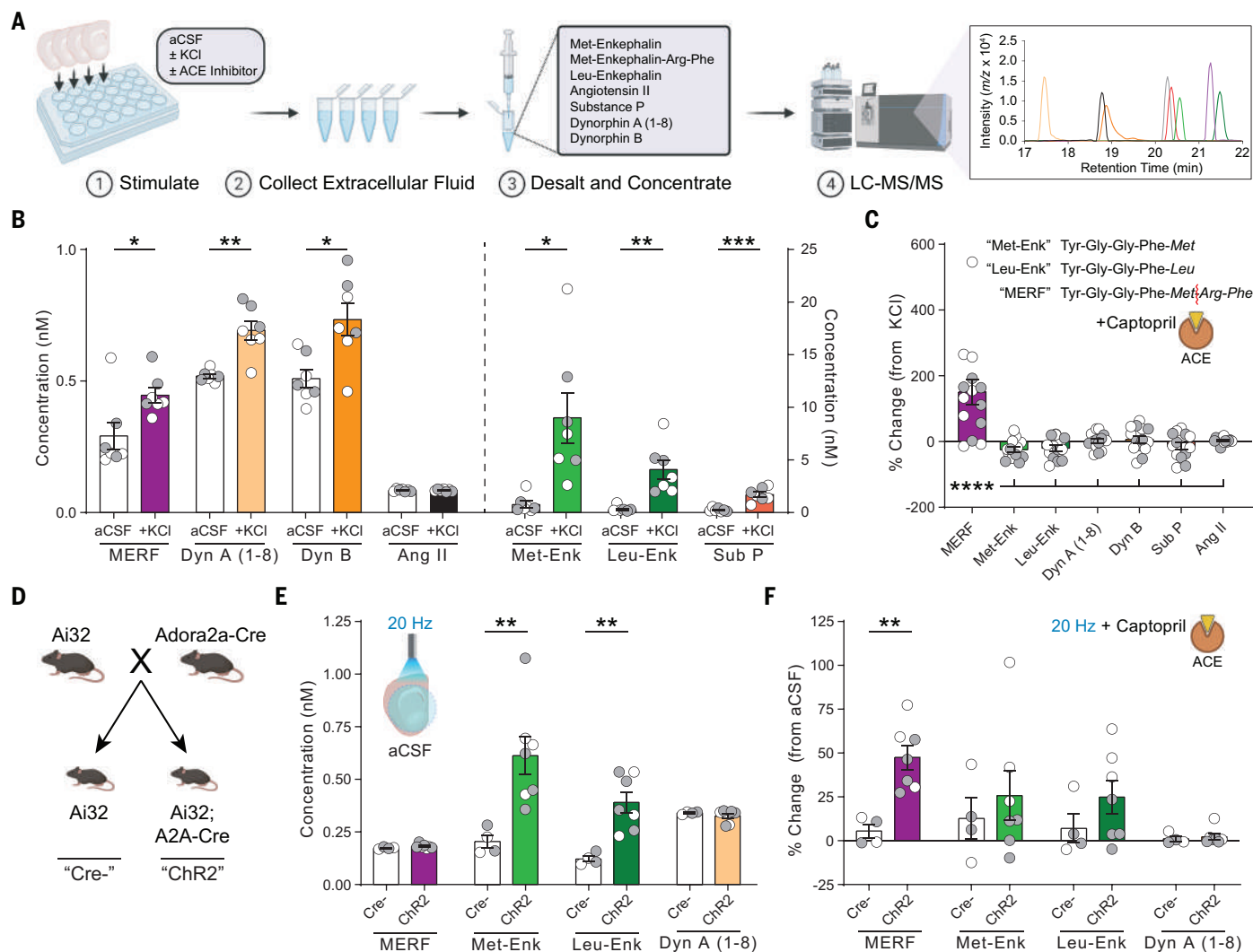
Fig. 2. ACE selectively degrades MERF in the extracellular space.
(**A**) Quantification of neuropeptide release from brain slices using LC-MS/MS.
(**B**) Extracellular neuropeptide levels from slices submerged in normal aCSF
or 50 mM KCl. (**C**) Percent change in extracellular neuropeptide levels after KCl
stimulation in presence versus absence of 10 μM captopril. Inset: Enkephalin
amino acid sequences and site of enzymatic cleavage of MERF by ACE (red line).
(**D**) Breeding strategy to generate mice expressing channelrhodopsin-2 in

D2-MSNs. (**E**) Extracellular neuropeptide levels from slices after optogenetic
stimulation at 20 Hz. (**F**) Percent change in extracellular neuropeptide levels after
optogenetic stimulation in presence versus absence of 10 μM captopril. Data
are means ± SEM for all panels; open and solid circles indicate samples from
female and male mice, respectively. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P <
0.0001$ [ANOVA followed by simple effect test in (B), (E), and (F) or Fisher's LSD
post hoc test in (C)]; see data S1 for complete statistics.

in mEPSC amplitude (Fig. 3F and fig. S14, A
to D). To confirm the role of MOR, we crossed
Drd1-tdTomato reporter mice with constitutive
MOR (*Oprm1*) knockout mice, generating
offspring that lacked functional MOR (Oprm1$^{-/-}$)
as well as littermate controls (Oprm1$^{+/+}$) (fig.
S14E). The synergistic effect of captopril (10 μM)
and threshold MERF (100 nM) on mEPSC fre-
quency in D1-MSNs was absent from Oprm1$^{-/-}$
mice (Fig. 3, G and H, and fig. S14, F to I).
Captopril-LTD of evoked EPSCs in D1-MSNs
was also absent from Oprm1$^{-/-}$ mice (Fig. 3, I
and J). MOR is also expressed postsynaptically
by most D1-MSNs and fewer D2-MSNs (*16, 17*).
In current-clamp recordings, 1 μM MERF de-
creased action potential firing in D1-MSNs

(Fig. 3K) but not D2-MSNs (fig. S15A). In com-
bination with 10 μM captopril, a threshold
MERF concentration (100 nM) also decreased
action potential firing in D1-MSNs (Fig. 3L)
but not D2-MSNs (fig. S15B).

Our experiments in brain slices have shown
that ACE inhibition reduces excitatory synap-
tic input to D1-MSNs (Fig. 4A). To complement
these analyses, we used fiber photometry
in vivo and found that systemic captopril ad-
ministration reduced the sensitivity of D1-MSNs
to optogenetic stimulation of excitatory input
arising from the medial prefrontal cortex (mPFC;
Fig. 4, B to D, and fig. S16). Because the re-
warding effects of addictive drugs are driven
by D1-MSN activity and strengthening of ex-

citatory synaptic input (*18–22*), we used an
unbiased place-conditioning assay to determine
whether systemic captopril administration could
counteract the rewarding properties of fenta-
nyl (Fig. 4E). Mice exhibited robust conditioned
place preference (CPP) for a fentanyl-paired
context [0.04 mg/kg subcutaneously (s.c.)],
but the magnitude of CPP was significantly
attenuated when captopril [30 mg/kg intra-
peritoneally (i.p.)] was injected before fentanyl
(Fig. 4, F and G). Trandolapril (the prodrug
form of trandolaprilat) had a similar effect on
fentanyl CPP (fig. S17). Captopril itself was not
rewarding or aversive in the place-conditioning
assay (Fig. 4, H to J) and did not alter loco-
motion during conditioning (fig. S18). In a test

Fig. 3. Captopril enhances MERF effects on presynaptic and postsynaptic opioid receptors.
(**A**) Top: mEPSCs from D1-MSNs (left) and D2-MSNs (right) before and after bath perfusion of 10 μM MERF. Bottom: Cumulative fraction plots of inter-event interval (left) and amplitude (right) of mEPSCs at increasing MERF concentrations (0.01 to 10 μM). (**B**) MERF causes a dose-dependent decrease in mEPSC frequency in D1-MSNs (left, orange, n = 8) and D2-MSNs (right, green, n = 9). (**C**) Sigmoidal interpolation of MERF dose response normalized to maximal frequency change at 10 μM (IC$_{50}$, 438 nM; 95% confidence interval, 279 to 690 nM; n = 17). (**D** and **E**) mEPSC frequency (D) and amplitude (E) after combined captopril (10 μM) and/or threshold MERF (100 nM) in D1-MSNs (left, n = 14) and D2-MSNs (right, n = 12). (**F**) Combined effect of captopril and threshold MERF in the presence of selective antagonists of δ (SDM25N, 500 nM, blue, n = 9), κ (NOR-BNI, 100 nM, green, n = 11), or μ (CTAP, 1 μM, orange, n = 12) opioid receptors. (**G** and **H**) Combined effect of captopril and threshold MERF on mEPSC frequency (G) and amplitude (H) in Oprm1$^{-/-}$ knockout mice (gray, n = 8) and Oprm1$^{+/+}$ littermates (purple, n = 8). (**I** and **J**) EPSC amplitude time course (I) or average during last 5 min (J) of captopril-LTD in Oprm1$^{+/+}$ (orange, n = 8) and Oprm1$^{-/-}$ mice (gray, n = 9). Inset shows traces before captopril (black lines) and during last 5 min (colored lines). (**K**) Action potential firing rate of D1-MSNs (n = 5 to 7) before and after exposure to MERF (0.1 to 1 μM). (**L**) Change in action potential firing rate of D1-MSNs (n = 3 to 7) at 120 pA after combined captopril (10 μM) and/or threshold MERF (100 nM). Data are means ± SEM for all panels; open and solid circles indicate recordings from female and male mice, respectively. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 [concentration main effect in (B), treatment simple effect in D1-MSNs in (D), genotype or treatment main effect in (F), (J), (K), and (L), two-sample t test in (G)]; see data S1 for complete statistics.
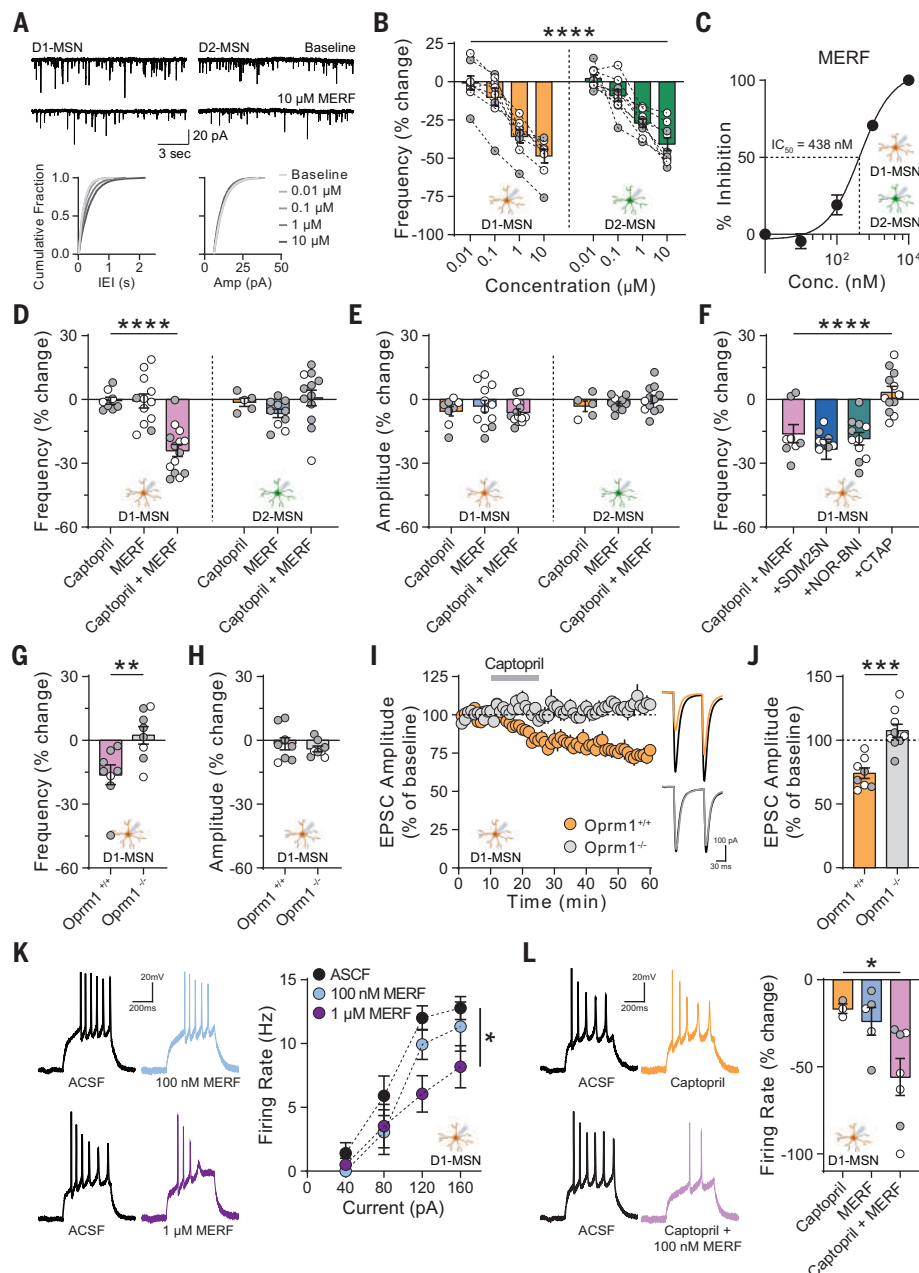
of social interaction between two freely moving mice, captopril administration increased the amount of social interaction (Fig. 4, K to O), which is consistent with enhanced MOR signaling in the NAc (22) and rules out a general disruption of motivated behavior.

Targeting molecules with enriched expression in specific circuit elements is one strategy for translating an increasingly precise understanding of neural circuit function into therapeutic advances (1). Our data show that enriched expression of ACE by D1-MSNs can be leveraged to induce synaptic plasticity in a brain circuit–specific fashion (fig. S19). Pharmacological inhibition of ACE prevents degradation of MERF, thereby enhancing endogenous MOR signaling in the NAc (Fig. 4A). This resembles the

effects of selective reuptake inhibitors for other neurotransmitters, which have substantial therapeutic value for brain disorders. The circuit specificity of these effects likely results from the high levels of MERF in the NAc, combined with enriched expression of ACE by D1-MSNs. ACE inhibition did not induce synaptic plasticity at excitatory synapses onto D2-MSNs or layer V pyramidal cells in the ACC, even though these synapses were sensitive to exogenous MERF application (Fig. 3B and fig. S3). By selectively and locally enhancing endogenous opioid signaling in the vicinity of D1-MSNs, central ACE inhibition may limit abuse liability by avoiding MOR activation in other brain circuits. Indeed, systemic ACE inhibition significantly reduced the rewarding effects of

fentanyl and increased reciprocal social interaction. Conversely, rodents that exhibit reduced social interaction after chronic social stress have up-regulated ACE expression in NAc tissue (23) and D1-MSNs (24). This behavioral phenotype is reversed by treatment with antidepressant drugs (25), and human patients taking centrally active ACE inhibitors can experience relief from depression (26–28) as well as improved quality of life (29, 30) and slower cognitive decline (31). Together, this evidence suggests that central ACE inhibition could have therapeutic potential for a variety of brain conditions. Our findings may thus herald a new era of repositioning and redesigning ACE inhibitors with central activity as a brain circuit–specific pharmacotherapy.
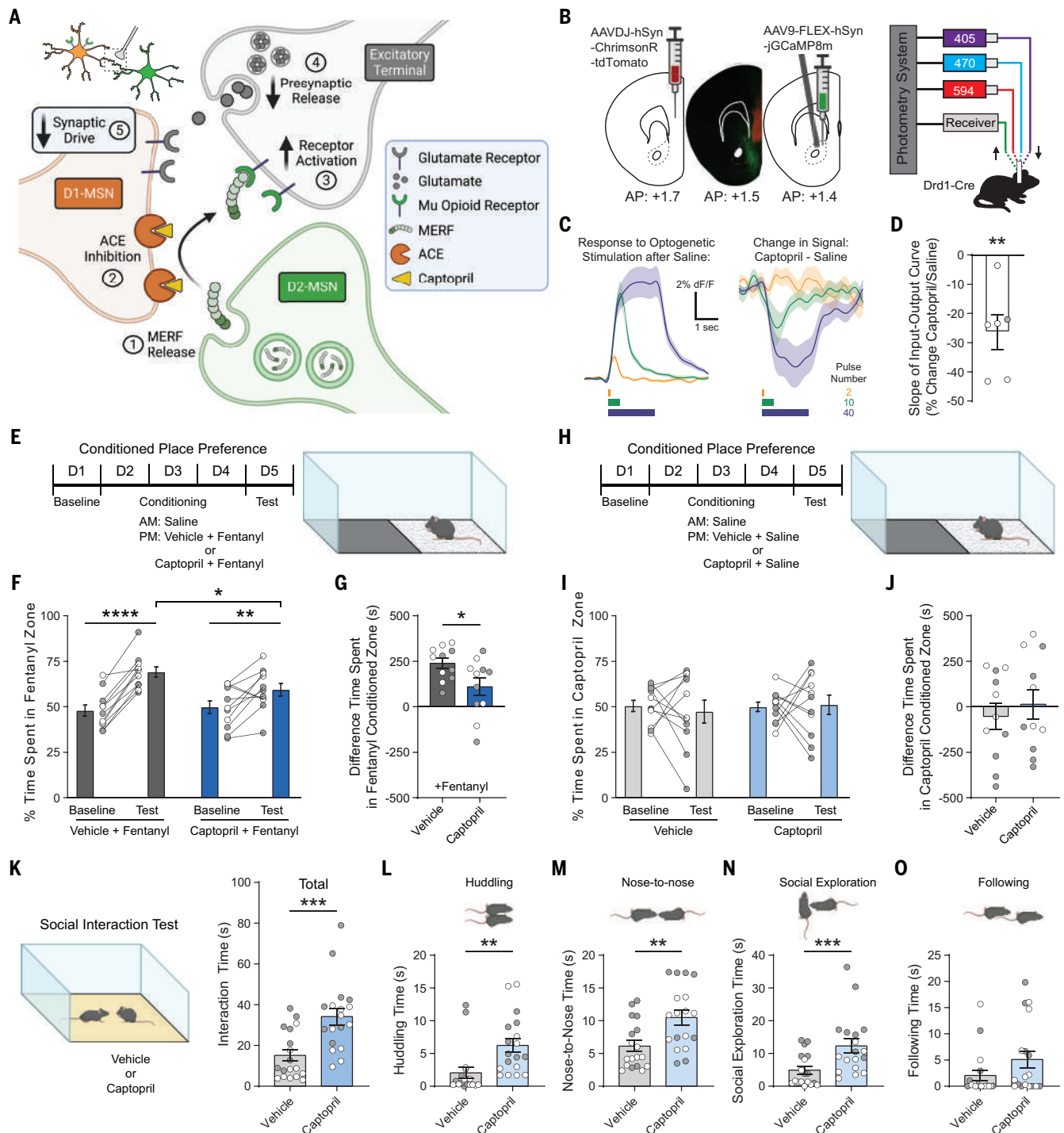
Fig. 4. Systemic captopril reduces excitatory input to D1-MSNs, counteracts fentanyl reward, and increases sociability. (**A**) Proposed mechanism by which captopril regulates glutamate release onto D1-MSNs via MERF. (**B**) Left: Schematic showing viral injection of ChrimsonR-tdTomato in mPFC and Cre-dependent GCaMP8m in NAc, separated by fluorescent image showing viral expression. Right: Setup for simultaneous optogenetic stimulation (594 nm) and fiber photometry recording (405/470 nm). (**C**) Left: Traces showing average response to 2, 10, and 40 pulses of red light at 20 Hz after injection of saline. Right: Average change in response after injection of captopril (30 mg/kg i.p.). (**D**) Percent change in slope of the input-output curve after injection of captopril versus saline ($n = 6$). (**E** to **G**) Schematic of unbiased place conditioning procedure (E), with percent time on fentanyl side (F) and CPP score (G) for groups receiving fentanyl (0.04 mg/kg s.c.)

preceded by vehicle ($n = 11$, dark gray) or captopril (30 mg/kg i.p.; $n = 11$, dark blue). (**H** to **J**) Schematic of unbiased place conditioning procedure (H), with percent time on fentanyl side (I) and CPP score (J) for groups receiving saline preceded by vehicle ($n = 11$, gray) or captopril (30 mg/kg i.p.; $n = 11$, blue). (**K**) Left: Schematic of reciprocal social interaction test after injection of vehicle or captopril (30 mg/kg i.p.). Right: Total social interaction time after vehicle ($n = 18$, gray) or captopril ($n = 18$, blue). (**L** to **O**) Time spent huddling (L), interacting nose-to-nose (M), socially exploring (N), or following (O) the partner mouse throughout the assay. Data are means ± SEM for all panels; open and solid circles indicate female and male mice, respectively. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ [one-sample t test in (D), simple effect of session/treatment in (F), treatment main effect in (G) and (K) to (N)]; see data S1 for complete statistics.

REFERENCES AND NOTES

1. J. A. Gordon, *Nat. Neurosci.* **19**, 1385–1386 (2016).
2. P. E. Rothwell *et al.*, *Cell* **158**, 198–212 (2014).
3. N. Schwartz *et al.*, *Science* **345**, 535–542 (2014).
4. M. Creed, V. J. Pascoli, C. Lüscher, *Science* **347**, 659–664 (2015).
5. S. M. Strittmatter, M. M. Lo, J. A. Javitch, S. H. Snyder, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1599–1603 (1984).
6. A. Saunders *et al.*, *Cell* **174**, 1015–1030.e16 (2018).
7. B. K. Atwood, D. A. Kupferschmidt, D. M. Lovinger, *Nat. Neurosci.* **17**, 540–548 (2014).
8. M. A. Ondetti, B. Rubin, D. W. Cushman, *Science* **196**, 441–444 (1977).
9. C. P. Blomeley, E. Bracci, *J. Neurosci.* **31**, 13346–13356 (2011).
10. B. P. Roques, M. C. Fournié-Zaluski, M. Wurm, *Nat. Rev. Drug Discov.* **11**, 292–310 (2012).
11. K. Ploj, E. Roman, L. Gustavsson, I. Nylander, *Brain Res. Bull.* **53**, 219–226 (2000).
12. A. Mansour, M. T. Hoversten, L. P. Taylor, S. J. Watson, H. Akil, *Brain Res.* **700**, 89–98 (1995).
13. C. E. Inturrisi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5512–5514 (1980).
14. M. Benuck, M. J. Berg, N. Marks, *Neurochem. Int.* **4**, 389–396 (1982).
15. R. Al-Hasani *et al.*, *eLife* **7**, e36520 (2018).
16. M. R. Banghart, S. Q. Neufeld, N. C. Wong, B. L. Sabatini, *Neuron* **88**, 1227–1239 (2015).
17. P. Charbogne *et al.*, *Biol. Psychiatry* **81**, 778–788 (2017).
18. M. K. Lobo *et al.*, *Science* **330**, 385–390 (2010).
19. J. W. Koo *et al.*, *Neuropsychopharmacology* **39**, 2646–2653 (2014).
20. V. Pascoli *et al.*, *Nature* **509**, 459–464 (2014).
21. E. S. Calipari *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2726–2731 (2016).
22. V. Trezza, R. Damsteegt, E. J. Achterberg, L. J. Vanderschuren, *J. Neurosci.* **31**, 6362–6370 (2011).
23. H. Nam *et al.*, *Neuropsychopharmacology* **44**, 1876–1885 (2019).
24. H.-D. Kim *et al.*, *Mol. Psychiatry* **26**, 7316–7327 (2021).
25. O. Berton *et al.*, *Science* **311**, 864–868 (2006).
26. G. S. Zubenko, R. A. Nixon, *Am J. Psychiatry* **141**, 110–111 (1984).
27. R. F. Deicken, *Biol. Psychiatry* **21**, 1425–1428 (1986).
28. L. Germain, G. Chouinard, *Biol. Psychiatry* **23**, 637–641 (1988).
29. S. H. Croog *et al.*, *N. Engl. J. Med.* **314**, 1657–1664 (1986).
30. M. A. Testa, R. B. Anderson, J. F. Nackley, N. K. Hollenberg, Quality-of-Life Hypertension Study Group, *N. Engl. J. Med.* **328**, 907–913 (1993).
31. K. M. Sink *et al.*, *Arch. Intern. Med.* **169**, 1195–1202 (2009).

CELL BIOLOGY

Live cell tracking of macrophage efferocytosis during *Drosophila* embryo development in vivo

Michael H. Raymond[1,2,3]†, Andrew J. Davidson[4]†, Yi Shen[5], Daniel R. Tudor[6], Christopher D. Lucas[4], Sho Morioka[1,2,7], Justin S. A. Perry[8], Julia Krapivkina[9], David Perrais[9], Linus J. Schumacher[6], Robert E. Campbell[5,10], Will Wood[4]*, Kodi S. Ravichandran[1,2,11,12]*

Apoptosis of cells and their subsequent removal through efferocytosis occurs in nearly all tissues during development, homeostasis, and disease. However, it has been difficult to track cell death and subsequent corpse removal in vivo. We developed a genetically encoded fluorescent reporter, CharON (Caspase and pH Activated Reporter, Fluorescence ON), that could track emerging apoptotic cells and their efferocytic clearance by phagocytes. Using *Drosophila* expressing CharON, we uncovered multiple qualitative and quantitative features of coordinated clearance of apoptotic corpses during embryonic development. When confronted with high rates of emerging apoptotic corpses, the macrophages displayed heterogeneity in engulfment behaviors, leading to some efferocytic macrophages carrying high corpse burden. Overburdened macrophages were compromised in clearing wound debris. These findings reveal known and unexpected features of apoptosis and macrophage efferocytosis in vivo.

E fferocytosis—the uptake and degradation of apoptotic cells by phagocytes—is essential for tissue development, homeostasis, and resolution of inflammation (*1*, *2*). Previous work has characterized efferocytic receptors on phagocytes and their corresponding ligands on dying cells; however, tracking efferocytosis in vivo is challenging (*3–5*). Even detecting apoptotic cells in vivo is difficult owing to their swift removal and the lack of tools with which to track them as they emerge.

To track both apoptosis and efferocytosis in vivo, we engineered two probes. For the apoptosis reporter, we tested multiple designs in which the cleavage by executioner caspases 3 and 7 induced fluorescence of a green fluorescent protein (GFP) (Fig. 1A; figs. S1, A to G, and S2, A to D; and supplementary ma-

terials, materials and methods) (*6–9*). A limitation of GFP-based probes is photoquenching in the acidic environment of lysosomes. To circumvent this, we engineered pH-CaspGFP to be pH-tolerant through mutation (Q204H, in which gluatmine at position 204 is replaced with histidine), based on the parent GC3ai molecule (Fig. 1B) (*10*). pH-CaspGFP faithfully reported apoptosis after different triggers [ultraviolet (UV)–, staurosporine-, or ABT-737–induced apoptosis) and in different cell types (Jurkat T cells, THP-1 monocytes, and LR73 fibroblasts) (figs. S1, A to G, and S2, A to D). The pH-CaspGFP signal induction aligned well with annexin V staining of phosphatidylserine exposure on the apoptotic cells and was sensitive to caspase inhibition (fig. S1G).

To detect corpse acidification and digestion within phagosomes, we engineered a red-fluorescent pH sensor, pHlorina (pronounced "florina"), that exhibits increasing fluorescence with decreasing pH (Fig. 1C; fig. S3, A to E; and materials and methods). When shifted from pH 7.5 to pH 5, pHlorina exhibits >20-fold excitation ratiometric change and greater than sixfold red fluorescence intensity increase (excitation, 560 nm; emission, 620 nm). To assess pHlorina during efferocytosis, we stably transduced Jurkat cells, stimulated apoptosis, and incubated these as targets with mouse J774 macrophages (Fig. 1D and fig. S4A). Upon corpse internalization, we detected a 3.4-fold increase in pHlorina fluorescence (Fig. 1D), which was inhibited by the acidification inhibitor Bafilomycin or *Rubicon* suppression (Fig. 1D and fig. S4, B and C). Thus, pHlorina can track internalized apoptotic corpses during acidification and digestion within phagocytes.

Next, we combined the above two probes to generate "CharON" (Caspase and pH Activated

¹Center for Cell Clearance, University of Virginia, Charlottesville, VA, USA. ²Department of Microbiology, Immunology, and Cancer Biology, University of Virginia, Charlottesville, VA, USA. ³Neuroscience Graduate Program, University of Virginia, Charlottesville, VA, USA. ⁴Centre for Inflammation Research, University of Edinburgh, Edinburgh BioQuarter, Edinburgh, UK. ⁵Department of Chemistry, University of Alberta, Edmonton, Canada. ⁶Centre for Regenerative Medicine, University of Edinburgh, Edinburgh BioQuarter, Edinburgh, UK. ⁷Department of Medicine, Division of Nephrology and CIIR, University of Virginia, Charlottesville, VA, USA. ⁸Immunology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. ⁹University of Bordeaux, CNRS, Interdisciplinary Institute for Neuroscience (IINS), UMR 5297, Bordeaux, France. ¹⁰Department of Chemistry, School of Science, The University of Tokyo, Bunkyo-ku, Tokyo, 113-0033, Japan. ¹¹VIB/UGent Inflammation Research Centre, and Biomedical Molecular Biology, Ghent University, Belgium. ¹²Division of Immunobiology, Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA.
*Corresponding author. Email: ravi@virginia.edu (K.S.R.); w.wood@ed.ac.uk (W.W.)
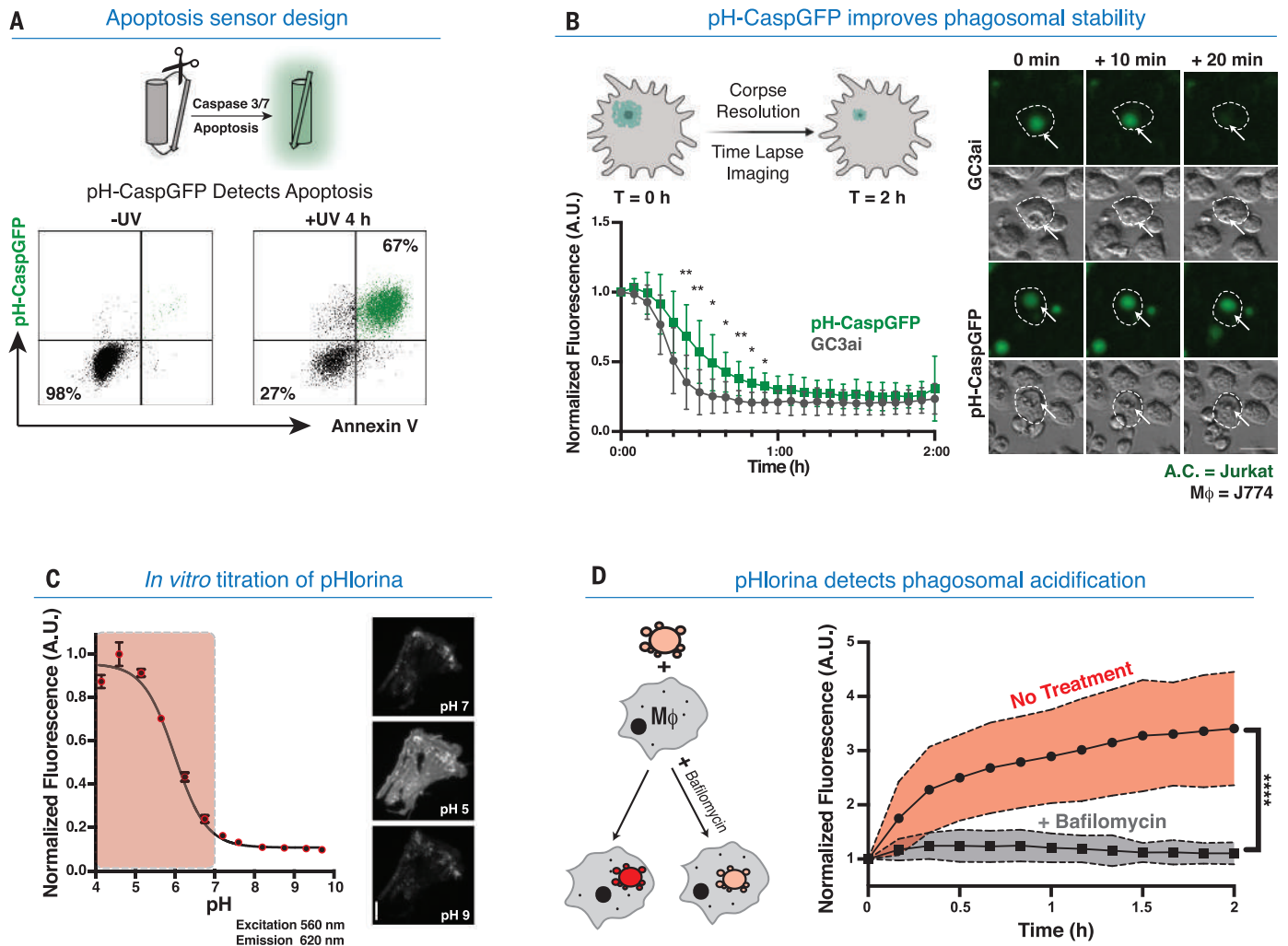†These authors contributed equally to this work.

Fig. 1. Engineering a pH-stable apoptosis reporter and a red fluorescent protein pH sensor. (**A**) Apoptosis sensor design. (Top) Upon apoptosis, a Caspase-3/7 linker is cleaved, promoting GFP fluorescence. (Bottom) GFP/Annexin V positivity of pH-CaspGFP–expressing Jurkat cells ± apoptosis (4 hours after UV-C). (**B**) pH-CaspGFP exhibits improved phagosomal stability. (Top left) Schematic of in vitro engulfment assay. GC3ai or pH-CaspGFP apoptotic Jurkat cells were cocultured with J774 macrophages. (Bottom left) Upon internalization, GFP fluorescence intensity was tracked. (Right) Time-lapse images of (top) GC3ai or (bottom) pH-CaspGFP apoptotic Jurkat cells engulfed by J774 macrophages (arrows).

Three independent experiments, two-way analysis of variance (ANOVA), *$P < 0.0332$, **$P < 0.0021$. (**C**) In vitro validation of pHlorina. (Left) pH titration of purified pHlorina from pH 4 to 10 (excitation/emission = 560/620 nm). (Right) pHlorina fluorescence at pH 7.0, 5.0, and 9.0. Three independent experiments. (**D**) pHlorina detects phagosomal acidification during efferocytosis. (Left) Schematic of in vitro engulfment assay to track pHlorina fluorescence during efferocytosis ± bafilomycin (to inhibit phagosomal acidification). (Right) pHlorina signal in apoptotic Jurkat cells after engulfment by J774 macrophages ± bafilomycin. Three independent experiments, unpaired Student's t test, **** = $P < .0001$. All scale bars = 50 μm.

Reporter, Fluorescence ON, referring to the character in Greek mythology who ferries the deceased across the river to the afterlife) (Fig. 2A and materials and methods). Using CharON, an apoptotic cell first turns GFP[+]. After engulfment by a phagocyte and acidification within the phagolysosome, the intensity of pHlorina increases as the GFP gradually quenches (Fig. 2A). When apoptotic CharON-expressing Jurkat cells were mixed with blue fluorescent protein–positive (BFP[+]) mouse J774 macrophages or mouse bone marrow–derived macrophages, CharON permitted all the stages of efferocytosis to be visualized (figs. S4, D and E, and movie S1). Both fluorescent components of CharON exhibited wide dynamic ranges (Fig. 2B), with a 17-fold increase in average pHlorina/pH-CaspGFP ratio within 2 hours of corpse internalization (Fig. 2B). Furthermore, CharON visualized successive efferocytic events, in which macrophages engulfed multiple corpses (fig. S4F and movie S2). Thus, CharON is a powerful tool for tracking apoptosis and efferocytosis.

To track apoptosis and efferocytosis in vivo, we generated CharON transgenic *Drosophila*. During mid-late *Drosophila* embryogenesis (stages 12 to 16), there is a wave of apoptosis in the developing central nervous system (CNS) (*11, 12*). CharON allowed visualization of developmental apoptosis and efferocytosis

throughout the CNS (Fig. 2C; fig. S5, A to D; and movie S3). The apoptotic burden in embryonic *Drosophila* CNS is shared by the phagocytic glia and the dispersing ventral hemocytes (macrophages) (Fig. 2C and fig. S6A) (*13, 14*). The highly motile macrophages disperse through the hemocoel ("blood-cavity"), clearing apoptosis at the CNS interface (*15, 16*). Efferocytosis deep within the CNS was sporadic, representing constant autonomous engulfment by phagocytic nonmotile glia, and conferred a weak trend for increasing CharON acidic fluorescence (Fig. 2D). By contrast, the macrophage-mediated efferocytosis at the CNS interface was visualized as a choreographed wave of
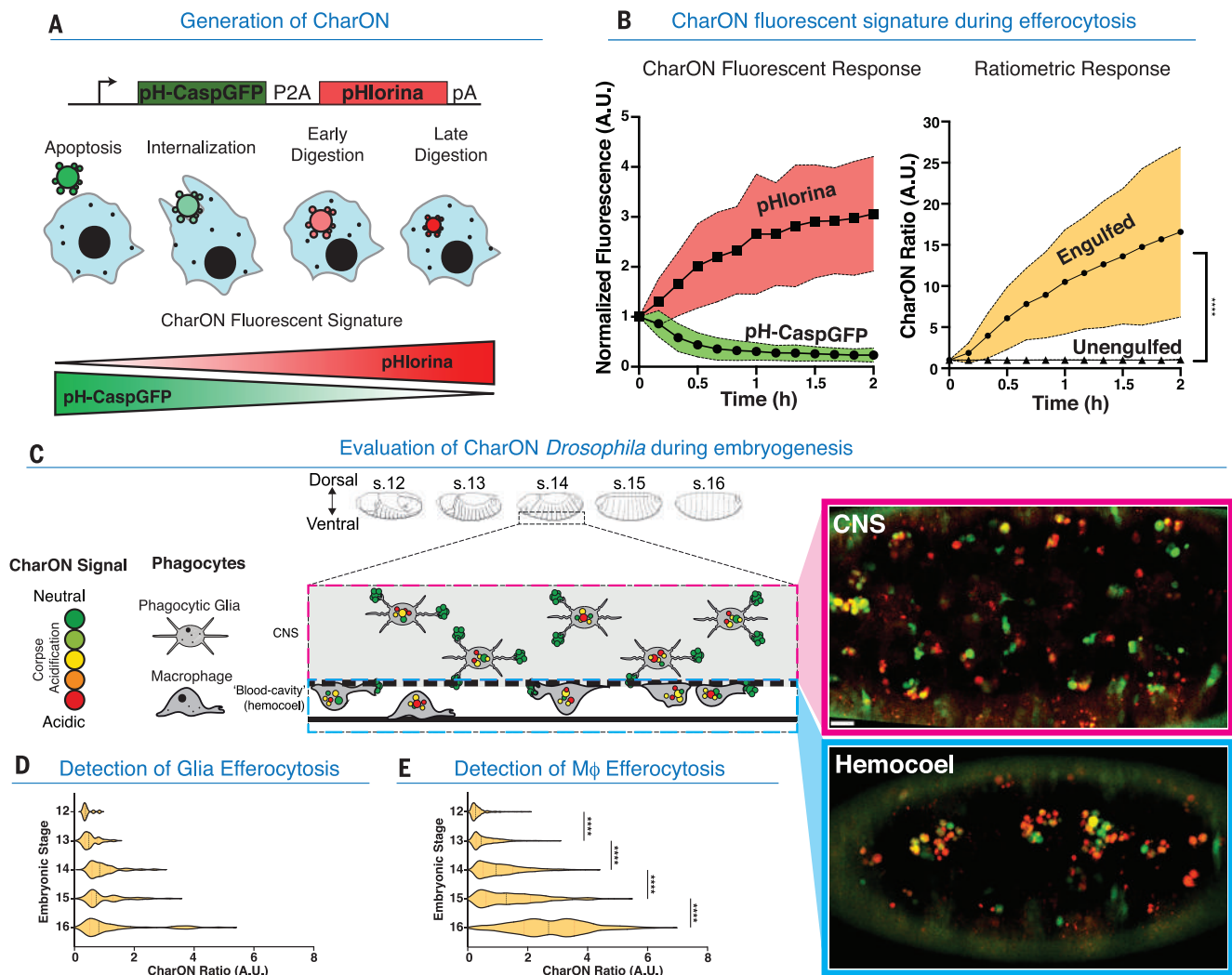
Fig. 2. CharON detects apoptosis and efferocytosis in vitro and in vivo. (**A**) CharON design and rationale. (Top) CharON construct design. (Bottom) Fluorescence of an apoptotic CharON-expressing cell (green or red) engulfed by a macrophage (blue). (**B**) CharON fluorescence during efferocytosis. (Left) CharON pH-CaspGFP and pHlorina fluorescence during engulfment. (Right) Ratiometric CharON (pHlorina/pH-CaspGFP) signal during engulfment. Three independent experiments, two-way ANOVA, Šidák's multiple comparison test, ****$P < 0.0001$. (**C**) (Left) CharON activity during *Drosophila* embryogenesis. Embryo outlines highlight morphological changes during mid- to late-stage embryogenesis (stages 12 to 16)

(*36*). Extensive apoptosis occurs within the developing CNS (dashed black box), which is cleared by phagocytic glia (dashed pink box). The ventral-most corpses at the interface between the CNS and the underlying blood-cavity (hemocoel) are cleared by the macrophages (dashed blue box). (Right) CharON visualizes efferocytosis within (top) the CNS (pink box) and (bottom) the hemocoel of a stage 14 embryo. (**D** and **E**) Efferocytosis increases during CNS development. CharON ratio (pHlorina signal/pH-CaspGFP signal) of individual corpses within (D) CNS and (E) hemocoel during embryogenesis (stages 12 to 16). Five embryos per stage, one-way ANOVA, ****$P < 0.0001$. Scale bars, 10 μm.

apoptosis and efferocytosis, yielding a strong increase in pHlorina signal (Fig. 2E and movie S3). CharON also highlighted differences in phagosome size between phagocytic glia and macrophages (fig. S5, E and F), with more "compacted" corpse sizes in macrophages.

We focused on macrophage efferocytosis owing to its synchronized pattern of CharON signal and the importance of macrophages in clearing cellular debris during human development and disease. A macrophage marker was combined with CharON to visualize macrophage-mediated efferocytosis in vivo (movie S4). Although the GFP signal of the macrophage marker overlapped with the pHCaspGFP of CharON, apoptotic cells were easily distinguished by differences in morphologies and signal intensity (movie S4). We also used an mCherry macrophage marker to distinguish early interactions between phagocyte and apoptotic corpses (fig. S6B). CharON permitted the full efferocytic program to be observed in vivo, including apoptosis, macrophage recruitment, and target binding ("Find Me"); internalization ("Eat Me"); and corpse acidification and degradation ("Digest Me") (Fig. 3A, fig. S6C, and movie S5). Quantification of pHlorina-positive corpses within mac-

rophages revealed that corpse burden peaked after dispersal (stage 15, after clearance) (Fig. 3, B and C). It was only after an initial lag that internalized corpses increased pHlorina intensity and decreased in size (degradation) (Fig. 3, D and E).
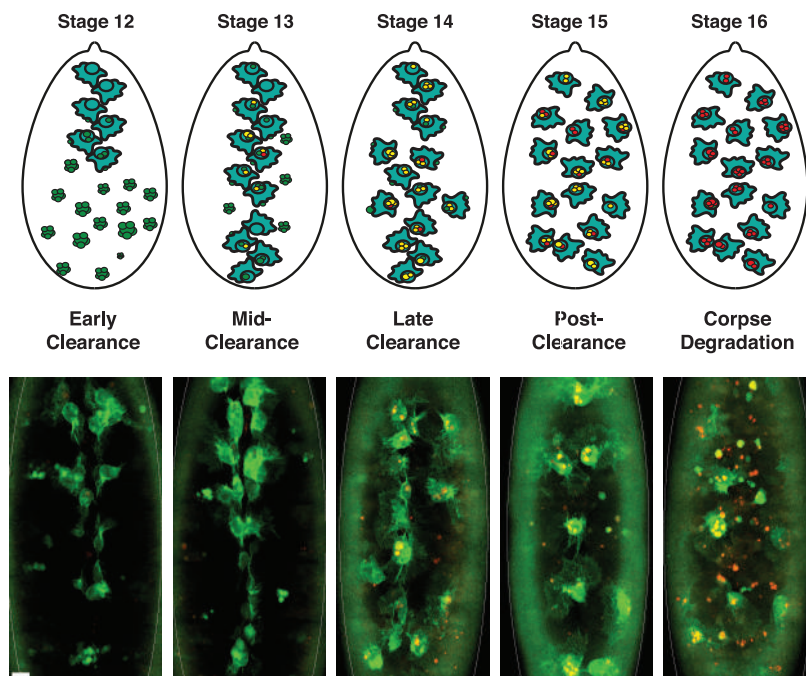
Tracking macrophages during efferocytosis revealed that macrophages sensed and migrated up to 8% of the embryo length toward apoptosis, regardless of preexisting corpse burden (figs. S6, D to F) (*17*). At times, multiple macrophages were required to clear a single, fragmenting apoptotic corpse (movie S6). The recruitment and uptake of apoptotic corpses
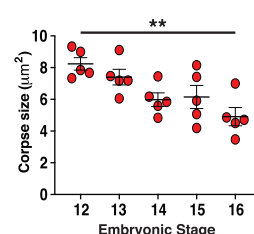
Fig. 3. CharON illuminates macrophage-mediated efferocytosis in vivo. (**A**) CharON visualizes the different stages of efferocytosis. A GFP-labeled macrophage (Mφ; blue dashed outline) engulfs a CharON-labeled apoptotic corpse (A.C.; dashed outline) within the *Drosophila* embryo (stage 12). Apoptosis induces pH-CaspGFP (green) activation and macrophage (green) attraction, leading to target binding and uptake. After internalization, acidification and degradation of the corpse is detected through increasing pHlorina signal (red). (**B**) Efferocytosis increases during macrophage dispersal. (Top) Diagrams and (bottom) images highlighting the ventral dispersal of *Drosophila* macrophages (GFP; green) within the embryo (stages 12 to 16). During their stereotyped dispersal, macrophages clear CharON-labeled apoptotic corpses (green and red). The additional pHlorina-positive corpses are within unlabeled phagocytic glia deeper within the CNS. (**C** to **E**) Mean corpse number per macrophage [(C), burden], pHlorina intensity per macrophage [(D), acidification], and corpse area per macrophage [(E), size] across stages 12 to 16 (embryo averages, 5 embryos per stage). One-way ANOVA, **$P < 0.0021$, ***$P < 0.0002$, ****$P < 0.0001$, error bars = SEM. Scale bars, 10 μm.

took less time (~5 to 7 min) than the time required for maximal corpse acidification (~25 min), suggesting that corpse degradation may be rate limiting for further uptake of additional apoptotic corpses (Fig. 3A and fig. S6C). Macrophages maintained their high motility throughout efferocytosis and readily moved toward and engulfed successive corpses before acidification of their existing or newly acquired corpse (fig. S6, D to F, and movie S4). Thus, macrophages adopt an "eat-first, digest-later" strategy, prioritizing debris uptake over corpse acidification and degradation, which in part explains why apoptotic cells are rarely seen in vivo, even in tissues with high cell turnover.

The first "pioneer" macrophages (early stage 12), which initially disperse from the head region, were confronted by a dense field of apoptotic corpses (fig. S7A and movie S4). The rapid uptake of these corpses by the dispersing macrophages resulted in large disparities in macrophage corpse burden (Fig. 4A and fig. S7, B and C). We classified this functional heterogeneity between macrophages as having either no, low (one to three corpses), medium (four to six corpses), or high seven or more corpses) burden (Fig. 4B). Examples of macrophages with extreme burden (10 or more; sometimes up to 20 corpses) were also evident. High burden eventually led to increased corpse acidification, suggesting efferocytic adaptation, as shown previously (fig. S7D) (18–20).

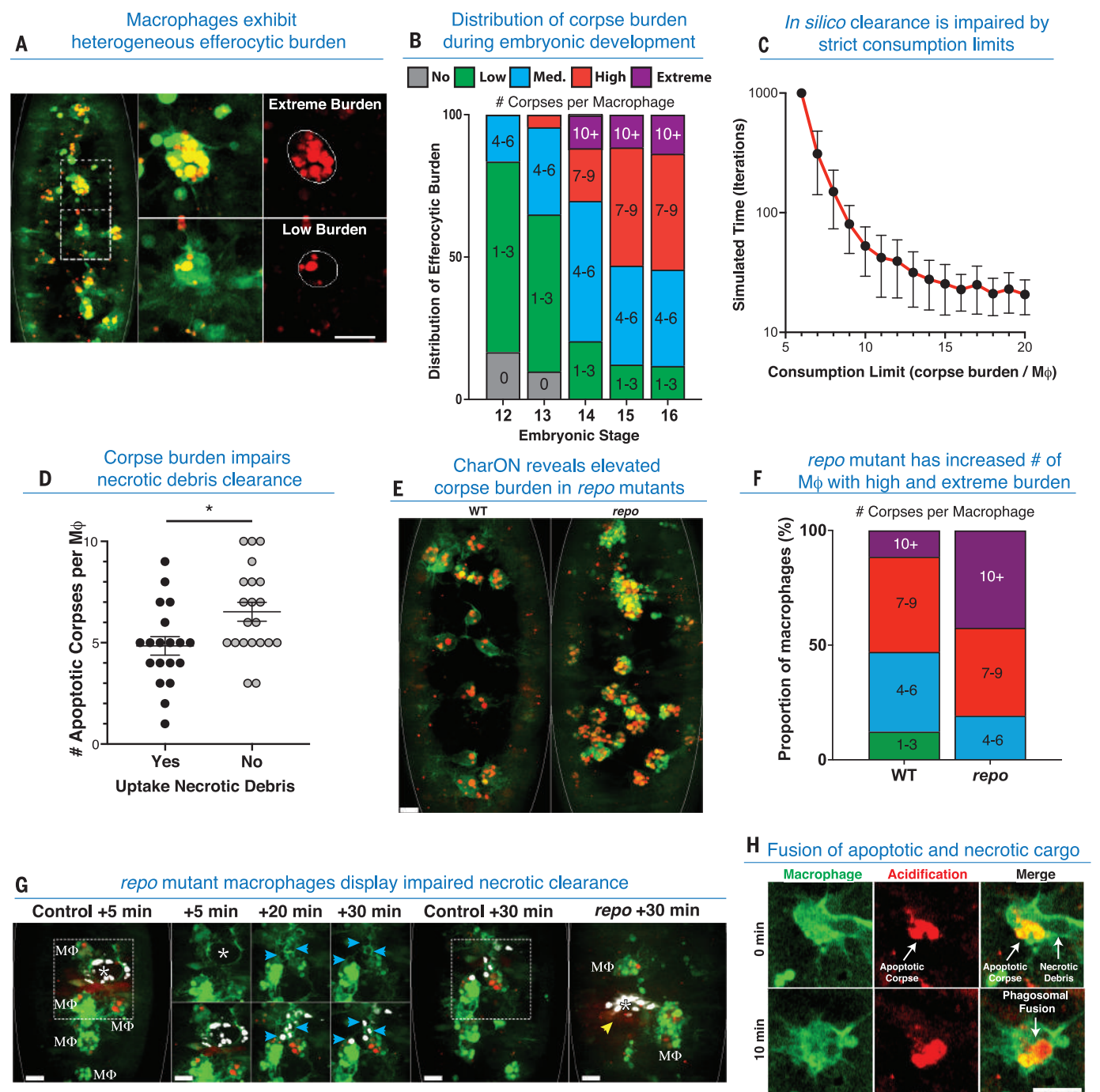These data suggested that when a high concentration of apoptotic cells emerge,

Fig. 4. Macrophage efferocytic heterogeneity. (**A**) Macrophages exhibit variable corpse burden. (Left) A stage-16 embryo with GFP-labeled macrophages (green) and CharON-labeled apoptotic corpses (green and red). Two adjacent macrophages with contrasting corpse burdens are highlighted (dashed boxes), magnified, and outlined (middle and right). (**B**) Percentage of macrophages with indicated corpse burdens during embryogenesis (stages 12 to 16, five embryos per stage). (**C**) In silico clearance time is exponentially increased when equal sharing of corpses is enforced through consumption limits. Simulations were run until clearance was completed or 1000 iterations max and repeated 50 times to yield standard deviations (error bars). (**D**) Macrophages with higher developmental apoptotic corpse burdens are significantly less likely to engulf necrotic debris at wounds (unpaired Student's t test, *$P < 0.0332$, five wounded embryos, error bars = SEM). (**E** and **F**) Macrophages in *repo* mutants have elevated corpse burdens. (E) Wild-type (WT) and *repo* mutant stage-15 embryos (outlined)

expressing CharON (green and red) and macrophage-specific GFP (green). (F) Percentage of macrophages with indicated corpse burdens in wild-type (WT) or *repo* mutant embryos (stage 15, five embryos per genotype). (**G**) *repo* mutant macrophages display impaired necrotic debris clearance. A necrotic stain (DRAQ7; white) was injected into control or *repo* mutant embryos (stage 15) expressing CharON-specific (green and red) and macrophage-specific (MΦ) GFP (green). After laser-wounding (indicated with an asterisk), control macrophages cleared necrotic debris within 30 min (blue arrows). By contrast, *repo* macrophages with extreme corpse burden failed to clear necrotic debris (yellow arrow). (**H**) Fusion of phagosomes containing apoptotic and necrotic cargo. A GFP-labeled macrophage (green) containing CharON-labeled apoptotic corpses (green and red) engulfs necrotic debris at a wound. Acidification of necrotic corpse was detected with pHlorina, occurring rapidly after interaction with an acidified apoptotic corpse. Scale bars, 10 µm.

unrestrained uptake ensures rapid clearance but results in unequal corpse burden among macrophages. We tested this concept using agent-based modeling of macrophage efferocytosis, the parameters for which were derived from our in vivo observations (fig. S8, A and B, and materials and methods). When virtual phagocytes and corpses were randomly distributed, corpse capacity limits exponentially increased clearance time (Fig. 4C, fig. S8C, and movie S7). Conversely, releasing the phagocytes from strict consumption limits led to rapid clearance in silico, at the expense of unequal corpse burden among the macrophages. A similar pattern emerged from simulations in which phagocytes chemotaxed through a field of corpses, in which a strict consumption limit again impaired complete clearance (fig. S8, D and E, and movie S7).

We next explored consequences of varying corpse burden for in vivo macrophage behavior. Tracking of macrophages after clearance (stage 15) demonstrated that corpse burden did not affect their basal motility, despite low and high corpse burdens representing different "physical" loads (fig. S9A). Similarly, macrophages with high corpse burden readily migrated toward laser-induced wounds. There was no significant difference between responding and nonresponding macrophages in terms of prior apoptotic corpse burden (figs. S9, A to D, and S10A and movie S8).

Laser-induced wounds in *Drosophila* embryos are entirely necrotic, as demonstrated by the lack of pH-CaspGFP fluorescence (movie S9). Macrophage-mediated clearance of necrotic debris was observed as fluorescence-negative particle uptake (lacking pH-CaspGFP signal) against the GFP-labeled macrophages. Although recruited at a comparable rate, macrophages with high apoptotic corpse burden exhibited an impaired ability to engulf necrotic debris at the wound, implying a "phagocytic satiety" (Fig. 4D). Although high and extreme corpse burden compromises subsequent inflammatory efferocytosis, the heterogeneity between macrophages in corpse burden leads to sufficient macrophages with low corpse burden to ensure clearance.

In many pathologies, macrophages function in a background of increased cell death and are thus bound to carry a higher corpse burden. To model this, we used the *repo* mutant (which lacks phagocytic glia and their efferocytic contribution), which increased the corpse burden on the macrophages (Fig. 4, E and F, and fig. S10B) (*20–22*). This resulted in more macrophages with high and extreme corpse burdens and a near absence of macrophages with low burden. Next, using DRAQ7 dye to label the laser-induced necrotic corpses, we visualized inflammation in wild-type and *repo* embryos by means of three-color imaging. Although *repo* mutant macrophages have

impaired inflammatory chemotaxis, sufficient macrophages are recruited to analyze efferocytosis (*22*). Wild-type macrophages cleared all DRAQ7-labeled necrotic corpses within 30 to 60 min (Fig. 4G and movie S9). By contrast, overburdened *repo* macrophages struggled to engulf necrotic debris, even when directly contacting the wound (Fig. 4G, fig. S4C, and movie S9).

Last, CharON allowed us to track the fate of apoptotic and necrotic cargo within the same macrophage. Rapid acidification of necrotic corpses sometimes occurred through fusion with phagolysosomes that contained an apoptotic corpse (Fig. 4H and movie S10). This suggested that macrophages may rapidly degrade necrotic debris to alleviate high corpse burden during times of heightened efferocytosis, such as inflammation.

CharON should prove instrumental in dissecting efferocytosis in vivo, as genetically encoded probes provide cell type–specific benefits (*23, 24*). Whether macrophages arrest after they encounter, or ingest, an apoptotic cell has been debated (*18, 19, 25–27*). Using CharON, we tracked developmental apoptosis and efferocytosis within the *Drosophila* embryo, in real time. However, even macrophages that had recently engulfed multiple corpses remained highly motile and phagocytic, implying that engulfment and motility are not mutually exclusive. Furthermore, we propose that macrophages prioritize unrestrained corpse uptake, as opposed to corpse degradation, to ensure the rapid clearance of extensive apoptosis (such as during development). Although this results in variable (including extreme) corpse burdens, this efferocytic strategy should reduce the risk of uncleared apoptotic corpses undergoing secondary necrosis. Furthermore, enforced equal sharing of corpses would presumably require additional regulatory complexity.

In many human inflammatory diseases, there is increased and diverse forms of cell death, which can perturb efferocytosis (*1, 28*). It is unknown whether the same macrophage can engulf both apoptotic and necrotic corpses in vivo. In our model, we observed macrophages filled with apoptotic corpses seamlessly transition to clearing necrotic debris. However, we found that overburdened macrophages were severely impaired in their ability to engulf further necrotic debris. Thus, the efferocytic strategy adopted by macrophages, which maximizes clearance under homeostasis, may be counterproductive during pathology and ultimately exacerbate disease.

REFERENCES AND NOTES

1. S. Morioka, C. Maueröder, K. S. Ravichandran, *Immunity* **50**, 1149–1162 (2019).
2. J. M. Kinchen, K. S. Ravichandran, *Nat. Rev. Mol. Cell Biol.* **9**, 781–795 (2008).
3. L. Fourgeaud *et al.*, *Nature* **532**, 240–244 (2016).
4. N. A-Gonzalez *et al.*, *J. Exp. Med.* **214**, 1281–1296 (2017).
5. P. Ayata *et al.*, *Nat. Neurosci.* **21**, 1049–1060 (2018).
6. L. Galluzzi *et al.*, *Cell Death Differ.* **25**, 486–541 (2018).
7. Q. Zhang *et al.*, *J. Am. Chem. Soc.* **141**, 4526–4530 (2019).
8. T.-L. To *et al.*, *Cell Chem. Biol.* **23**, 875–882 (2016).
9. J. Zhang *et al.*, *Nat. Commun.* **4**, 2157 (2013).
10. T. M. Roberts *et al.*, *Sci. Rep.* **6**, 28166 (2016).
11. J. M. Abrams, K. White, L. I. Fessler, H. Steller, *Development* **117**, 29–43 (1993).
12. A. Rogulja-Ortmann, K. Lüer, J. Seibert, C. Rickert, G. M. Technau, *Development* **134**, 105–116 (2007).
13. A. J. Davidson, W. Wood, *Cold Spring Harb. Perspect. Biol.* **12**, a036350 (2020).
14. E. Kurant, S. Axelrod, D. Leaman, U. Gaul, *Cell* **133**, 498–509 (2008).
15. W. Wood, C. Faria, A. Jacinto, *J. Cell Biol.* **173**, 405–416 (2006).
16. A. J. Davidson, W. Wood, *Cell Rep.* **31**, 107692 (2020).
17. T. A. Markow, S. Beall, L. M. Matzkin, *J. Evol. Biol.* **22**, 430–434 (2009).
18. Y. Wang *et al.*, *Cell* **171**, 331–345.e22 (2017).
19. A. Yurdagul Jr. *et al.*, *Cell Metab.* **31**, 518–533.e10 (2020).
20. G. Trébuchet *et al.*, *J. Neurosci.* **39**, 238–255 (2019).
21. W. C. Xiong, H. Okano, N. H. Patel, J. A. Blendy, C. Montell, *Genes Dev.* **8**, 981–994 (1994).
22. E. L. Armitage, H. G. Roddie, I. R. Evans, *Cell Death Dis.* **11**, 627 (2020).
23. T.-W. Chen *et al.*, *Nature* **499**, 295–300 (2013).
24. A. S. Abdelfattah *et al.*, *Science* **365**, 699–704 (2019).
25. N. A-Gonzalez *et al.*, *Immunity* **31**, 245–258 (2009).
26. H. G. Roddie, E. L. Armitage, J. A. Coates, S. A. Johnston, I. R. Evans, *PLOS Biol.* **17**, e2006741 (2019).
27. D. Park *et al.*, *Nature* **477**, 220–224 (2011).
28. A. C. Doran, A. Yurdagul Jr., I. Tabas, *Nat. Rev. Immunol.* **20**, 254–267 (2020).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4430
Materials and Methods
Figs. S1 to S10
References (*29–39*)
MDAR Reproducibility Checklist
Movies S1 to S10

SURFACE CHEMISTRY

Visualization of on-surface ethylene polymerization through ethylene insertion

Weijun Guo[1,2]†, Junqing Yin[3]†, Zhen Xu[2], Wentao Li[2], Zhantao Peng[2], C. J. Weststrate[4], Xin Yu[1], Yurong He[3], Zhi Cao[1,3], Xiaodong Wen[1,3], Yong Yang[3], Kai Wu[2]*, Yongwang Li[1,3], J. W. Niemantsverdriet[1,4], Xiong Zhou[1,2]*

Polyethylene production through catalytic ethylene polymerization is one of the most common processes in the chemical industry. The popular Cossee-Arlman mechanism hypothesizes that the ethylene be directly inserted into the metal–carbon bond during chain growth, which has been awaiting microscopic and spatiotemporal experimental confirmation. Here, we report an in situ visualization of ethylene polymerization by scanning tunneling microscopy on a carburized iron single-crystal surface. We observed that ethylene polymerization proceeds on a specific triangular iron site at the boundary between two carbide domains. Without an activator, an intermediate, attributed to surface-anchored ethylidene ($CHCH_3$), serves as the chain initiator (self-initiation), which subsequently grows by ethylene insertion. Our finding provides direct experimental evidence of the ethylene polymerization pathway at the molecular level.

Polyethylene (PE) is a vital ingredient for many indispensable materials because of its light weight, inertness, and relatively large mechanical strength (1, 2). In industry, PEs are formed through catalytic polymerization with either Ziegler-Natta or Phillips catalysts (3, 4). The former requires active Ti species precombined with an activator such as methyl aluminoxane, whereas the Phillips counterpart in its simplest form consists of dispersed Cr^{3+} species. According to the Cossee-Arlman mechanism (5) for catalytic polymerization, the polymer is supposed to elongate at one end by ethylene insertion into the metal–C bond. Surface science studies of olefin polymerization by planar models of Ziegler-Natta catalysts (6–8), Cr/SiO_2 catalysts (9), and anchored organometallics (10) have mainly addressed the chemical states of the active species and the morphologies of the polymers in their early stages. Polymerizations of other monomers such as alkyne (11, 12), alkane (13), and halogenate arenes (14–16) have been observed by scanning tunneling microscopy (STM). However, on-surface ethylene polymerization has been missing, leaving the ethylene insertion process elusive at the molecular level.

A number of other elements, such as Ni and Fe, also have catalytic activity for olefin polymerization (17–19). Iron carbide is generally considered as the active phase for polymerization

of CH_x species in Fischer-Tropsch synthesis (FTS) (20–22). Moreover, it has generally been accepted that ethylene can be incorporated into hydrocarbon chains under FTS conditions (23, 24). Such a similarity in the processes implies that an atomically flat iron carbide surface might be suitable for exploration of on-surface ethylene polymerization. In this work, ethylene polymerization was visualized by in situ STM imaging at room temperature (RT-STM) on a partially carburized Fe(110) surface. The active site for the polymerization was revealed at the boundary between two carbide domains (fig. S1). Our experimental measurements indicate that ethylene at these sites rearranges into a surface species and acts as the chain initiator that subsequently triggers polymer growth by ethylene insertion.

The iron carbide surface is reproducibly prepared by C aggregation from the bulk of a Fe(110) single crystal after thermal annealing at 573 K (25). Details of carbide formation can be found in fig. S2. Except for signals of Fe

and C, other frequently observed contaminants such as O, N, and S are not detected in the x-ray photoelectron spectrum (XPS) (fig. S3). Angle-resolve XPS experiments (fig. S4) show that the measured Fe:C ratio decreases with the take-off angle and reaches a stable minimum value of 2.9:1, below the critical angle of 13°. The binding energy (BE) of the Fe $2p_{3/2}$ feature shifts from 707.0 to 707.4 eV, whereas the BE of the C $1s$ peak remains at 282.9 eV, confirming the chemical state of the iron carbide film. The thickness of the grown iron carbide overlayer is estimated to be ~9 Å.

The STM image in Fig. 1A shows the iron carbide surface consisting of parallel domain strips, which contain periodical subunits with an approximate rectangular unit cell of 6.8 Å × 9.7 Å (Fig. 1B). The boundary strip separating the domains has a constant width of 6.0 Å. The atomically resolved STM image in Fig. 1C reveals both Fe and C as protrusions after tip modification. All possible exposed surfaces (6335 in total) of the low Miller index (h, k, l ≤ 5) planes (716 in total) of common iron carbides (FeC, ε-Fe_2C, χ-Fe_5C_2, and θ-Fe_3C) were investigated, and the θ-Fe_3C(102) surface resembles our observed domain structure best (fig. S5). The surfaces share rectangular units, side-by-side arranged hexagonal structures, as well as interval ridges and valleys, although the unit cell of the θ-Fe_3C(102) surface is slightly larger (6.7 Å × 11.1 Å). The simulated STM image (fig. S5G) is consistent with this and indicates that the dark depressions in the STM image shown in Fig. 1B are due to the C atoms, similar to those on nickel carbide in previous reports (26, 27). Carbide formation in both on-surface and subsurface layers of Fe(110) is thermodynamically favorable according to previous density functional theory (DFT) calculations (28, 29). As shown in fig. S6, the Fe(110) surface can be divided into super cells of 7.0 Å × 9.9 Å along the [1̄11] direction, which are very close to the domain unit cells and consistent with the orientation of the carbide stripes.

[1]SynCat@Beijing, Synfuels China Technology Co., Ltd., Beijing 101407, China. [2]Beijing National Laboratory for Molecular Sciences (BNLMS), College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. [3]State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan 030001, China. [4]SynCat@DIFFER, Syngaschem BV, 5600 HH Eindhoven, Netherlands.
*Corresponding author. Email: kaiwu@pku.edu.cn (K.W.), zhouxiong@synfuelschina.com.cn (X.Z.)
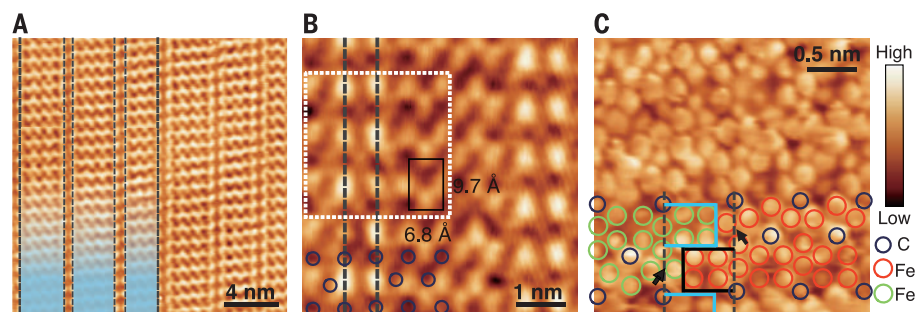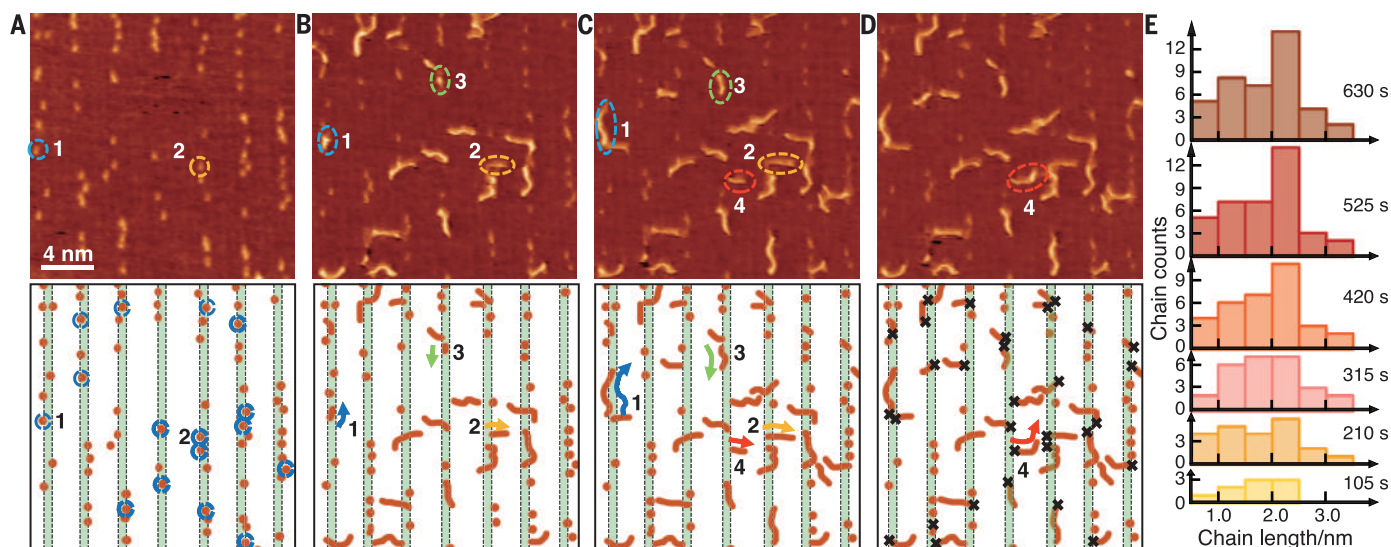†These authors contributed equally to this work.

Fig. 1. LT-STM (10 K) images of the carburized Fe(110) surface. (**A**) Parallel domains of variable widths (light blue) separated by narrow boundary stripes. Imaging conditions were as follows: bias voltage = 40 mV; feedback current = 3 nA. (**B**) The approximate rectangular unit cell within the domain (10 mV, 8 nA). (**C**) Atomically resolved STM image of the white frame in (B) with a superimposed structure model. Blue and black lines indicate the the zipper-like boundary. Black arrows point to the joint positions of neighboring domains (−150 mV, 9 nA).

Fig. 2. STM snapshots of the carburized Fe surface exposed to C_2H_4 at RT. (**A** to **D**) Video frames selected from the STM movie at 35, 245, 385, and 420 s, respectively, along with the corresponding schematic structures (bottom). The circled and numbered areas mark several initial intermediates that are growing into chains (RT, 35 mV, 0.8 nA). Crosses in (D) highlight the roots of the chains at the boundary. C_2H_4 pressure was 1×10^{-8} mbar. (**E**) Statistical length distribution of in situ formed polymer chains.

Fe atoms in two adjacent domains are highlighted by the green and red circles in Fig. 1C. It is apparent that the atomic arrangement of the boundary strip is an intermittent extension of the neighboring domains, which resembles a zipper in its locked state. At the joint endings of the two zipper teeth rows, specific sites appear and interrupt the periodicities of the domains, as highlighted by the arrows in Fig. 1C.

In situ STM images of the carburized Fe(110) surface under 1×10^{-8} mbar C_2H_4 at RT were continuously recorded at a speed of one frame every 35 s (Fig. 2). The bias was adjusted to be as low as 35 mV to avoid possible tip-induced effects. Movie S1 contains STM images showing the instant growth of the oligomers, and Fig. 2 shows four representative frames as a function of the exposure time. The first image, taken during the initial 35 s of reaction, shows that the protrusions are nearly exclusively located at the boundary (Fig. 2A and fig. S7A). Subsequent images (Fig. 2, B to D) illustrate growing chains. The number and length of the chains increase as the reaction proceeds. Figure 2E and fig. S8 depict the length distribution as a function of the reaction time, indicating that the most abundant chain length is 2 to 2.5 nm, corresponding to the C_{16} to C_{20} oligomers. The longest one is ~3.5 nm within the reaction duration. Another in situ experiment (fig. S9) shows a similar length distribution. Some chains grow directly from the protrusions, as shown in movie S1 and marked by the blue dashed cycles in Fig. 2A, implying that these boundary-anchored species are the initiators for the polymerization. For instance, chains 1 and 2 start from the protrusions in
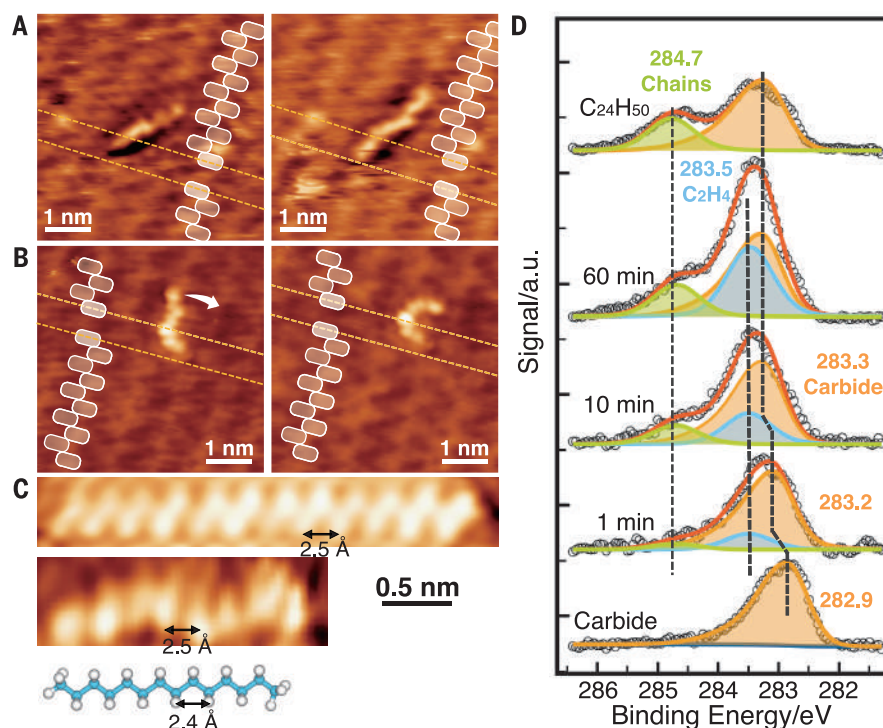

Fig. 3. Characterizations of the polymer chains on the carburized Fe surface. (**A**) Continuous in situ STM monitoring of a growing single chain (RT, 2.5 V, 1.3 nA). (**B**) STM tip manipulation of a grown chain (RT, 1.7 V, 0.3 nA). (**C**) STM images of a single $C_{24}H_{50}$ chain (top) (10 K, −150 mV, 0.5 nA) and a representatively grown chain (RT, 30 mV, 1 nA) (bottom) out of C_2H_4 at RT. (**D**) C 1s XPS spectra of the carburized Fe(110) during its exposure to 1×10^{-8} mbar C_2H_4 at RT.

Fig. 2A, develop into short chains in Fig. 2B, and eventually grow into longer ones in Fig. 2C. Continuous STM monitoring of chains 1 to 4 in Fig. 2 and a growing single chain from 1.8 to 2.1 nm in Fig. 3A confirms that one end of the chain is immobile and remains anchored at the boundary. Once the polymerization is initiated, its subsequent chain growth becomes quite fast, which echoes the characteristics of the chain reaction polymerization mechanism.

Fig. 4. LT-STM images of the carburized Fe(110) after exposure to ethylene at RT.
(**A**) At 0.12 L, C_2H_4 molecules preferentially adsorb at the boundary. (**B**) At 0.6 L, triangular species appear. Yellow and blue cycles highlight species at the boundary and domain, respectively. (**C**) Zoom-in STM image of the white dashed box in (B) with superimposed structure models. (**D**) Boundary geometric model (left) and its corresponding simulated STM image (right). The unit cell (white rectangle) is 1.10 nm × 2.69 nm. (**E** and **F**) Enlarged STM images of an individual C_2H_4 molecule (E) and ethylidene ($CHCH_3$) (F), along with their calculated structure models (middle) and simulated STM images (right). Imaging size: 0.75 × 0.75 nm². (**G**) LT-STM image of grown short PE chains. (**H**) Schematic illustrations of the polymer chain growth at the boundary from C_2 to C_{10} and the experimental STM images of

C_8 and C_{10}. Imaging conditions were as follows: 10 K, −30 mV, 1.4 nA [except for (A), for which conditions were 10 K, 100 mV, 1 nA]. (**I**) Geometric models and calculated adsorption energy differences (ΔE) of upright and horizontal C8 and C10 chains on the iron carbide overlayer.
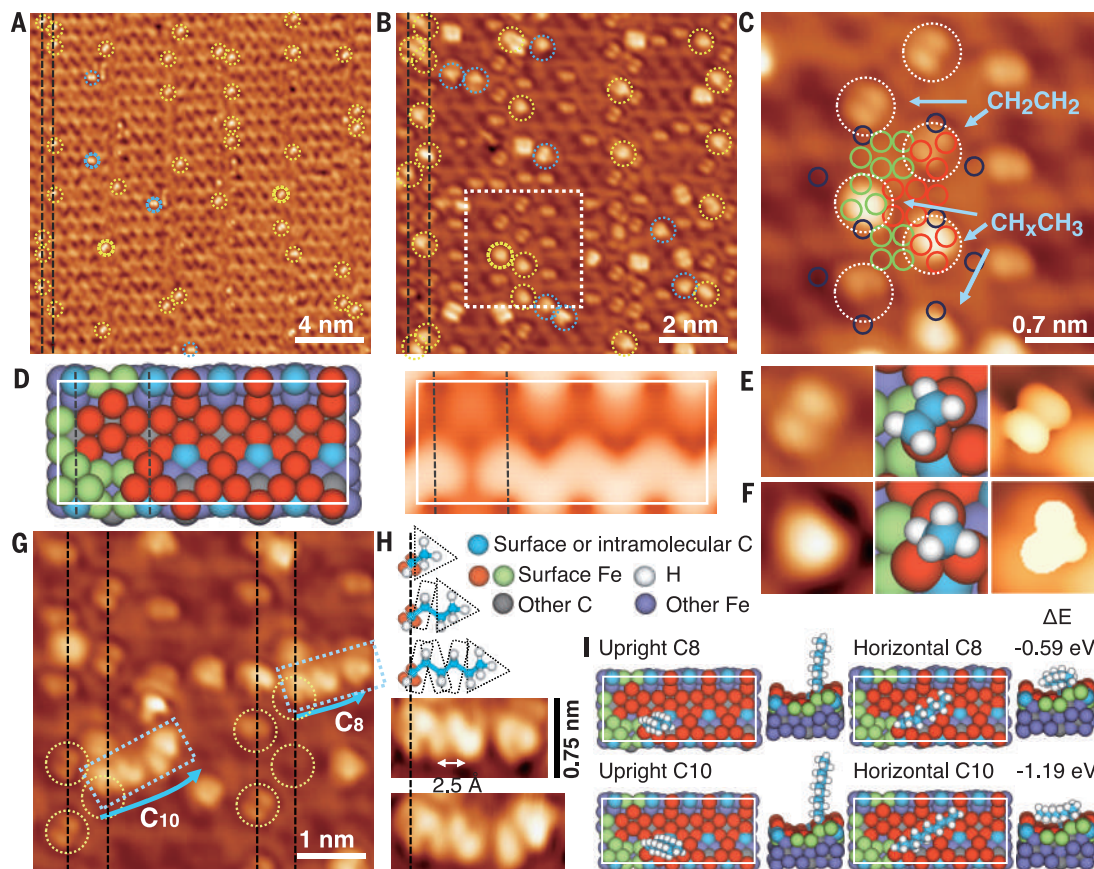
The imaged area contains 33 chains in total after 420 s of reaction (Fig. 2D), during which 16 chains grow directly at the protrusion locations in Fig. 2A. In addition, most chains (31 of 33) exist with one end anchored at the boundary, reinforcing the idea that the polymerization indeed takes place at the boundary sites. Furthermore, the movable ending of a single chain is bendable by tip manipulation (Fig. 3B), but the ending anchored at the boundary is unbendable. The high-resolution STM image in fig. S10 further confirms that the chains do originate from boundary sites, which excludes the possible participation of the C atoms from the carbide surface.

It needs to be clarified that low-coverage adsorbed ethylene molecules could not be imaged by RT-STM, presumably because of their high mobility on the surface. However, under a higher pressure such as 10^{-7} mbar, the ethylene molecules cover the whole surface quickly and prevent them from moving (fig. S11). In this way, the randomly crowded ethylene molecules become visible. At ~100°C, ethylene dehydrogenation kicks off (fig. S12). The generated chains adopt an armchair configuration, different from the zigzag chains grown at RT. Ex situ STM imaging of the sample upon exposure

to ethylene at RT reveals a similar chain length distribution (fig. S13), further excluding the tip effect. The polymers cannot grow very long because the chain propagation needs more energy to make the whole chain creep on the surface (fig. S13C). In a reference experiment, no chains grew on the carburized Fe(100) single crystal (fig. S14).

For comparison, the tetracosane ($C_{24}H_{50}$) molecules are thermally deposited onto the iron carbide surface. Because they form a rather strong van der Waals interaction with the substrate, the tetracosane molecules can be clearly imaged by both RT-STM and low-temperature STM (LT-STM) (fig. S15). Their appearance (Fig. 3C) in STM closely resembles those of the on-surface–grown polymer chains. The experimentally measured periodicity, 2.5 Å, for both the polymerized ethylene and tetracosane molecule corresponds to the interval spacing for each additional –CH_2– segment and is very close to the spacing distance of 2.4 Å in free alkanes. The C 1s feature in XPS (Fig. 3D) of the iron carbide exposed to the C_2H_4 gas at RT provides spectroscopic evidence for PE formation. To analyze the XPS data, the contribution from the carbide C 1s is deconvoluted and subtracted by fitting the XPS peaks with

the same asymmetric line type and intensity (orange). The carbide C 1s feature shifts upward in BE from 282.9 to 283.3 eV because of the so-called adsorbate-induced core-level shift (30, 31) caused by adsorbed C_2H_4 molecules. After being exposed to C_2H_4 for 1, 10, and 60 min, an additional common feature gradually develops at 283.5 eV (blue), which we attribute to the adsorbed C_2H_4 molecules (31, 32). At exposures to C_2H_4 for 10 and 60 min, the intensified feature at 284.7 eV (green) clearly stems from PE. Moreover, the C 1s of the adsorbed $C_{24}H_{50}$ positions at 284.7 eV is very similar to that of polymerized ethylene on surface.

To explore the initiation of the polymer growth, LT-STM imaging at 10 K and DFT calculations are used. Mobile adsorbed C_2H_4 molecules can be clearly imaged by LT-STM, as shown in Fig. 4A, at an exposure of 0.12 Langmuir (L) (2×10^{-9} mbar for 60 s) of ethylene at RT. Both CH_2 groups in C_2H_4 are clearly distinguishable as a pair of protrusions. About 90% of the adsorbed C_2H_4 molecules are located at the boundary and adopt the same orientation with respect to the underlying substrate lattice. Most boundary sites become occupied as the C_2H_4 exposure increases to 0.6 L (1×10^{-8} mbar for 60 s) (Fig. 4B). A few

rectangular species are captured, which are tentatively attributed to paired C_2H_4 molecules in parallel arrangement. It is noteworthy that a new type of adsorbate in approximate triangular shape dominates at the boundary, accounting for ~75% of the adsorbates. These species have an apparent height of 115 pm, substantially higher than that of the initially adsorbed C_2H_4 molecules (65 pm). These intermediates are also visible with RT-STM (Fig. 2A), meaning that they bind to the substrate more strongly than do C_2H_4 molecules. The triangular protrusion in STM suggests a terminal CH_3 group. Possible species include ethylidene ($CHCH_3$), ethylidyne (CCH_3), and ethyl (CH_2CH_3). Species such as CCH_3 and $CHCH_3$ are frequently formed upon exposure of transition metal surfaces to C_2H_4 (31–34). It is therfore proposed that the triangular species are $CHCH_3$ formed by isomerization of C_2H_4. Figure 4C shows that the C_2H_4 and $CHCH_3$ species reside on a triangle consisting of three Fe atoms at the boundary between two carbide domains.

The boundary model used in our calculation was constructed from θ-Fe_3C (102) with a 1×4 supercell (Fig. 4D and fig. S16) by simplifying the central symmetric boundary structure into mirror symmetric. The theoretically optimized adsorbate geometries and simulated STM images of adsorbed C_2H_4 and $CHCH_3$ shown in Fig. 4, E and F, and fig. S17 are in good agreement with the experimental results; e.g., the calculated center-to-center distance of CH_2 groups in C_2H_4 (2.3 Å) and the triangular shape of simulated $CHCH_3$ image are consistent. DFT calculations of the adsorbed C_2H_4 and $CHCH_3$ also indicate an adsorption preference at the triangular site. These triangular sites at the boundary are not just the preferred ones for ethylene adsorption but also promote the transformation of ethylene into the proposed ethylidene, which acts as the chain initiator for the ethylene polymerization. The activity of the boundary site (35, 36) stems from the Fe undercoordination to C because of the abrupt breaking in their periodicity at this ensemble site, which is both supported by the theoretically calculated charge distribution and the experimentally measured local work function mapping (fig. S18).

At higher ethylene exposures, it became increasingly difficult to distinguish all surface species in the LT-STM images (fig. S19), and some relatively short chains could be clearly identified. Figure 4, G and H, shows the C_8 and C_{10} chains rooted from the boundary, confirming that they initiate there. These short chains retain the triangularly shaped feature at their ends farther away from the boundary. Such a triangular ending is the terminal CH_3 group, which is inactive for further reaction with any incoming C_2H_4 molecules. The DFT calculations for the C4/C6 and C8/C10 chains on the iron carbide surface are given in fig. S20 and Fig. 4I, respectively, and show stronger adsorption for the horizontal ones against their upright counterparts and are consistent with the experimental results. All of our results fit the scenario that a C_2H_4 molecule is inserted into the initial $CHCH_3$ intermediate at the triangular site to form a chain. This explains PE formation through ethylene self-initiation over the Phillips catalyst without the necessity of an activator.

In summary, this study reports in situ STM visualization of the activator-free polymerization of ethylene on a partially carburized Fe surface. The ethylene molecules adsorb on the carbide surface and rearrange into an intermediate, presumably ethylidene, on the specific triangular ensemble site located at the boundary of two neighboring carbide domains. The polymerization initiates on the boundary ensemble sites and propagates by insertion of the activated ethylene molecule in the neighborhood into the Fe–CH–R species. During each run of the ethylene insertion, hydrogen is supposed to transfer among the formed carbon chains to retain the metal–CH–R configuration so that the PE chain keeps growing.

REFERENCES AND NOTES

1. P. D. Hustad, *Science* **325**, 704–707 (2009).
2. J. M. Eagan et al., *Science* **355**, 814–816 (2017).
3. M. P. McDaniel, in *Advances in Catalysis*, B. C. Gates, H. Knözinger, Eds. (Academic, 2010), vol. 53, pp. 123–606.
4. V. C. Gibson, S. K. Spitzmesser, *Chem. Rev.* **103**, 283–315 (2003).
5. E. J. Arlman, P. Cossee, *J. Catal.* **3**, 99–104 (1964).
6. E. Magni, G. A. Somorjai, *Surf. Sci.* **345**, 1–16 (1996).
7. S. H. Kim, G. A. Somorjai, *J. Phys. Chem. B* **105**, 3922–3927 (2001).
8. A. Andoni, J. C. Chadwick, S. Milani, J. W. Niemantsverdriet, P. C. Thüne, *J. Catal.* **247**, 129–136 (2007).
9. P. C. Thüne, J. Loos, P. J. Lemstra, J. W. Niemantsverdriet, *J. Catal.* **183**, 1–5 (1999).
10. W. Han, C. Müller, D. Vogt, J. W. Niemantsverdriet, P. C. Thüne, *Macromol. Rapid Commun.* **27**, 279–283 (2006).
11. S. Wang et al., *Nat. Chem.* **11**, 924–930 (2019).
12. Y. Okawa, M. Aono, *Nature* **409**, 683–684 (2001).
13. D. Zhong et al., *Science* **334**, 213–216 (2011).
14. L. Lafferentz et al., *Science* **323**, 1193–1197 (2009).
15. J. Cai et al., *Nature* **466**, 470–473 (2010).
16. D. P. Goronzy et al., *ACS Nano* **12**, 7445–7481 (2018).
17. H. Makio, H. Terao, A. Iwashita, T. Fujita, *Chem. Rev.* **111**, 2363–2449 (2011).
18. M. W. Bouwkamp, E. Lobkovsky, P. J. Chirik, *J. Am. Chem. Soc.* **127**, 9660–9661 (2005).
19. O. L. Sydora, *Organometallics* **38**, 997–1010 (2019).
20. J. van de Loosdrecht et al., in *Comprehensive Inorganic Chemistry II: From Elements to Applications*, J. Reedijk, K. Poeppelmeier, Eds. (Elsevier, 2013), pp. 525–557.
21. E. van Steen, M. Claeys, *Chem. Eng. Technol.* **31**, 655–666 (2008).
22. X. Zhou et al., *ACS Catal.* **8**, 7326–7333 (2018).
23. Y. Zhang et al., *Eng. Rep.* **2**, e12232 (2020).
24. J. Yang et al., *Appl. Catal. A Gen.* **598**, 117564 (2020).
25. Materials and methods are available as supplementary materials.
26. C. Klink et al., *Phys. Rev. Lett.* **71**, 4350–4353 (1993).
27. R. T. Vang et al., *Surf. Sci.* **600**, 66–77 (2006).
28. R. Gao et al., *Catal. Lett.* **149**, 645–664 (2019).
29. X. W. Liu, C. F. Huo, Y. W. Li, J. Wang, H. Jiao, *Surf. Sci.* **606**, 733–739 (2012).
30. W. F. Egelhoff Jr., *Surf. Sci. Rep.* **6**, 253–415 (1987).
31. C. J. Weststrate, I. M. Ciobîčă, J. van de Loosdrecht, J. W. Niemantsverdriet, *J. Phys. Chem. C* **120**, 29210–29224 (2016).
32. S. Lizzit, A. Baraldi, *Catal. Today* **154**, 68–74 (2010).
33. G. A. Somorjai, A. M. Contreras, M. Montano, R. M. Rioux, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10577–10583 (2006).
34. H. Kirsch, Y. Tong, R. K. Campen, *ChemCatChem* **8**, 728–735 (2016).
35. J. H. K. Pfisterer, Y. Liang, O. Schneider, A. S. Bandarenka, *Nature* **549**, 74–77 (2017).
36. R. T. Vang et al., *Nat. Mater.* **4**, 160–162 (2005).

SIMONS FOUNDATION

The Simons Foundation is pleased to announce the appointment of our new class of Junior Fellows of the Simons Society of Fellows, a community of scholars created toencourage intellectual interactions across disciplines and research centers in New York City.

Daniel Alabi
Columbia University
Computer Science

Morris Ang
Columbia University
Probability and Mathematical Physics, Conformal Field Theory and Liouville Quantum Gravity

Eric Arsenault
Columbia University
Chemistry

Beatrice Barra
New York University Langone Health
Sensory and Motor Systems Neuroscience

Michael Chapman
New York University
Mathematics, Group Theory and Quantum Information Theory

Sophie Huiberts
Columbia University
Theoretical Computer Science, Optimization

Megan Kirchgessner
New York University
Molecular Systems Neuroscience

Lachlan Lancaster
Columbia University
Astrophysics

Oliver Philcox
Columbia University
Astrophysics

Cynthia Steinhardt
Columbia University
Biomedical Engineering, Neuroscience, and Neuroengineering

Andre Toussaint
Columbia University
Neuroscience

Lynn Yap
Columbia University
Molecular, Cellular, and Systems Neuroscience

For more information:
🐦 @Simons_SOF • https://www.simonsfoundation.org/simons-society-of-fellows

The support I needed

Eleven months into my postdoc, my 4-year-old daughter was diagnosed with cancer. After hearing the news, I stepped outside her hospital room to make two phone calls: one to my mom and the other to my adviser. He didn't answer—it was Easter Sunday—so I left him a voicemail message. "Emery has leukemia. I'm not coming back to the lab. Please give someone else my project and clear my bench." Later, he told me he supported me setting everything aside. But he also never gave up hope that I would rejoin the lab. Unbeknownst to me at the time, my adviser did everything possible to keep things afloat for my return. I'm so grateful he did.

I had joined his lab to work on a project that was a perfect fit for me. I had worked as a veterinarian before completing my Ph.D., and the research—a cancer project that could benefit dogs—leveraged both sets of experience. I worked hard that first year to gather preliminary data. By 11 months in, we had some exciting findings and it was becoming clear that the project had real potential. Then everything came to a screeching halt.

In the months that followed, days came and went and my family and I spent an inordinate amount of time at the hospital. My daughter lost her hair and became extraordinarily weak and thin. I sometimes checked her breathing at night because I was worried it had stopped. In the midst of the stress, I never thought about work.

My adviser stayed in touch with supportive messages. He only called twice about work. The first time was to ensure I had received paperwork I needed to complete to qualify for family medical leave. The papers had been lying on my kitchen counter for days and I could not bring myself to fill them out. I just wanted to quit. "What is the easiest exit strategy for me?" I asked him. "Can't you just say I stopped showing up for work?" He told me that no, he couldn't. "Can't you just fire me then?" He didn't want to do that either. He gently persisted until he convinced me to fill out the paperwork.

The second time he called was to discuss a fellowship I had been awarded. It was an extraordinary honor, and one I had been working so hard for until a couple months earlier. But by that point I didn't care.

Nine months after the initial diagnosis, however, my mindset started to change. My daughter, who was entering the final phase of her chemotherapy, started to regain her strength and longed to return to school. I missed science as well. But the thought of returning to work left me feeling guilty and scared. Had I missed some symptoms when I was busy rushing around as both a scientist and a mom? What if my daughter needed me when I was at work? What if the cancer came back?

My adviser suggested I start working a few hours per week on my project—which, I was surprised to learn, he'd never given to another person. That allowed me to come into the lab on my own time. And as I slowly re-entered my lab life, I discovered he had been laying the groundwork to ease my return. He had arranged to defer my fellowship, which now provided me with funding as well as unparalleled collaborative and networking opportunities.

I was glad to get back to my project. But I wasn't the same scientist as before. I prioritized family time to a greater degree, limiting the amount of work I brought home at night. I also put up some walls at work because I was surrounded by cancer researchers. I avoided seminars about pediatric and blood cancers. When those topics came up unexpectedly in other talks, I would sometimes sit in the audience crying quietly—or I would leave the lecture room entirely.

I am now in the fourth year of my postdoc, and I am thankful every day for my adviser's support. He gave me the space I needed when I was dealing with a personal crisis. He also cared about my career when I didn't give it a second thought. For any professors out there who are juggling projects, grants, and deadlines, please remember to take the time to support your lab members through difficult periods because life—like science—can be messy. ■

> **"My adviser did everything possible to keep things afloat for my return."**

Morgan Schrock is a postdoc at Ohio State University, Columbus. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE
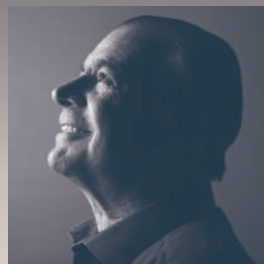
Since 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science,* so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

cut here ✂

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:
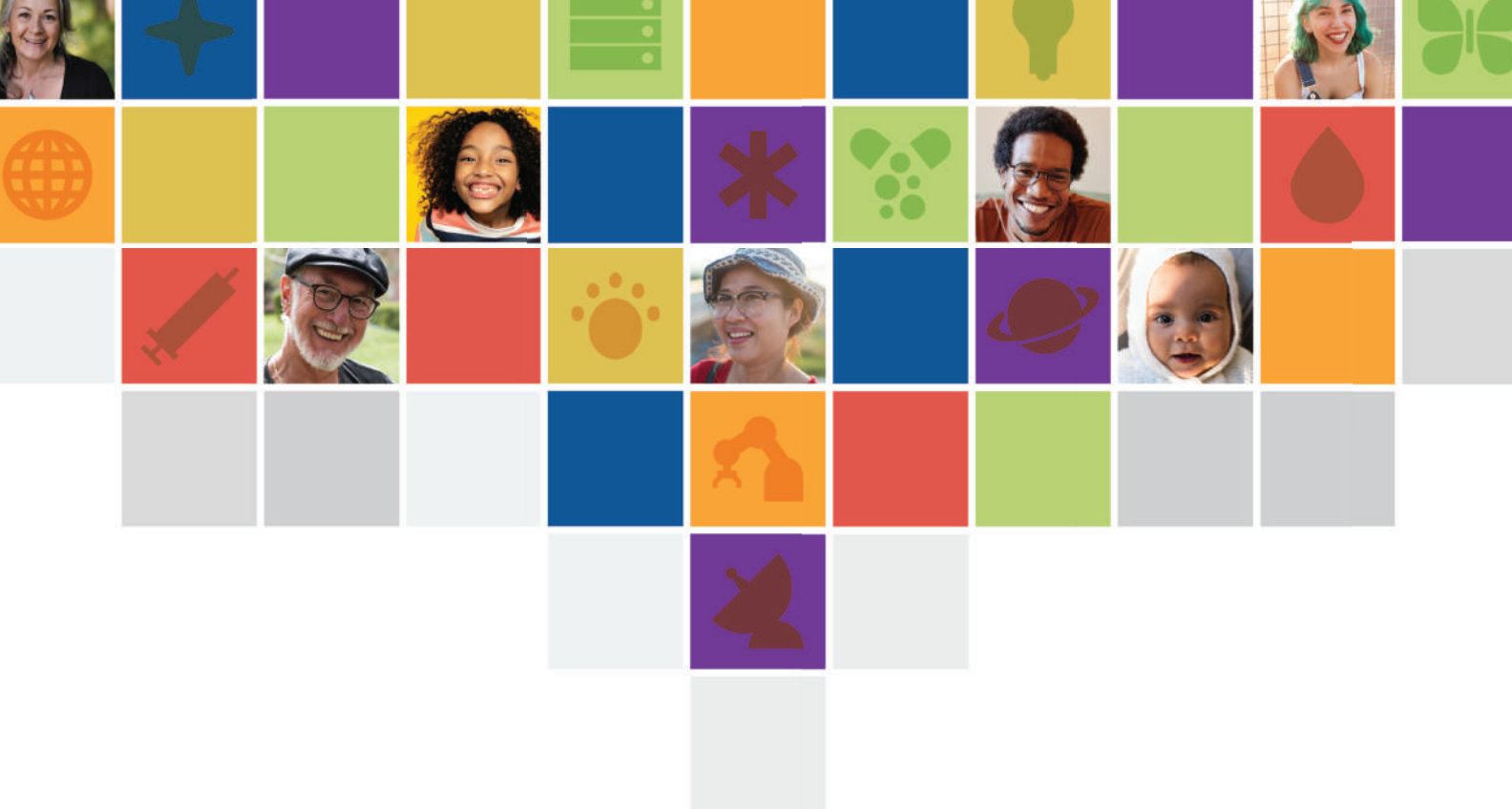
Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

SCIENCE FOR HUMANITY

AAAS | ANNUAL MEETING

The scientific endeavor aims to provide findings, models, tools, and advances to better understand and address the myriad complex challenges facing society. Meeting this goal requires collective objectivity determined by a wealth of individual and diverse perspectives and experiences—a sample as broad as the diversity of disciplines in the scientific endeavor itself.

Highlighting the importance and just integration of our multiplicity, the 2023 AAAS Annual Meeting will feature groundbreaking multi-disciplinary research—research that advances knowledge and responds to the needs of humanity. Drawing from work ranging from astronomy to zoology, the program committee seeks proposals that highlight breakthroughs in science and technology and, in particular, those that incorporate the importance of diversity—in its investigators, subjects of study, and translational implications.

aaas.org/meetings | #AAASmtg