A bid to commercialize a rare natural blue—ethically p. 1100

Beryllium bonding in a crystalline compound pp. 1106 & 1147

Changing transcription over the life span of *Drosophila* p. 1145

Science

$15
16 JUNE 2023
SPECIAL ISSUE
science.org
AAAS

LIGHT POLLUTION

LOSING THE DARKNESS

By **Keith T. Smith**, **Bianca Lopez**, **Sacha Vignieri**, *and* **Brad Wible**

For most of history, the only lights made by humans were naked flames. Daily life was governed by the times of sunrise and sunset, outdoor nighttime activities depended on the phase of the Moon, and viewing the stars was a common and culturally important activity. Today, the widespread deployment of outdoor electric lighting means that the night is no longer dark for most people—few can see the Milky Way from their homes. Outdoor lighting has many legitimate uses that have benefited society. However, it often leads to illumination at times and locations that are unnecessary, excessive, intrusive, or harmful: light pollution.

This special issue examines the effects of light pollution on the natural world, human health, and the night sky. It discusses how the level of light pollution can be measured and what could be done about it.

Both the amount of light pollution and its geographic extent are increasing rapidly, causing worsening impacts on the environment. This wasted light emission consumes vast amounts of electricity, with associated financial costs and greenhouse gas emissions. Although streetlights are the most obvious form of outdoor lighting, light pollution often arises from buildings, vehicles, advertising, sports facilities, and many other sources.

Fortunately, light pollution does not accumulate in the environment; it can be halted by simply turning off the lights, although that is not always practical. Often, those responsible for poor lighting do not realize that it is causing pollution that harms the environment. Careful design, appropriate use of technology, and effective regulation can ensure that we retain the benefits of artificial light at night while minimizing its harmful effects. If we fail to do so, we will lose what little darkness remains.

10.1126/science.adi4552

POLICY FORUM

REVIEWS

RELATED ITEM PODCAST

PHOTO: JIM RICHARDSON

The city of Chicago, seen from above through scattered clouds with the shoreline of Lake Michigan at the bottom. Extensive light pollution is visible from sources including streetlights, commercial buildings, vehicles, and leisure facilities. The light escaping into the sky performs no useful function but is bright enough to reveal the layout of the city and penetrate through the clouds.

A national law regulating artificial light affects m
Koreans, particularly in cities such as Seoul,
where local authorities have power to impose fines.

POLICY FORUM

Regulating light pollution: More than just the night sky

Impacts on ecology, health, energy, and climate are critical

By **Martin Morgan-Taylor**

Artificial light at night (ALAN), or light pollution, causes a broad spectrum of problems. Emerging scientific research describes energy waste exacerbating the climate emergency, and harm to human (*1*) and ecological health (*2*). Despite these broad effects, the problems of light at night are usually couched in terms of the loss of the night sky in the media, leading the public to believe that light pollution is an issue primarily concerning astronomy and star gazing, rather than being more immediately relevant to the average person through broader adverse effects. This public perception may be contributing to the relatively modest priority thus far placed on addressing ALAN in many jurisdictions. Yet all of these issues warrant regulation, which is evolving in response to research findings, across the world. We examine below the types of regulation in existence, and their merits, then offer suggestions to aid the call for further regulation.

Leicester De Montfort Law School, De Montfort University, Leicester, UK. Email: mart@dmu.ac.uk

The types of regulation that exist today vary depending on the source of the law, and how it goes about regulation. Sources vary from "hard law"—binding law from supranational, national government, or regional bodies—to "soft law," which may take the form of non–legally binding guidance. Laws explicitly dedicated to light pollution may be drafted, or ALAN-oriented elements may be "bolted on" to existing legislation that is not ALAN specific. Regulation may act proactively to stop problems in the first place—by, for example, preventing the sale of or use of certain types of lighting—or limit light outputs or spill. Or it may act reactively to lighting problems once they exist, such as light shining into windows. It is possible to combine a hard law core with soft law supporting guidance, involving a mix of the above. Notably, any regulation requires enforcement, which in turn requires the resources for and the buy-in of the bodies tasked with enforcement.

Policy-makers must view regulation as worthwhile, and this means that voters must see the value. The burden imposed by regulation must be set against the negative effects of ALAN. So, regulation must be based on a full understanding of how the benefits and disadvantages of light at night might be effectively balanced. It must cover the spectrum of problems caused by light at night and be enforced in practice. And it must be as simple, easy to understand, and as cheap as possible to enforce. Underpinning this is a need for a proper understanding of the problems and practical solutions, which may be aided with scientific evidence. A general understanding is required of the wider problems that ALAN may cause us all, not just the loss of the night sky. Perceptions of need and economics, in the eyes of the public, policy-makers, and enforcement bodies, are all critical.

EXAMPLE REGIMES

Supranationally, the United Nations Environment Programme has recently published draft light pollution guidance for birds and bats (*3*). Also, a growing number of nations are adopting national laws, with France and the Republic of Korea serving as excellent examples of jurisdictions that have adopted hard law approaches. Some US states have hard law, and other jurisdictions, such as those in the UK, have adopted "bolt on" approaches. We will examine initiatives in the European Union (EU), France, Korea, and the UK as examples.

European Union

Light pollution was on the agenda of the Czech Republic Presidency of the EU in 2022. An outline of European initiatives was commissioned (*4*), listing 18 out of 32 European nations as having some form of national legislation, and the emerging underlying physical, ecological, social, and health scientific research that states the case for regulation. Light pollution is also beginning to be addressed by the EU block itself; the EU's Green Public Procurement Criteria for Street Lighting and Traffic Signals provides street-lighting recommendations, which includes addressing light pollution (*5*), for policy-makers and lighting professionals. This report recommends dimming and selective switching off, subject to the EN13201 lighting standard. Primarily designed to save energy, these standards also serve to reduce all other negative effects of ALAN. The Zero Pollution Action Plan of the European Green Deal names light pollution as an emerging pollutant for monitoring (*6*). Similarly, light pollution is named as a cause of pollinator decline in the EU Pollinators Initiative (*7*). Although such recognition is good, progress is slow in actually reducing light pollution. Continuing scientific research of these fields will support calls for further regulation.

France

France adopted a decree in 2018 that covers the broad spectrum of ALAN problems, including disturbing humans, fauna, and ecology, as well as cutting energy waste and protecting the night sky. Framed in terms of energy waste and the environment, this decree can appeal to businesses and the public as a response to financial, energy, and climate challenges. The decree augments France's existing environmental code, as well as commitments to buildings and town planning, and broader commitments to the Paris agreement and the EU Green Deal. It also defines and gives extra protection to 11 astronomically important areas. Aiming to be preventive, the decree provides a scheme of design and use, for all lighting, but does not cover the sale of lighting. Unlike its 2013 precursor, it applies to private consumer lighting as well as that from the commercial and public sectors. It is built on a review and understanding of the underlying scientific research into the broad problems, provided by the French astronomical community.

The decree includes curfews on external and internal lighting, which are easy to understand and to enforce; limiting blue-rich light at night (implicated in harm to human and ecological health) to a color temperature of 2400 to 3000 K; and banning laser lights that exceed 100,000 lumens in many areas. The decree also limits upward light into the night sky and cutting glare, which will reduce astronomical light pollution and intrusive light, and limits "excessive intrusive light into dwellings" from all lights, including domestic "security" lights. But "excessive" is not defined, leaving tension over the right to self-determination.

Republic of Korea

The Republic of Korea has also adopted a dedicated law (*8*), mainly to cut energy waste and nuisance lighting, especially from billboards. It uses metrics based on the International Commission on Illumination's international standard (CIE 150). These metrics are clear and transparent, and because they set maximum lighting levels, provide less scope for subjective judgments about what is reasonable. This is good for lighting designers and commercial users as it helps minimize enforcement conflict that might arise as a result of uncertainty and ambiguity. The transparency means that consumers have access to metrics to determine whether they have a case, rather than risk financial resources on uncertain legal action. However, metrics require the use of light meters, which provides a cost and training burden on regulators, and this law does not prevent the sale of certain lighting types. As such this is not a fully preventive law.

The metrics classify areas into four environmental zones (E zones) ranging from E1 (the darkest) to E4 (the brightest, city areas). This law is likely to be seen as directly relevant to many city-dwelling Koreans, such as those facing illuminated billboards shining into their windows. A disadvantage of a zone-based policy is that like noise, light travels across different zones. This can be a particular problem in densely populated countries, such as many in Europe and Korea. Similarly, cities include mixed-use areas containing both commercial and residential buildings. These areas often have illuminated billboards. The Korean law uses the same maxima recommended under CIE 150 (*9*). However, it is possible that the law could in time also adopt the curfews used by the French law, so that illuminated billboards and other lights must be switched off. Such a law would combine the best elements of the curfew and metrics system.

Effective regulation needs enforcement powers with penalties for violators. Under the Korean law, local authorities have enforcement powers to fine and/or impose a compliance order. Fines rise with the severity of the breach, and the number of offences, ranging from roughly USD$50 to roughly USD$1000. Increasing fines is an effective way to ensure compliance. It remains to be seen whether local authorities will allocate resources to the lowest levels of breach.

United Kingdom

The UK is not a single jurisdiction, but rather consists of several "home countries" with autonomy in specified areas, including the environment (devolution). The law in England will be discussed here, as it has the highest population (*10*). All UK jurisdictions have adopted the bolt-on approach, rather than dedicated laws, because it was the quickest, cheapest method, and possibly because it was not deemed to warrant the time and cost of a dedicated law. Consideration of light pollution has been added to the planning stage requirements for new buildings, and for the nuisance effects under statutory nuisance laws. These methods offered the best fit to existing laws. There has been a Royal Commission in 2009 (*11*), and there is a current all-party parliamentary group (APPG for Dark Skies) (*12*), with both recommending regulatory changes. These were in response to representations from the astronomical community, including star counts showing the impact on the night sky, and so do not reflect the full spectrum of the problems with ALAN, which might explain the lack of a dedicated law.

Light pollution is now a factor for consideration when applying for planning permission for land development under the National Planning Policy Framework, to "limit the impact of light pollution from artificial light on local amenity, intrinsically dark landscapes and nature conservation." However, this framework tends to highlight dark landscapes and nature conservation, which may conflict with the aims of the lighting to address safety and security, commercial interests, and so on. So this measure is not especially helpful in tackling the broader problems of light at night.

This bolt-on approach has weaknesses, as it tries to incorporate light pollution into an existing framework for something else, which may reduce effectiveness. For example, planning law only covers "development," which is limited to lighting in existing buildings that affects daytime visual appearance. This is because the planning system is interested in the visual appearance of the light fittings, rather than the problems they might cause.

Supplementary guidance from the lighting industry and other environmental interest groups can also play a substantial role in bolt-on regulation. For example, the UK's Institution of Lighting Professionals regularly updates its Guidance Notes for the Reduction of Obtrusive Light (GN01, 2021), with clear guidance on avoiding overlighting, and provides diagrams on how to design and install light fittings correctly. This guidance is well respected in the UK.

ALAN was added to the list of possible statutory nuisances in 2006. These are criminal offences to protect society generally, but it requires an act or omission that is either "prejudicial to health" or a "nuisance," under the Environmental Protection Act 1990. Scientific evidence clearly indicates that harming sleep harms human health, and the emerging research indicates that exposure to light at night might harm sleeping. So the health element might cover loss of sleep caused by light shining into bedroom windows. However, the most fundamental problem with bolting light onto existing nuisance legislation is that light pollution causes problems far wider than the narrow scope of the nuisance regime's second (nuisance) element. This only covers interference relating to everyday use of land, not the broader environmental ecological or night sky concerns that ALAN may cause. So, nuisance has had little effect in combating light pollution.

A WAY FORWARD

Light pollution is an issue that can elicit a response in favor of light at night that is just as emotional as the response to the problems that it can cause. This may to some degree be because a lot of the emotion derives from myths and misconceptions, or rests on beliefs not necessarily borne out by the evidence. These issues may center on an emotional fear

of the dark and a belief that light offers safety and security. In this view, more light is always better, possibly explaining instances in which residents object to the dimming or switching-off of some streetlights. Moreover, it might be that light is regarded as an intrinsic part of liberty, where one is entitled to string up as many security lights as desired. The perception that light is good has led to the common interpretation of "well lit" to mean brightly lit. Concerns over attempts to reduce lighting may also center on the safety of women at night. These all are very valid concerns. And this state of affairs may explain the somewhat piecemeal regulatory response to light pollution. Continued research can help support or refute these beliefs.

One notable gap in the literature is the lack of recent, reliable economic analyses of the financial costs of light pollution. Such analyses could better inform the levels and timescales of change required, and consideration of possible financial incentives such as subsidies. A 2010 paper indicates that ALAN cost USD$7 billion annually in the US alone (*13*). But what is "waste" light? Recent satellite mapping research has mapped upwardly escaping light, but not as yet quantified its cost. Nevertheless, greater understanding and awareness of the long-term savings offsetting the costs of retrofitting lighting is possible. For example, in 2002, Calgary retrofitted its streetlighting to "full cut off lights" (which do not emit light above the horizontal), and recouped the costs within 6 to 7 years, before saving roughly USD$2 million a year (*14*).

Calls for regulation often center on the night sky. For example, a leading international body in raising awareness and pushing for change (including excellent templates and guidance for regulation) is named the International Dark-Sky Association. The UK All-Party Parliamentary Select Committee is named the APPG for Dark Skies. Although the night sky is clearly the most graphic victim of ALAN, the average person may be unlikely to prioritize it among the reasons to light in the first place. Hence, a night-sky focus may be limiting public awareness and political will. An unintended consequence of the growing number of well publicized dark-sky accredited places has led to suggestions that if the night sky is protected in reserves, then there is no need to protect it in city and urban areas, with their perceived higher lighting needs for safety, security, commerce, and so on.

A slight change in tactics might better put ALAN on the radar of the general public (and so the politicians and regulators). The issue could be reframed around aspects more relevant to the lives of the general public. Education can describe how addressing light pollution is a readily achievable balance between the competing interests of the need to light against the adverse consequences, not simply a call to turn off all the lights to look at the night sky. Education could emphasize how this involves cutting energy waste and carbon emissions in a time of an energy and climate crisis, may reduce harm to human and ecological health, and may aid safety and security by cutting glare. "Well lit" might come to be understood not to mean brightly lit, but instead the use of more suitable and efficient lighting, aiming to light to a level actually required and that is on when, and only illuminating where, needed. These are issues that society is more likely to see as relevant to everyday life, and more worthy of action; and the night sky gets protected as a direct consequence.

What sort of regulation is likely to be the most effective? Dedicated laws, specifically drafted to address light pollution, such as in France and Korea, offer the most effective control. These could be supported by bolt-on laws where relevant. However, light pollution also needs preventive, guidance, and educational elements. Lighting should be considered at the design and build stage, to avoid light pollution from occurring in the first place, which in turn reduces the regulatory need and cost. Design and build also should be based on a full understanding of the entire light pollution problem, particularly the levels of lighting, the color spectra, light spill controls, and the way that the lighting can be installed physically (to avoid direct glare and upwardly escaping light).

Education is critical so that all light users and regulators are cognizant of the problems of ALAN and how a balance may be struck. Ownership of the balance may be best placed in the hands of stakeholder interest groups on all sides, who can contribute to the debate, to the creation and evolution of regulation, and to the drafting of guidance that might supplement hard law. There is also a need for further research to address the myths and misconceptions surrounding light at night. Education must break down preconceptions such as "the brighter the better" and "well lit" means brightly lit. Or that lighting is always needed, all the time, and that any form of lighting control is an unjustifiable fetter on personal autonomy. These attitudes could be replaced with understanding, namely, "the right amount of light, of the right type, where needed, when needed." Education should also lead to ALAN issues being taken more seriously by enforcement bodies. Education may help to shift some existing attitudes to lighting shown by the public and commerce. For example, changing recognition of the cost and value of lighting should lead to changing purchasing habits, for example, reducing the demand for lighting such as ultrabright floodlighting. The International Dark-Sky Association offers a wealth of excellent exemplar ordinances and underpinning education for regulators (*15*).

The public and regulators must recognize the benefits and value of the restrictions and burden that come with regulation. To this end, the most fruitful approach might be to lead on the adverse effects of light at night that are most relevant to everyday life, rather than the loss of the night sky. This needs to be bolstered with ongoing interdisciplinary research and active co-operation with stakeholders. In particular, education and research might center on the emotive matters that promote the use of light at night, such as the belief that light at night always cuts crime, and so the more light the better. Estimating the financial and carbon cost of light pollution would be beneficial as well. The result will be an evidence-based, and not an emotion-based, use of lighting. ■

REFERENCES AND NOTES

1. K. Zielinska-Dabkowska, *Nature* **553**, 274 (2018).
2. A. C. S. Owens *et al*., *Biol. Conserv.* **241**, 108259 (2020).
3. Convention on the Conservation of Migratory Species of Wild Animals, guidance document (2022); https://www.cms.int/en/document/draft-light-pollution-guidelines.
4. Light Pollution Reduction Measures in Europe, Working paper for the international workshop Light Pollution 2022, during the Czech Presidency of the Council of the European Union; https://czech-presidency.consilium.europa.eu/media/oo5cwlmn/light-pollution-reduction-measures-in-europe.pdf.
5. S. Donatello *et al*., Revision of the EU green public procurement criteria for road lighting and traffic signals–*Technical report and criteria proposal*. Publications Office of the European Union, Luxembourg, 2019.
6. COM/2019/640 final, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, The European Green Deal.
7. European Commission, COM/2018/395 final, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, EU Pollinators Initiative.
8. M. Morgan-Taylor, J. T. Kim, *J. Sustain. Light.* **18**, 21 (2016).
9. International Commission on Illumination, Guide on the Limitation of the Effects of Obtrusive Light from Outdoor Lighting Installations, 2nd Edition, CIE 150:2017; https://cie.co.at/publications/guide-limitation-effects-obtrusive-light-outdoor-lighting-installations-2nd-edition.
10. M. Morgan-Taylor, *J. Plan. Environ. Law*, 1114 (2006).
11. Royal Commission on Environmental Pollution (UK), Artificial Light in the Environment, (2009); https://www.gov.uk/government/publications/artificial-light-in-the-environment.
12. All Party Parliamentary Group, (UK Parliament), APPG for Dark Skies; https://appgdarkskies.co.uk/.
13. T. Gallaway, R. N. Olsen, D. M. Mitchell, *Ecol. Econ.* **69**, 658 (2010).
14. "Enviromart Streetlights Retrofit Project (Northwest Calgary): Protecting the Environment--One Streetlight at a Time," The Transportation Factor 2003. Annual Conference and Exhibition of the Transportation Association of Canada. (Congres et Exposition Annuels de l'Association des transport du Canada), 2003; https://trid.trb.org/view/700213.
15. International Dark-sky Association, https://www.darksky.org/our-work/lighting/public-policy/lighting-ordinances/.

ACKNOWLEDGEMENTS

M.M.-T. is a member of the board of Dark Sky UK, the UK chapter of the International Dark Sky Association.

10.1126/science.adh7723

REVIEW

Measuring and monitoring light pollution: Current approaches and challenges

Miroslav Kocifaj[1,2]*, Stefan Wallner[1,3], John C. Barentine[4]

Understanding the causes and potential mitigations of light pollution requires measuring and monitoring artificial light at night (ALAN). We review how ALAN is measured, both from the ground and through remote sensing by satellites in Earth orbit. A variety of techniques are described, including single-channel photometers, all-sky cameras, and drones. Spectroscopic differences between light sources can be used to determine which are most responsible for light pollution, but they complicate the interpretation of photometric data. The variability of Earth's atmosphere leads to difficulty in comparisons between datasets. Theoretical models provide complementary information to calibrate experiments and interpret their results. Here, we identify several shortcomings and challenges in current approaches to measuring light pollution and suggest ways forward.

Preserving the environment and ensuring sustainability are worldwide challenges. They include the phenomenon of light pollution caused by artificial light at night (ALAN). Light pollution primarily consists of misdirected light emission, illuminating outdoor areas not intended or required to be lit. It also includes overillumination—the use of lights with much higher brightness than necessary—and the use of harmful light colors, such as lighting that emits radiation at short optical wavelengths (blue light). Light pollution produces "light domes" visible in the night sky near cities, brightening the sky over wide areas and reaching into otherwise dark areas, such as protected natural spaces (*1*). The adverse consequences of light pollution include detrimental effects to flora and fauna and to human health (*2*–*4*). Increased night sky brightness (NSB) also impairs astronomical observations of celestial objects (*5*, *6*).

Reducing the negative impacts of light pollution requires environmentally responsible urban development. This is often taken to include the widespread conversion of lighting systems from inefficient incandescent or high-intensity gas discharge lamps to light-emitting diodes (LEDs) (*7*, *8*). However, current trends in the spatial and temporal distribution of ALAN show that switching to LEDs has been counterproductive for light pollution, with observations showing continuous growth in illuminated areas and upward-directed radiance worldwide, both being ~2% per year (*9*). In inhabited locations, the rate of increase can be even higher, with contemporaneous citizen-science data pointing to an increase in observed NSB of nearly 11% per year (*10*). Mapping NSB across the globe provides a baseline for investigating the worldwide emergence of lighted areas (*11*).

It is necessary to identify sources and quantify the impact of ALAN, particularly to guide regulations and other mitigation strategies (*12*, *13*). A multitude of measurement techniques are available, providing either single snapshots of lighting conditions and influences or long-term monitoring and remote sensing of ALAN. Many measurement devices are available, all of which have advantages and shortcomings (*14*–*17*). We review the methods behind quantifying light pollution and ALAN, focusing on the diverse functions. We consider current challenges in determining light pollution influences at arbitrary places and predicting how it will change over time. We also provide recommendations on how the measurements can be used more comprehensively in the future.

"...outdoor lighting design should minimize the amount of blue light emitted..."

Quantifying light pollution

Light pollution research uses a large variety of measurement techniques and devices. The right method, instrument, and analytical approach must be chosen for each application in analyzing ALAN and its effects.

Some light pollution parameters can be measured directly from the night sky itself. The enhancement of NSB caused by ALAN is generally called skyglow (Fig. 1A), most commonly occurring in and around densely inhabited areas. A clear night sky background without any ground-based light pollution has a luminance of ~200 μcd m^{-2} (*15*), equivalent to a value of 22.0 mag arc sec^{-2} in astronomical magnitudes (an inverted logarithmic scale) (*18*). Observations that include luminous celestial objects within the instrumental field of view show appreciably higher figures; for example, brightest parts of the Milky Way are ~2.5 times as bright as the surrounding night sky (*19*). Although the highest NSB values are measured within the light domes above cities, absolute measurements are highly dependent on the distance from (and proliferation of) individual light sources as well as the observed field of view on the night sky. The zenith—the point on the sky directly overhead—is often used as a local reference direction to characterize the approximate sky quality.

Figure 2 shows a comparison of different NSB measurement techniques. Photometric measurements are usually one-dimensional, having no angular resolution (*20*), and may be either portable for single readings or permanently installed as part of monitoring networks (*21*, *22*). Figure 2 includes an illustration of an NSB measurement process using devices with a specified field of view directed toward the zenith to collect continuous NSB data throughout a night. This technique is widespread—used by researchers and activists—because of its generally low data-acquisition cost and high accuracy. However, to collect information about the entire night sky rather than small fields of view, additional techniques are required to analyze skyglow. Two methods predominate. First, the horizontal illuminance of the overall radiation field can be measured using a simple light-to-frequency counter. Second, all-sky imaging techniques measure the entire hemisphere of the sky simultaneously (*23*, *24*) (also illustrated in Fig. 2). All-sky imaging has the advantage of not only measuring any NSB increases over time but also identifying the spatial distribution and relative contributions of individual light domes around the horizon. Combined with calibration software (*25*, *26*), the resulting night sky luminance matrices provide sufficient information to identify light pollution sources and the night sky quality at the time of observation.

ALAN directed toward the sky (directly or indirectly) can also be measured by spaceborne instruments (Fig. 1B). Whereas the ground-based techniques discussed above provide data on local conditions, satellite observations probe much larger spatial scales (*27*, *28*). Satellite remote sensing measures upward-directed radiance from light sources on Earth's surface (Fig. 2) with the goal of analyzing whole cities, countries, or other large areas. These data are particularly useful for studying extensive conversions of existing lighting systems, including potential changes in their total luminous flux, radiation angles, and other properties (*29*). They can also identify the type of lights installed on the ground (*30*).

Returning to smaller observational scales on the ground, ecological light pollution is widespread. The techniques used to study it depend on both the light source and the organism or ecosystem being investigated. When these conditions are clearly defined, light pollution

[1]Department of Optics, Institute of Construction and Architecture, Slovak Academy of Sciences, 845 03 Bratislava, Slovakia. [2]Department of Experimental Physics, Faculty of Mathematics Physics and Informatics, Comenius University, 842 48 Bratislava, Slovakia. [3]Department of Astrophysics, University of Vienna, 1180 Wien, Austria. [4]Dark Sky Consulting, Tucson, AZ 85730, USA.
*Corresponding author. Email: miroslav.kocifaj@savba.sk

measurements can be used to assess impacts on selected species of flora and fauna, caused by either single or multiple light emissions into the nighttime environment (*31*, *32*) (Fig. 1C). These measurements must consider the specific detector parameters required, for example, to provide analogs of animal eyes and the sensitivity to radiation of different origins (*33*, *34*). Light traps are often used to observe ground-based wildlife, and unmanned aerial vehicles (drones) are used to recreate the influence of ALAN on flying animals (*35*) (Fig. 2).

In urban management, outdoor lighting is a tool used to improve safety, provide orientation, and improve wayfinding. However, its use is not usually subject to meaningful public oversight. During the planning process, luminance and illuminance analyses (Fig. 1D) can be used to determine how to achieve those goals (*36*). Parameters, including the spatial distribution and illuminance, are adjusted as needed for site-specific requirements. However, influences such as light scattering effects and nonideal construction of light fixtures potentially lead to unexpected results and can cause light pollution. Therefore, determining luminance and illuminance is necessary in urban engineering. Within cities, individual light fixtures can be optimized on the basis of such analyses (*37*), and in dark areas, ecological light pollution can be explored using the same approach (*38*).

The impact of ALAN on the environment depends not only on its luminous flux but also its spectral power distribution. Exposure to short-wavelength (blue) light at night has negative consequences for many organisms (*3*, *39*). Even at very low illuminances, blue light disrupts the human sleep-wake cycle and suppresses secretion of the hormone melatonin, whose dysregulation is associated with metabolic diseases and certain cancers (*40*). To minimize potential harm, outdoor lighting design should minimize the amount of blue light emitted wherever possible. There are two approaches for quantifying this aspect of light pollution. First, individual lamps are characterized by their spectra (Fig. 1E), and the emission at shorter wavelengths is analyzed. Second, skyglow retains information about the light sources on the ground that generated it, which allows spectrographic measurements of ALAN (Fig. 2).

To forecast changes in light pollution, or its effect under varying ground-based conditions (e.g., lighting conversions and meteorological changes), theoretical modeling is applied. Several computational methods have been developed (*25*, *41*, *42*) to simulate ALAN and its spatial and temporal distribution at arbitrary locations and chosen input parameters. The accuracy of these models is limited by the (typically) large number of lights in cities, which differ in lumen outputs, spectral compositions, directional emissions, and spatial distributions, resulting in a nontrivial cumulative light emission pattern. Modeling of the angular distribution of urban photons has sometimes used a simple analytical formula, combining direct upward emissions with an assumed fraction of ground-reflected light (*43*). This approach has been improved by adding an extinction factor that accounts for light-blocking obstacles near the horizon (*44*). Although this approach works for some cities, the angular distribution of emitted light is generally more complex, requiring a combination of several model functions (*11*). The spectral composition of light escaping an urban area changes with direction as a result of the variety of light sources. These limitations in modeling source emissions affect NSB across the modeling domain.

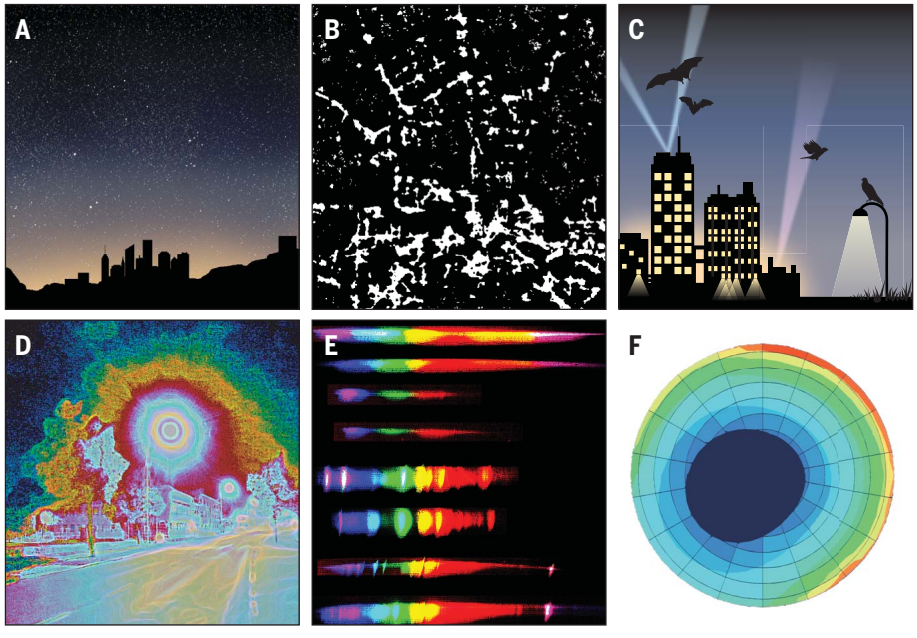


Fig. 1. Topical areas that require quantification of light pollution. Six areas that quantify the impacts of ALAN are illustrated: (**A**) Higher levels of NSB, also called skyglow. (**B**) Satellite remote sensing of upward-directed radiance from Earth's surface. (**C**) Ecological light pollution. (**D**) Surface illuminance resulting from individual light sources—the colors indicate the brightness levels on an illuminated building. (**E**) Emission spectra of individual lighting fixtures. (**F**) Computational modeling. For each application, different techniques and analysis methods are necessary.

Current challenges in measuring light pollution

Although there are many different methods of quantifying the impact of ALAN, they each have shortcomings. Despite the variety of instruments available to observe skyglow, radiance, or other characteristics, there are several common problematic aspects. Because light pollution is affected by the scattering of light in the lower atmosphere, it is influenced by the spatial and temporal instability of the atmosphere. Each night is potentially different from its predecessor or successor, depending on atmospheric and meteorological conditions, so measurements obtained on different dates might not be directly comparable. This primarily affects long-term analyses from monitoring stations because both NSB and atmospheric conditions must be recorded simultaneously and considered together. This applies to all measurement techniques; for example, fog affects satellite remote sensing as well as ground-based illuminance measurements of a single lighting fixture. Ecological light pollution studies are particularly affected because the atmospheric conditions change the spatial distribution of ALAN (*45*). Solutions to this problem can use data from atmospheric monitoring stations, if present near the measurement sites, but those are usually not available, so researchers must rely heavily (or solely) on theoretical models (*46*).

Other factors also cause light pollution measurements to vary. The types and flux density of lighting sources change over time—even during a single night, as lights are switched on and off—and vary spatially as a result of the influence of local topographical features. Parameters, including the albedo of the illuminated surface and shadowing by physical obstructions, affect the resulting light pollution and its analysis (*47*). These variations can be incorporated into a mathematical description known as the emission function (EF) (*48*). It is not possible to obtain sufficient data to determine the EF completely, but it is commonly used as an input for data analysis and processing. Several approaches have been developed to approximate the EF of urban areas (*49*, *50*).

Technical limitations of available devices also affect light pollution measurements. The

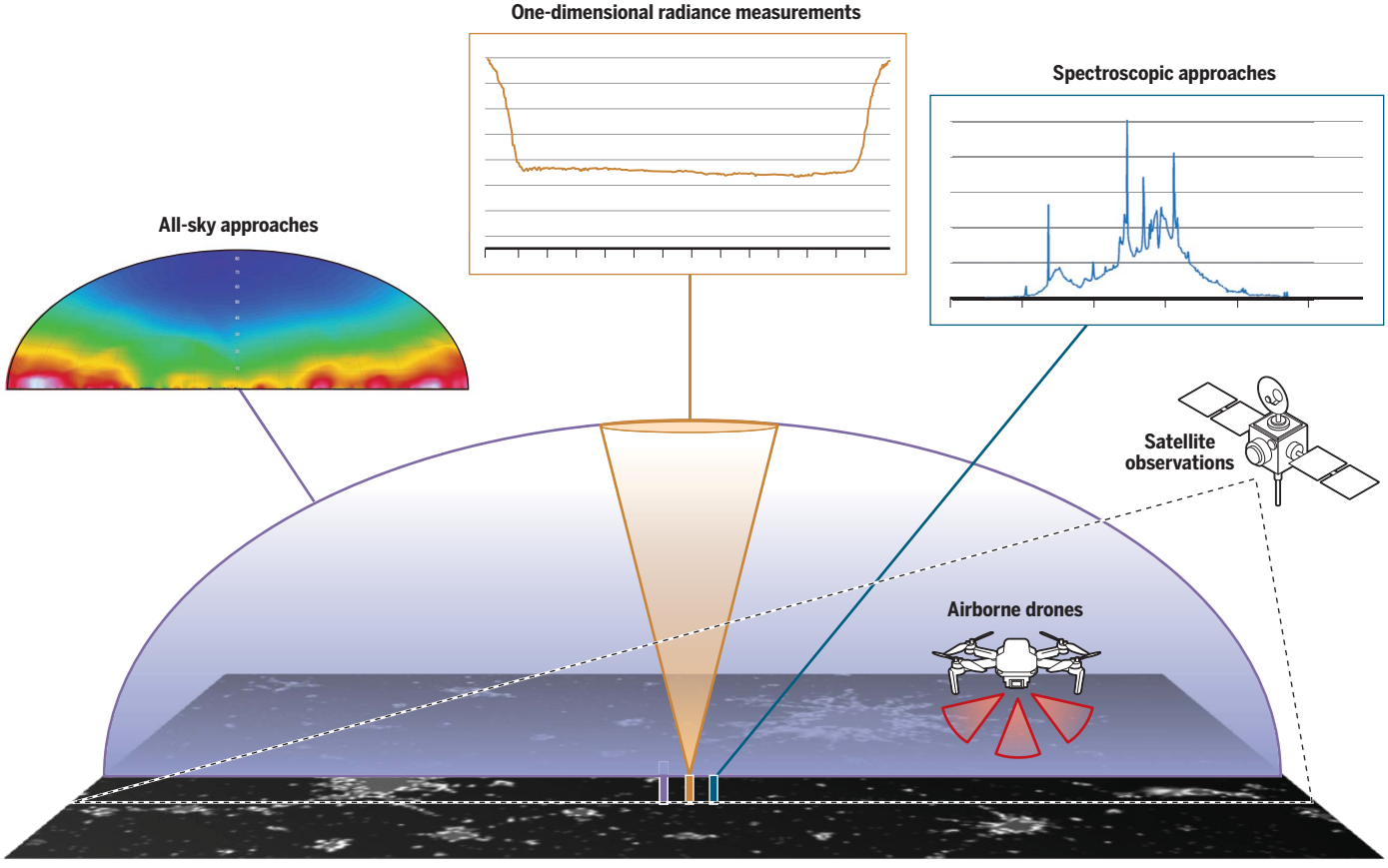


Fig. 2. Illustration of available measurement techniques for light pollution analyses. The variety of methods for measuring light pollution, their domains of applicability, and related observables are shown. All-sky approaches (purple semicircle) collect information originating in the upper hemisphere; spectroscopic approaches (blue line) provide information on the wavelengths of light; one-dimensional radiance measurements (orange cone) typically sample a small region around the zenith; satellite observations (black dashed lines) observe radiance from sources on the ground escaping to space; and airborne drones (red triangles) measure (spectral) radiance and irradiance in multiple directions.

differing spectral sensitivities of different devices combine with the potentially varying spectral power distributions of light sources (discussed above) in complex ways that affect the measurement. If the types of ground-based light sources are unknown and not included as a variable in the analysis, then the measurement conditions before and after a lighting conversion might not be directly comparable (*51*). This is also true for comparisons between different instruments. Satellite observations are particularly affected by this issue. For example, the main source of orbital ALAN data, with global coverage and nightly temporal cadence, is insensitive to light with wavelengths shorter than 500 nm (*52*). Space-based measurements are also influenced by the emission angle of ground-based light sources. Ground-based measurements have shown that zenithal observations can differ substantially from those obtained at low elevation angles (*53*, *54*). The lower signal-to-noise ratio of measurements taken in low-light environments makes them less reliable than locations with high levels of light pollution, and complex corrections because of natural night airglow are required (*55*). Portable instruments, such as cameras or drones, can be easily maintained, but permanently installed devices are susceptible to meteorological conditions, such as rain and snow. Solar radiation during the daytime has been shown to affect the optics of instruments, causing an aging effect that influences nighttime measurements (*56*).

There is a wide range of instrumental and environmental influences that affect light pollution measurements, which must be considered during data processing. The wide range of techniques can be advantageous for tackling different research goals, but the lack of measurement standards contributes to difficulty in comparing results and the need for complex interpretations. Yet, measurements made consistently with well-designed protocols over long periods of time can yield information of distinct value to light pollution researchers and dark-sky activists alike. Provided that data are obtained with care and the instrumental limitations are understood, light pollution measurements can be confidently applied to situations involving urban planning, land management, natural resource conservation, and more.

Using data more comprehensively

There is great potential for extracting more information from measurement data than is typical at present, for example, through long-term observations of NSB, which are scarce and generally only consist of zenith radiance data. Monitoring networks routinely operate single-channel optical instruments to gather time series of zenith radiance for trend analysis (*57*). However, such data contain more information than is inferred from simple trend statistics. Exploiting more of the information content of zenith radiance measurements has been demonstrated in nighttime monitoring of atmospheric aerosols using differential photometry (*58*). Zenith radiances obtained in rough or irregular terrain, from two measuring stations separated in elevation, have been used to characterize the turbidity of the atmospheric layer between the stations. Conventional measurements, when taken under suitable configurations or spatial arrangements, can therefore provide additional information about the nighttime environment (*58*).

Aerosols—tiny particles suspended in the air—are a large source of uncertainty in quantifying the impacts of ALAN. Several measuring techniques

and tools for retrieving aerosol properties have been developed to determine the aerosol optical depth (AOD), a parameter used in, for example, modeling the influence of ground-based light sources. Multiple techniques are in use to determine AOD, but only a few of them are applicable at night, and most are difficult for inexperienced experimentalists to use. One method useful for light pollution measurements relies on an empirical relationship between the zenith brightness and AOD; it can be implemented with low-cost optical devices during moonless nights (*21*).

Simple measurement techniques are preferable for use in monitoring programs at many locations. As discussed above, local atmosphere data are highly advantageous to interpret light pollution measurements, but the instrumentation for measuring atmospheric conditions is not present at most sites. For example, ceilometers—devices that use lasers or other light sources to determine the height of a cloud ceiling or cloud base—can provide useful information on local atmospheric conditions that can be used as inputs for light pollution analyses and skyglow modeling (*59*). When systematically used, measurements of cloud base altitude and backscatter from aerosols provide complementary information to light pollution data.

Spectral data are usually required to characterize light sources (*60*) and to quantify light field distributions for a broad range of atmospheric conditions (*61*). Such measurements are rarely available. Ground-based spectral measurements are infrequent (*62*) because optical systems with the required sensitivity are expensive. Space-based spectral measurements require highly sensitive detectors with high spectral and spatial resolutions. Orbital remote sensing of ALAN mostly uses the Day-Night Band—part of the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument aboard the Suomi spacecraft (*16*)—which can be used to map directional outputs from cities for a range of zenith viewing angles (*63*). The collective effect of the whole-city lights (i.e., its cumulative angular emission pattern) is a required input for, for example, ecological light pollution measurements or forecast approaches, but is difficult to determine experimentally. Exploitation of satellite data for this purpose is a challenge for both experimentalists and theorists. VIIRS does not have a multispectral capability with both sufficient sensitivity and panchromatic response in the optical range. Multiple-angle remote sensing could be used for extended diagnosis of the atmosphere and artificial lights (*53*).

Conclusion and outlook

Experiments and theoretical studies are equally useful to investigate light pollution because they provide complementary information about the nighttime environment. Field experiments can never be performed under fully controlled conditions; the data gathered by optical systems are therefore not free of errors or the effects of other physical phenomena. Measurements are only possible in discrete locations because data acquisition in an arbitrary spatial pattern is impractical. Theoretical studies are required to address these issues. Models are also useful in determining the isolated effect of single parameters on the light field, improving our understanding of their impacts on the measured quantities. However, the models are accurate only within the limitations of the theories used. Experimental data and theoretical models are complementary, providing incomplete information if isolated from each other. Theories can fill data gaps, whereas experimental data are necessary to test the theories.

The development of new models and experimental techniques should go hand in hand because the outcome of one drives progress in the other and can generate new applications. Understanding the processes of light emission and propagation allows for more-specialized field experiments that more fully use the information content of nighttime light measurements. For example, the polarization of light at night is largely unexplored in light pollution research.

Light pollution has drawn increasing attention from the scientific community in recent decades, and we expect that trend to continue. There is a need for more-accurate devices, data acquisition, and study management—all activities that have high technical demands. As the number and diversity of instruments available for field light pollution measurements continue to increase, we question whether a technical standard for absolute calibration of their data can be achieved. Given the need for more global collaboration in the interdisciplinary field of light pollution research, we feel that standardization of measurement protocols will be necessary.

REFERENCES AND NOTES

1. D. Duriscoe, C. Luginbuhl, C. Elvidge, *Light. Res. Technol.* **46**, 35–49 (2013).
2. R. G. Stevens, Y. Zhu, *Phil. Trans. R. Soc. B* **370**, 20140120 (2015).
3. F. Falchi, P. Cinzano, C. D. Elvidge, D. M. Keith, A. Haim, *J. Environ. Manage.* **92**, 2714–2722 (2011).
4. T. Longcore, C. Rich, *Front. Ecol. Environ.* **2**, 191–198 (2004).
5. A. Crumey, *Mon. Not. R. Astron. Soc.* **442**, 2600–2619 (2014).
6. R. F. Green, C. B. Luginbuhl, R. J. Wainscoat, D. Duriscoe, *Astron. Astrophys. Rev.* **30**, 1 (2022).
7. C. Baddiley, *J. Quant. Spectrosc. Radiat. Transf.* **267**, 107574 (2021).
8. J. C. Barentine *et al*., *J. Quant. Spectrosc. Radiat. Transf.* **212**, 10–23 (2018).
9. C. C. M. Kyba *et al*., *Sci. Adv.* **3**, e1701528 (2017).
10. C. C. M. Kyba, Y. Ö. Altıntaş, C. E. Walker, M. Newhouse, *Science* **379**, 265–268 (2023).
11. F. Falchi *et al*., *Sci. Adv.* **2**, e1600377 (2016).
12. S. Schroer, B. J. Huggins, C. Azam, F. Hölker, *Sustainability* **12**, 2551 (2020).
13. F. Hölker *et al*., *Ecol. Soc.* **15**, 13 (2010).
14. S. Mander, F. Alam, R. Lovreglio, M. Ooi, *Sustain Cities Soc.* **92**, 104465 (2023).
15. J. C. Barentine, *Nat. Astron.* **6**, 1120–1132 (2022).
16. N. Levin *et al*., *Remote Sens. Environ.* **237**, 111443 (2020).
17. A. Hänel *et al*., *J. Quant. Spectrosc. Radiat. Transf.* **205**, 278–290 (2018).
18. S. Bará, M. Aubé, J. Barentine, J. Zamorano, *Mon. Not. R. Astron. Soc.* **493**, 2429–2437 (2020).
19. P. Deverchère, S. Vauclair, G. Bosch, S. Moulherat, J. H. Cornuau, *Sci. Rep.* **12**, 17050 (2022).
20. S. Bará, C. E. Tapia, J. Zamorano, *Sensors* **19**, 1336 (2019).
21. C. Marseille, M. Aubé, A. Barreto, A. Simoneau, *Remote Sens.* **13**, 4623 (2021).
22. T. Posch, F. Binder, J. Puschnig, *J. Quant. Spectrosc. Radiat. Transf.* **211**, 144–165 (2018).
23. A. Müller, G. Wuchterl, M. Sarazin, *Rev. Mex. Astron. Astrofis.* **41**, 46–49 (2011).
24. A. Jechow, C. C. M. Kyba, F. Hölker, *J. Quant. Spectrosc. Radiat. Transf.* **250**, 106988 (2020).
25. Z. Kolláth, D. Száz, K. Kolláth, *Remote Sens.* **13**, 3653 (2021).
26. A. Jechow *et al*., *J. Quant. Spectrosc. Radiat. Transf.* **209**, 212–223 (2018).
27. C. D. Elvidge *et al*., *Front. Remote Sens.* **3**, 919937 (2022).
28. M. Zhao *et al*., *Remote Sens.* **11**, 1971 (2019).
29. C. C. M. Kyba *et al*., *Remote Sens.* **7**, 1–23 (2015).
30. N. Rybnikova, A. Sánchez de Miguel, S. Rybnikov, A. Brook, *Remote Sens.* **13**, 4413 (2021).
31. A. E. Aulsebrook *et al*., *Biol. Lett.* **18**, 20220035 (2022).
32. J. Bennie, T. W. Davies, R. Inger, K. J. Gaston, *Methods Ecol. Evol.* **5**, 534–540 (2014).
33. V. J. Alaasam, M. E. Kernbach, C. R. Miller, S. M. Ferguson, *Integr. Comp. Biol.* **61**, 1170–1181 (2021).
34. B. M. Seymoure, C. Linares, J. White, *J. Zool.* **308**, 93–110 (2019).
35. T. Degen, Z. Kolláth, J. Degen, *Ecol. Evol.* **12**, e9608 (2022).
36. M. T. Vaaja *et al*., *Remote Sens.* **7**, 11389–11402 (2015).
37. S. Wallner, *J. Imaging* **5**, 86 (2019).
38. A. Jechow, C. C. M. Kyba, F. Hölker, *J. Imaging* **5**, 46 (2019).
39. K. J. Gaston, *Science* **362**, 744–746 (2018).
40. Y. Cho *et al*., *Chronobiol. Int.* **32**, 1294–1310 (2015).
41. M. Kocifaj, *Appl. Opt.* **46**, 3013–3022 (2007).
42. M. Aubé, L. Franchomme-Fosse, P. Robert-Staehler, V. Houle, in *Atmospheric and Environmental Remote Sensing Data Processing and Utilization: Numerical Atmospheric Prediction and Environmental Monitoring*, H.-L. A. Huang, H. J. Bloom, X. Xu, G. J. Dittberner, Eds. (SPIE, 2005), p. 589012.
43. R. H. Garstang, *Publ. Astron. Soc. Pac.* **98**, 364–375 (1986).
44. C. Luginbuhl *et al*., *Publ. Astron. Soc. Pac.* **121**, 204–212 (2009).
45. S. Wallner, M. Kocifaj, *J. Environ. Manage.* **335**, 117534 (2023).
46. J. Puschnig, S. Wallner, A. Schwope, M. Näslund, *Mon. Not. R. Astron. Soc.* **518**, 4449–4465 (2023).
47. S. Wallner, M. Kocifaj, *J. Quant. Spectrosc. Radiat. Transf.* **239**, 106648 (2019).
48. H. A. Solano Lamphar, *J. Quant. Spectrosc. Radiat. Transf.* **211**, 35–43 (2018).
49. B. R. Espey, *Remote Sens.* **13**, 3827 (2021).
50. J. Petržala, *J. Quant. Spectrosc. Radiat. Transf.* **213**, 86–94 (2018).
51. A. Sánchez de Miguel *et al*., *Mon. Not. R. Astron. Soc.* **467**, 2966–2979 (2017).
52. L. B. Liao, S. Weiss, S. Mills, B. Hauss, *J. Geophys. Res. Atmos.* **118**, 12705–12718 (2013).
53. C. C. M. Kyba *et al*., *J. Geophys. Res. Atmos.* **127**, e2021JD036382 (2022).
54. L.-W. Hung, S. J. Anderson, A. Pipkin, K. Fristrup, *J. Environ. Manage.* **292**, 112776 (2021).
55. C. C. M. Kyba, J. Coesfeld, *Int. J. Sustain. Light.* **23**, 51–57 (2021).
56. J. Puschnig, M. Näslund, A. Schwope, S. Wallner, *Mon. Not. R. Astron. Soc.* **502**, 1095–1103 (2021).
57. P. Fiorentin *et al*., *Remote Sens.* **14**, 5787 (2022).
58. M. Kocifaj, S. Bará, *Mon. Not. R. Astron. Soc.* **500**, L47–L51 (2021).
59. K. Kolláth, Z. Kolláth, *J. Quant. Spectrosc. Radiat. Transf.* **253**, 107158 (2020).
60. J. Puschnig, T. Posch, S. Uttenthaler, *J. Quant. Spectrosc. Radiat. Transf.* **139**, 64–75 (2014).
61. P. Cinzano, F. Falchi, *Mon. Not. R. Astron. Soc.* **427**, 3337–3357 (2012).
62. M. Kocifaj, F. Kundracik, D. M. Duriscoe, S. P. Balm, S. Wallner, *Mon. Not. R. Astron. Soc.* **506**, 2739–2745 (2021).
63. X. Li *et al*., *Remote Sens. Environ.* **271**, 112920 (2022).

ACKNOWLEDGMENTS

Funding: M.K. was funded by the Slovak Research and Development Agency (APVV) (grant APVV-18-0014). S.W. was funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie mechanism, grant agreement no. 945478 - SASPRO 2. **Author contributions:** M.K. and S.W. conceived the content of the article and substantially wrote the text. J.C.B. analyzed and commented on the findings and edited and expanded the text. **Competing interests:** The authors declare no competing interests.

Submitted 30 January 2023; accepted 24 April 2023
10.1126/science.adg0473

REVIEW

Effects of anthropogenic light on species and ecosystems

Annika K. Jägerbrand[1]* and Kamiel Spoelstra[2]

Anthropogenic light is ubiquitous in areas where humans are present and is showing a progressive increase worldwide. This has far-reaching consequences for most species and their ecosystems. The effects of anthropogenic light on natural ecosystems are highly variable and complex. Many species suffer from adverse effects and often respond in a highly specific manner. Ostensibly surveyable effects such as attraction and deterrence become complicated because these can depend on the type of behavior and specific locations. Here, we considered how solutions and new technologies could reduce the adverse effects of anthropogenic light. A simple solution to reducing and mitigating the ecological effects of anthropogenic light seems unattainable, because frugal lighting practices and turning off lights may be necessary to eliminate them.

Artificial light is ubiquitous in areas where humans are present and inevitably extends to our natural environment. This has far-reaching consequences for most species and their ecosystems. The ability of humankind to produce electric light has enabled us, a naturally diurnal species, to dispel darkness and extend our activities into the night. However, artificial light has serious side effects that are commonly referred to as light pollution. Light pollution is defined as the sum total of all of the adverse effects of artificial (hereafter referred to as anthropogenic) light (*1*). Ecological light pollution was originally defined as artificial light that alters the natural patterns of light and dark in ecosystems (*2*) that are caused by exposure to near and distant light sources and sky brightness. Light-polluted skies have become a global reality, affecting most of the world's economically developed areas (*3*). Sky brightness has increased over time, eroding natural darkness and encroaching on protected terrestrial and marine areas. Anthropogenic light not only worsens climate change through energy consumption but also poses serious challenges across species and ecosystems (*4*).

Over the past 15 years, there has been a substantial amount of research on the ecological effects of anthropogenic lighting ac[illegible] the globe (*5*, *6*). Most of these studies h[illegible] focused on direct light exposure. Research on broad-scale spatial patterns has become possible by combining remote sensing data on light emitted upward with digitalized biological data such as information on species occurrence and migration routes.

The deleterious effects of anthropogenic light have been reviewed for several species groups, including bats (*7*), insects (*8*–*11*), seabirds (*12*), fish (*5*), vertebrates (*13*), and marine, shoreline, and estuarine species (*14*–*16*). Effect sizes for numerous species have also been reviewed (*6*). These reviews line up numerous studies that have led to substantially more knowledge about the effects on different species groups, the diverse nature of such effects, and how they manifest across trophic levels, thereby increasing awareness about this environmental problem.

Commonly recommended solutions to mitigating the ecological effects of anthropogenic light include reductions or adaptations in light intensity, distribution, spectra, and duration. New technologies, such as light-emitting diodes (LEDs), can aid in reducing the effects of anthropogenic light on the natural environment. However, these solutions have limitations and may not safeguard against deleterious effects on all species.

Herein, we review the ways in which anthropogenic light affects species and ecosystems, discuss the research progress made in recent years, and describe various light pollution management solutions.

[1]Department of Electrical Engineering, Mathematics and Science, Faculty of Engineering and Sustainable Development, University of Gävle, 801 76 Gävle, Sweden. [2]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6700 AB Wageningen, Netherlands.
*Corresponding author. Email: annika.jagerbrand@hig.se

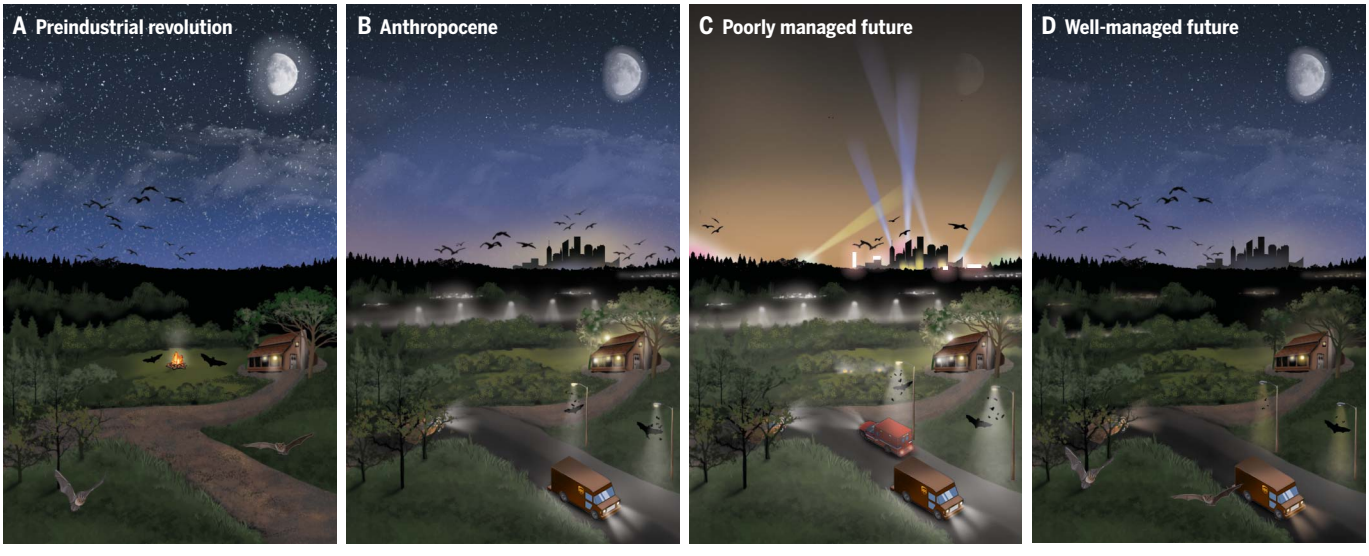


ILLUSTRATION: K.HOLOSKI/*SCIENCE*, BASED ON ANNIKA K. JÄGERBRAND AND KAMIEL SPOELSTRA

Fig. 1. Anthropogenic light through history and possible futures. (**A**) During the preindustrial revolution, few light sources were used for outdoor activities during darkness. Natural light sources dominated the natural environment, bats foraged and commuted along forest edges, migrating birds were undisturbed by strong light sources, and there was no sky glow from cities. (**B**) In the Anthropocene, anthropogenic lighting is used where it is needed to enhance human activities, attracting insects, migrating birds, and foraging bats and causing visible sky glow. (**C**) As anthropogenic lighting is increasingly used, light pollution will also increase and result in higher mortality in insects and migrating birds, habitat loss for light-repelled bats, and increased foraging opportunities for synanthropic bats. Barrier effects and fragmentation of dark ecosystems will decrease habitat quality. (**D**) A well-managed future conserves dark areas and limits light use in several ways, thereby minimizing the effects of anthropogenic light on birds, insects, and bats and numerous other species.

Effects of anthropogenic light on species groups

Birds

One of the most established effects of anthropogenic light on birds is their response to it during migration (Fig. 1). Many birds, including otherwise diurnal species, migrate at night. They are attracted to light and disoriented by it, especially strong light sources and bright spots in dark areas. This attraction to light can not only cause them to collide with buildings, lighthouses, oil rigs, and ships (*17*), but may also divert them from suitable stopover locations (*18*). Migratory routes are often close to illuminated urban areas (*19*). In areas with dark surroundings, such as islands, light sources attract seabird fledglings (*12*). These effects can directly result in high mortality and exhaustion. Many other effects are less pervasive but still problematic; for example, light may induce stress and disturb sleep (*20*, *21*). Many bird species in temperate zones depend on the accurate seasonal timing of breeding and the accurate daily timing of song activity and foraging, so light disturbance of this temporal organization is problematic [(*22*, *23*); Fig. 2].

Mammals

Most bat species are highly nocturnal and respond strongly to light. This response is driven by the fear of predators or the foraging ecology. For example, fast-flying, agile bats are less frightened by light and forage on accumulated insects around light sources, whereas slow-flying bats stay in sheltered locations (*24*) (Fig. 3). Bats are particularly vulnerable because they use linear landscape elements such as forest edges, hedgerows, and streams to guide them in foraging and commuting. Light along commuting and foraging routes can act as a barrier, thereby amplifying its negative effects by fragmenting habitat networks. Anthropogenic light in or near bat roosting sites can lead to delayed emergence or roost abandonment, which has negative consequences for the survival of populations (*7*).

A commonly accepted assumption is that, like bats, other nocturnal mammal species' responses to light are driven by fear of predators. A decrease in the activity of nocturnal species by light at night has been reported in numerous laboratory experiments. Many mammal species reduce activity in response to moonlight (*25*, *26*) and show a comparable response to anthropogenic light (*27*). Likewise, anthropogenic light can change the spatial behavior of rodents (*28*) and, even in the absence of predation, reduce longevity and reproduction (*29*). Larger mammal species such as deer and predators may also change their spatial activity (*30*). On a larger scale, the deterrent effects of light on infrastructure may have far-reaching effects because mammals may stop using passage structures (*31*). As is the case with many other species, anthropogenic light can have profound effects on mammals' daily and seasonal rhythms of activity, physiology, and reproduction (*32*).

Invertebrates

The attraction of insects to light is a well-known phenomenon (Fig. 1). In naturally dark environments, insects can be attracted to low light intensities (*33*), and the extent of attraction to a light source depends on the presence of surrounding lights (*34*). Insect's flight-to-light may be a maladaptive response to the original orientation toward moonlight (*35*) and is related to color composition. Blue light (<500 nm) attracts more insects than the yellow and red parts of the spectrum (*36*). The attraction of insects to anthropogenic light may cause mortality and exhaustion, which may play a substantial role in global insect declines. This has been corroborated by the fact that phototactic nocturnal species that fly toward light have shown strong population declines (*37*). Indeed, light posts have been found to affect local moth caterpillar abundance (*38*). Moreover, insect declines can be caused by the negative effects of light on reproduction and development (*10*). Insects that depend on bioluminescent signaling, such as fireflies, are especially vulnerable to anthropogenic light and are directly impaired in reproduction (*11*). Anthropogenic light deprives insects such as aquatic insects and dung beetles of their ability to use light cues for orientation (*39*). For several other invertebrate groups, the effects of light are well documented, including opportunistic foraging around light sources by spiders (*40*) and slugs (*41*). Finally, anthropogenic light can change species composition in invertebrate communities (*42*).

Amphibians

Early field observations have provided evidence that anthropogenic light affects the reproduction, visual performance, and activity patterns of amphibians (*2*). In toads, exposure to anthropogenic light has been found to cause reduced activity levels, altered energy allocation, and decreased juvenile growth and metamorphic duration (*13*). Anthropogenic light can also alter breeding behavior and reduce the fertilization success of toads (*43*). In frogs, although mate choice behavior appears to be unaffected by anthropogenic light, it has been shown to shorten the calling season and shift the daily calling period (*13*).

Reptiles

Knowledge about the impact of anthropogenic light on reptiles in general is limited, but its effect on marine turtle populations is widely known. Hatchlings are highly susceptible to disorientation caused by light when crossing the beach to reach the sea, leading to high mortality rates. Furthermore, even at low levels, anthropogenic light can disrupt the on-beach orientation of turtles, resulting in suboptimal selection of nesting sites (*14*). Turtle nests have declined in lit areas across several species, with turtles using lit beaches less frequently or avoiding them altogether (*15*). Some diurnal reptiles, such as geckos, are able to forage at night in the presence of anthropogenic light (*44*). Likewise, green anole lizards (*Anolis carolinensis*) express part of their normal daytime foraging and display activity during the night (*45*).

Fish

Compared with terrestrial organisms, fish have received less attention in studies investigating the effects of anthropogenic light (*5*). Such light may affect fish populations through changes in survival rates, spawning, hatching success, and physiology, and can alter the temporal and spatial activity of fish and increase their energy expenditure. This can stimulate various behaviors such as nest-guarding activity, higher nocturnal activity, overall activity during both day and night, increased time spent in open areas, and maintaining position in an area. Fish species can be attracted or repelled by light. For example, positive phototaxis can lead to the aggregation of smaller fish around lighting, thereby providing easy prey for predatory fish. Some species may benefit from foraging under brighter conditions, but whether the benefits outweigh the costs over time is still being determined (*5*). Very low light levels can have effects on fish behavior. For example, rainbow trout (*Oncorhynchus mykiss*) are attracted to low levels of bridge illumination, which further depends on various environmental factors, making it a complex issue (*46*).

Plants

Although the effects of anthropogenic lighting on commercially grown plants are well documented, there is a gap in the knowledge about its effects on plants and plant-mediated responses in the natural environment. Lighting close to trees can lead to increased photosynthesis and morphological changes through the relocation of biomass from roots to leaves (*47*). In deciduous trees, exposure to anthropogenic light advances the emergence of leaf buds and, together with temperature, delays the coloring of leaves (*48*). Species-specific variations in responses to anthropogenic light have also been demonstrated in herbs and grasses (*49*).

Ecosystem effects of anthropogenic light

Nocturnal environments are dominated by moonlight because it is the strongest and most abundant natural light source. Light from anthropogenic sources differs from natural light in magnitude, color composition, and temporal and spatial presence. Moonlight intensities at midlatitudes can be as high as 0.2 lux depending on the position above the horizon and the amount of cloud cover (*50*). In many areas, however, the amount of anthropogenic light surpasses moonlight.

Anthropogenic light sources usually have a spectral composition that differs substantially

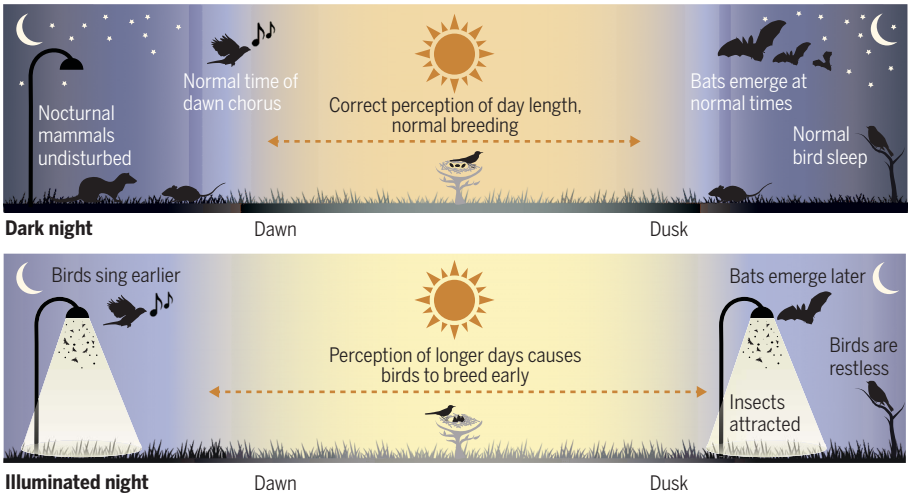


Fig. 2. Examples of disturbance of daily and seasonal rhythms by anthropogenic light. Nocturnal rodents are less active, and diurnal species can be restless in illuminated nights. Birds advance their dawn song, and bats delay emerging until the evening. Light at night can disrupt the perception of day length in birds, causing them to breed earlier in spring.

from natural nocturnal light: Outdoor LEDs typically have a peak in the blue spectral range, resulting in a cooler, white light compared with the yellowish light of the moon. Traditional light sources, such as high-pressure or orange low-pressure sodium lamps, are different from natural light because they have very little or no blue light. The intensity and distribution of natural nocturnal light vary as the night progresses and are also influenced by changes in the lunar phases and weather conditions. By contrast, anthropogenic light is often constant and concentrated around human settlements.

Terrestrial landscapes vary in topography and in the heights of physical objects, which result in variations in light distribution and exposure to organisms.

Forest ecosystems are generally darker as light is filtered through and absorbed by the vegetation, and only a small proportion of ambient light reaches the ground. Species that inhabit forests are believed to have adapted to darker nocturnal environments (*51*). In open environments, light can propagate far from the source, and even a single strong light can be a conspicuous element in the landscape, resulting in glare and difficulties in seeing details in the surroundings and stars in the night sky. In general, open environments are perceived as brighter when larger parts of the naturally lit sky and the ground can be viewed. Low-growing and light-green vegetation such as grasses, bushes, and herbs can reflect a high proportion of light, which increases the luminance of the landscape.

In aquatic ecosystems, anthropogenic light is reflected on water surfaces and thus has a high probability of propagating widely over open water and may affect large areas.

Many human structures in or near aquatic ecosystems, such as ports, ships, and oil rigs, use high-intensity lighting with insufficient restrictions, often resulting in large amounts of light spilling into the surroundings. In water, light changes with depth. Water and its particulates absorb and scatter light, reducing light intensity, altering color composition, and changing the degree of polarization. Clear ocean waters absorb ultraviolet, red, orange, and yellow wavelengths at the top of the water column, allowing blue light to penetrate the deepest, which results in a bluer color at greater depths. Coastal and freshwater systems often contain suspended particulates and phytoplankton, which selectively absorb the light, causing the water to appear more yellow-green, orange, or brown in color. Aquatic species have adapted their eye morphology to natural variations in intensity and color composition with photoreceptors that match the color of light, thereby increasing photon absorption (*52*). For example, deep-sea fishes have a visual pigment that matches the color of downwelling oceanic light, whereas fish species found in yellow-green coastal waters and inland freshwater lakes have photoreceptors with absorption peaks at longer wavelengths. Because light diminishes with depth, organisms in the mesopelagic zone have adapted to more dim light conditions with larger eye size and wider pupils. Aquatic species are expected to be vulnerable to anthropogenic light at night because of their high photosensitivity at low illuminance (*53*), which is an adaptation to decreased light intensity as light is filtered out by water.

Effects of anthropogenic light on temporal organization

Light is the key driver of the most important temporal niches in nature. The contrast between high light levels during the day and low levels at night enables species to share the same habitat within a 24-hour cycle. For most species, photic conditions are essential for their ability to survive (*24*). Organisms need to optimally schedule activity, rest, and sleep in (species-specific) natural light conditions.

The disturbance of rhythms in natural systems by light at night has attracted increasing interest (*54*), particularly in laboratory studies. For example, great tits (*Parus major*) exposed to low light levels during the dark phase of the light cycle were found to advance the start of their daily activities [e.g., (*55*)]. Rhythms in laboratory mice were shown to be weakened by dim light at night, and their hormonal rhythms were affected (*56*). Light at night has also been shown to disrupt circadian gene expression in great tits (*57*). In the field, an advanced onset of dawn song has been observed for several bird species, and individual blackbirds (*Turdus merula*) exposed to higher levels of anthropogenic light begin their activity earlier in the day (*58*). By contrast, nocturnal species respond to light by delaying their activity. For example, least horseshoe bats (*Rhinolophus pusillus*) emerge later, when their roost is illuminated (*59*). Bird species that normally start their dawn song relatively early during still low light levels, such as the robin (*Erithacus rubecula*), respond more strongly to the presence of artificial light compared with other birds (e.g., the blue tit, *Cyanistes caeruleus*) that start dawn song later in the morning [(*23*); Fig. 2].

In temperate zones, anthropogenic light interferes with the annual cycle of species. Trees delay shedding leaves from branches close to streetlights (*60*). Anthropogenic light can advance reproduction in birds (*22*), delay reproduction in mammals (*61*), and prolong yearly reproduction in insects (*62*).

Anthropogenic light affects interactions and trophic levels

Changes in the population of a species caused by anthropogenic light inevitably cause changes in the food web. Light can facilitate foraging for predators by concentrating prey species, such as for synanthropic bat species that catch light-attracted flying insects (Fig. 3) (*63*). Such changes in the predator–prey interaction may, however, be part of a trade-off for bats because they need to prevent exposing themselves too much (*64*). Foraging may be further facilitated by providing better visual detection of prey, as was shown for burrowing owls (*Athene cunicularia*), which could expand their foraging habitat into urbanized areas using the presence of anthropogenic light (*65*). Conversely, prey species may avoid illuminated, otherwise suitable habitat or change foraging behavior. Such changes have been shown in jerboas (*Allactaga sibirica*), which spend less time searching for food in illuminated conditions (*66*). Likewise, in aquatic systems, light at night can counteract the benefits of shelter material for amphipods (*Gammarus fossarum*) seeking shelter from the predatory Eurasian perch (*Perca fluviatilis*) (*67*).

Complex effects have been reported about the interaction of urbanization with the response to light. In urban areas, mule deer (*Odocoileus hemionus*) use illuminated foraging grounds but are then exposed to predation by cougars (*Puma concolor*) (*30*), which potentially results in an ecological "trap" in which species may do worse in an ostensibly beneficial situation.

Trophic interactions can form the basis of key ecosystem services such as pollination by insects. Nocturnal pollination is strongly affected by anthropogenic light, and reduced nocturnal pollination rates are compensated for by diurnal insects (*68*). The disturbance of trophic interactions by light at night in plant–herbivore communities can be complex and depend on species and color composition (*69*). To understand ecosystem-wide changes, a better understanding of trophic interactions is essential.

Anthropogenic light affects species at different time scales

The attraction of animals to light is a well-known direct response. Anyone who is attentive will notice the accumulation of insects around light sources. Attraction has been reported in many other species groups, such as marine turtles, birds, and amphibians, and often has direct consequences for survival (*8*, *12*, *17*) and subsequently for populations. Because the removal of light sources is very rare, it is important to assess effects that may only manifest after a long period of time. Several studies have shown indirect effects of anthropogenic light on many insect species (*9*). Adult fireflies survive light sources in proximity but fail to attract mates (*70*), which eventually results in population decline. Moth species, in addition to the agony of being trapped near light sources, are burdened by impaired reproduction and development (*38*). Light at night affects gonadal growth in blackbirds, which manifests only during the second year of exposure (*71*). Ultimately, light at night may cause adaptive changes, but evidence is limited to just a few studies (*9*). These include the examples of potential adaptation in spiders (*Larinioides sclopetarius*) with the innate preference for building webs around anthropogenic light sources (*40*) and in urban moths that show reduced flight-to-light behavior compared with rural conspecifics (*72*).

The way forward

Given the varied and substantial impacts of anthropogenic light, there are no simple solutions for its reduction and mitigation in natural systems. Continued increases in the nocturnal use of anthropogenic lighting will exacerbate its impact on our natural environment (Figs. 1C and 4, A to E), causing major changes to ecosystems, such as further declines in insect populations, loss of habitats of nocturnal mammals, and disruption of food web interactions. These changes will lead to the loss of biodiversity and potential feedback effects, including impaired ecosystem services such as the pollination of crops. Keeping natural areas dark, limiting light emissions from near and distant light sources, and thereby reducing sky brightness are therefore of utmost importance. Strategies to prevent light emissions into nature conservation areas are therefore proposed to be urgently implemented to ensure the long-term survival of protected species, preferably by combinations of different light mitigation measures (*53*, *73*). To effectively implement these measures, both national and international collaborations are essential. In protected areas, the aim should be to reduce light emissions from anthropogenic light sources to natural nocturnal light conditions. Current international guidelines, such as the "Guide on the limitation of the effects of obtrusive light from outdoor lighting installations" (*74*), can be used as the first step in ensuring that light emissions are below the recommended threshold values in protected areas.

Several technical and practical adaptations of lighting designs can be used to reduce their effects on local ecosystems (*53*, *73*). These include luminaire shielding, time-restricted and adaptive lighting, light intensity reductions, and tuning color composition.

Luminaire shielding or special optics can be used to prevent light being emitted in unwanted directions (*53*). Standardization to restrict spill light is urgently needed to prevent the negative effects of upward light and light outside of intentionally lit areas in vulnerable and exposed ecosystems such as open and aquatic environments (Fig. 4, A and B). Natural barriers, such as dense vegetation, can also be used to hinder light emission. The ecological benefit of time-restricted lighting may be limited, because the human demand for light during the first part of the night coincides with the peak activity of many nocturnal species. For example, bats are particularly active after dusk, reducing the potency of part-night lighting schedules (*75*). Adaptive road lighting is a promising solution to reduce ecological effects, but it is most effective for roads with low traffic. However, dedicated lighting schedules in more unique situations can be highly effective, for example, the intermittent off-switching of the lighting setup at the National 9/11 Museum's "Tribute in Light" in lower Manhattan to release thousands of light-trapped migratory birds (*17*). Decreasing light intensity is essential to preventing its ecological effects because many species are highly photosensitive even at extremely low illuminance levels due to their adaptation to dark nocturnal ecosystems. However, little is known about the intensity thresholds of many species, which may vary according to exposure duration, life history stages, and habitat structure. The installation of dimmable LED lighting is therefore very important because light levels can be adjusted after installation. A good example is the city of Rotterdam in the Netherlands, where all 100,000 light posts are currently fitted with LED fixtures that can be remotely adjusted for light intensity at any time of night. Because of the problematic identification of threshold light intensity levels, a pragmatic approach may be to keep light levels below lower moonlight illumination levels, which range from 0.05 to 0.1 lux (*76*). Finally, the color composition of light sources must be carefully chosen because it can modulate ecological

Fig. 3. Light at night drives species interactions. Many species are nocturnal because of fear of predators, for example by barn owls (*Tyto alba*). The middle and right species (Natterer's bat, *Myotis nattereri*, and the brown long-eared bat, *Plecotus auritus*) are examples of slow-flying bats that need to be extra wary of predators. These bats thrive in darkness, emerging late in the evening and foraging in sheltered places. The common pipistrelle (*Pipistrellus pipistrellus*) avoids light as well, but has a more agile flight and dares to forage on accumulated insects close to street lights.

ILLUSTRATION: KRISTOFFER NIKLASSON/*SCIENCE*. BASED ON ANNIKA K. JÄGERBRAND AND KAMIEL SPOELSTRA

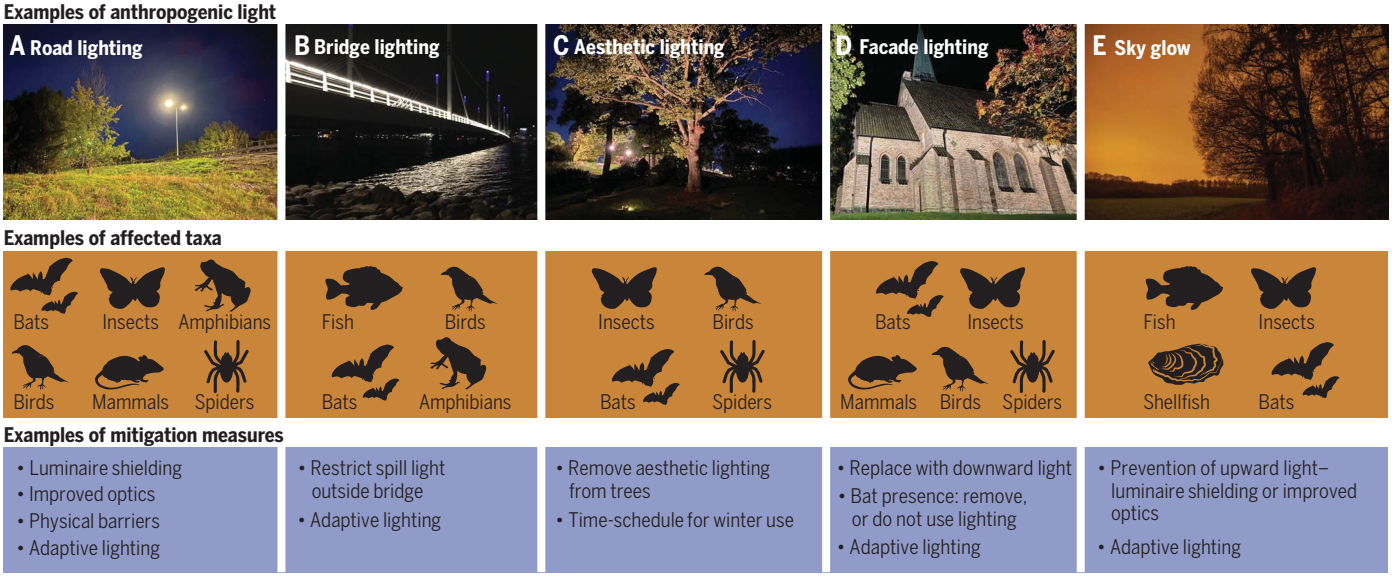


Fig. 4. Examples of anthropogenic light, affected taxa, and mitigation measures. (**A**) Terrestrial species can be affected by road lighting, which can be reduced through various measures to control light spill. (**B**) Bridge lighting can affect both aquatic and terrestrial species, and light spill outside the bridge should therefore be restricted. (**C**) Aesthetic lighting on trees can affect species that use trees for habitats or foraging, so this lighting should be removed or time restricted. (**D**) Upward façade lighting on buildings can affect several terrestrial species, and downward lighting should be used instead unless bats are present. (**E**) Even low levels of sky glow—reflected light from remote anthropogenic light sources—can affect species in natural habitat.

consequences. The current worldwide transition to LED lighting poses both challenges and opportunities. White LED lamps often contain a high proportion of blue light (~450 nm) and, although the responses of species are diverse, there is a general tendency to caution against the high emission of blue light. By utilizing diodes that produce different colors, and with the application of phosphor conversion techniques, the amount of blue light in a light source can be reduced. This is effective in protecting animal species because the red part of the color spectrum attracts fewer insects (*36*) and has a less disturbing effect on the activity of bats (*63*). Highly adapted spectra should, however, be used cautiously because they may create ecological traps for species that are unable to sense the light and therefore think they are in a safe environment. In addition, technical and practical lighting design adaptations presumably need adjustments to better compensate for variations in ecosystem properties and topography.

International initiatives have been formed to establish policy frameworks aimed at reducing light pollution, such as the Convention on the Conservation of Migratory Species of Wild Animals (CMS). Work is underway in the International Commission on Illumination (Commission Internationale de l'Eclairage, CIE) to develop guidelines for minimizing the effects of anthropogenic lighting on the natural environment. Furthermore, some countries have implemented national guidelines aimed at reducing ecological light pollution. Currently, outdoor lighting is not included in the 2030 Agenda for Sustainable Development, which has been adopted by all the members of the United Nations. This is very unfortunate because many Sustainable Development Goals are markedly affected by outdoor lighting and light pollution. For ecological sustainable development, it should be an urgent priority to keep naturally dark environments dark and to protect species from the adverse effects of anthropogenic light. This should be a strong motivator for national and local governments to include effects of anthropogenic lighting in planning and decision making toward a sustainable future.

Future recommendations for ecological protection should aim to establish numerical threshold values on the basis of illuminance or luminance of current light sources and technologies. This will facilitate the translation of new research findings into practical guidelines for lighting design and the upgrading of existing lighting systems. Given the variation in species responses to both spectral differences and intensities, a simple solution to mitigating the effects of anthropogenic light on all species may be challenging. Therefore, investigating different types of mitigation measures and their effectiveness in different species and environments is an important mission for future research.

It is important to recognize that even when the most advanced current technologies are used, light can still spill into the natural environment and sky because of reflections from surfaces. This reflected light may affect aerial species and contribute to sky glow (Fig. 4E). However, guidelines for outdoor lighting or light pollution rarely address restrictions on reflected light. Given the difficulty of effectively reducing the ecological effects of anthropogenic light, it may be necessary to adopt more frugal lighting practices and, in some cases, turn off lights altogether, despite the potential discomfort to humans that this may cause.

REFERENCES AND NOTES

1. International Commission on Illumination, "International lighting vocabulary" (CIE S017/E:2020 ILV, ed. 2, CIE, 2020); https://cie.co.at/publications/ilv-international-lighting-vocabulary-2nd-edition-0.
2. T. Longcore, C. Rich, *Front. Ecol. Environ.* **2**, 191–198 (2004).
3. F. Falchi *et al.*, *Sci. Adv.* **2**, e1600377 (2016).
4. A. K. Jägerbrand, *Sustainability (Basel)* **7**, 1028–1063 (2015).
5. A. Bassi *et al.*, *Fish Fish.* **23**, 631–647 (2021).
6. D. Sanders, E. Frago, R. Kehoe, C. Patterson, K. J. Gaston, *Nat. Ecol. Evol.* **5**, 74–81 (2021).
7. E. L. Stone, S. Harris, G. Jones, *Mamm. Biol.* **80**, 213–219 (2015).
8. A. C. S. Owens *et al.*, *Biol. Conserv.* **241**, 108259 (2020).
9. E. Desouhant, E. Gomes, N. Mondy, I. Amat, *Entomol. Exp. Appl.* **167**, 37–58 (2019).
10. D. H. Boyes, D. M. Evans, R. Fox, M. S. Parsons, M. J. O. Pocock, *Insect Conserv. Divers.* **14**, 167–187 (2021).
11. A. C. S. Owens, S. M. Lewis, *R. Soc. Open Sci.* **9**, 220468 (2022).
12. A. Rodríguez *et al.*, *Conserv. Biol.* **31**, 986–1001 (2017).
13. M. Grubisic *et al.*, *Sustainability* **11**, 6400 (2019).
14. L. F. B. Marangoni *et al.*, *Glob. Change Biol.* **28**, 5346–5367 (2022).
15. K. D. Lynn, P. A. Quijón, *Front. Ecol. Evol.* **10**, 863 (2022).
16. M. J. Zapata, S. M. P. Sullivan, S. M. Gray, *Estuaries Coasts* **42**, 309–330 (2019).
17. B. M. Van Doren *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11175–11180 (2017).
18. J. D. McLaren *et al.*, *Ecol. Lett.* **21**, 356–364 (2018).
19. S. A. Cabrera-Cruz, J. A. Smolinsky, J. J. Buler, *Sci. Rep.* **8**, 3261 (2018).
20. J. Q. Ouyang *et al.*, *Glob. Change Biol.* **23**, 4987–4994 (2017).
21. S. J. van Hasselt *et al.*, *Environ. Pollut.* **273**, 116444 (2021).
22. D. M. Dominoni, J. Kjellberg Jensen, M. de Jong, M. E. Visser, K. Spoelstra, *Ecol. Appl.* **30**, e02062 (2020).
23. A. Da Silva, M. Valcu, B. Kempenaers, *Anim. Behav.* **117**, 155–165 (2016).
24. G. Jones, J. Rydell, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **346**, 445–455 (1994).
25. D. M. Shier, A. K. Bird, T. B. Wang, *Environ. Pollut.* **263**, 114566 (2020).
26. L. Gordigiani *et al.*, *Mammal Res.* **67**, 39–49 (2022).
27. K. L. G. Russart, R. J. Nelson, *J. Exp. Zool. A Ecol. Integr. Physiol.* **329**, 401–408 (2018).
28. J. Hoffmann, R. Palme, J. A. Eccard, *Environ. Pollut.* **238**, 844–851 (2018).

29. H. Vardi-Naim, A. Benjamin, T. Sagiv, N. Kronfeld-Schor, Fitness consequences of chronic exposure to different light pollution wavelengths in nocturnal and diurnal rodents. Research Square [Preprint] (2022); https://doi.org/10.21203/rs.3.rs-1538025/v1.
30. M. A. Ditmer *et al*., *Ecography* **44**, 149–161 (2021).
31. L. L. Bliss-Ketchum, C. E. de Rivera, B. C. Turner, D. M. Weisbaum, *Biol. Conserv.* **199**, 25–28 (2016).
32. A. M. Dimovski, K. A. Robert, *J. Exp. Zool. A Ecol. Integr. Physiol.* **329**, 497–505 (2018).
33. T. Longcore *et al*., *Philos. Trans. R. Soc. London B Biol. Sci.* **370**, 20140125 (2015).
34. T. Degen *et al*., *J. Anim. Ecol.* **85**, 1352–1360 (2016).
35. K. J. Haynes, B. A. Robertson, *Curr. Opin. Insect Sci.* **45**, 91–96 (2021).
36. M. Donners *et al*., *J. Exp. Zool. A Ecol. Integr. Physiol.* **329**, 434–440 (2018).
37. F. van Langevelde *et al*., *Glob. Change Biol.* **24**, 925–932 (2018).
38. D. H. Boyes, D. M. Evans, R. Fox, M. S. Parsons, M. J. O. Pocock, *Sci. Adv.* **7**, eabi8322 (2021).
39. J. J. Foster *et al*., *Curr. Biol.* **31**, 3935–3942.e3 (2021).
40. A. M. Heiling, *Behav. Ecol. Sociobiol.* **46**, 43–49 (1999).
41. R. H. A. van Grunsven, D. Jähnichen, M. Grubisic, F. Hölker, *J. Exp. Zool. A Ecol. Integr. Physiol.* **329**, 429–433 (2018).
42. T. W. Davies, J. Bennie, K. J. Gaston, *Biol. Lett.* **8**, 764–767 (2012).
43. M. Touzot *et al*., *Environ. Pollut.* **259**, 113883 (2020).
44. J. Baxter-Gilbert *et al*., *Austral Ecol.* **46**, 501–506 (2021).
45. L. A. Taylor *et al*., *Physiol. Behav.* **257**, 113992 (2022).
46. T. R. Nelson *et al*., *Ecosphere* **13**, e4261 (2022).
47. M. T. Lockett, R. Rasmussen, S. K. Arndt, G. R. Hopkins, T. M. Jones, *Environ. Pollut.* **310**, 119803 (2022).
48. L. Meng *et al*., *PNAS Nexus* **1**, pgac046 (2022).
49. R. Heinen, *Basic Appl. Ecol.* **57**, 146–158 (2021).
50. J. Krieg, *Sci. Rep.* **11**, 20642 (2021).
51. C. C. Veilleux, M. E. Cummings, *J. Exp. Biol.* **215**, 4085–4096 (2012).
52. E. J. Warrant, S. Johnsen, *Curr. Biol.* **23**, R990–R994 (2013).
53. A. K. Jägerbrand, C. A. Bouroussis, *Sustainability* **13**, 5991 (2021).
54. N. A. Gilbert *et al*., *Trends Ecol. Evol.* **38**, 324–336 (2022).
55. K. Spoelstra, I. Verhagen, D. Meijer, M. E. Visser, *Proc. Biol. Sci.* **285**, 20172751 (2018).
56. K. L. G. Russart, R. J. Nelson, *Physiol. Behav.* **190**, 82–89 (2017).
57. D. M. Dominoni *et al*., *Sci. Rep.* **12**, 1553 (2022).
58. D. M. Dominoni, E. O. Carmona-Wagner, M. Hofmann, B. Kranstauber, J. Partecke, *J. Anim. Ecol.* **83**, 681–692 (2014).
59. B. Luo *et al*., *Environ. Pollut.* **288**, 117765 (2021).
60. J. Bennie, T. W. Davies, D. Cruse, K. J. Gaston, *J. Ecol.* **104**, 611–620 (2016).
61. K. A. Robert, J. A. Lesku, J. Partecke, B. Chambers, *Proc. Biol. Sci.* **282**, 20151745 (2015).
62. L. R. Fyie, M. M. Gardiner, M. E. Meuti, *J. Insect Physiol.* **129**, 104194 (2021).
63. K. Spoelstra *et al*., *Proc. Biol. Sci.* **284**, 20170075 (2017).
64. K. Barré *et al*., *Mov. Ecol.* **9**, 3 (2021).
65. A. Rodríguez, P. M. Orozco-Valor, J. H. Sarasola, *Landsc. Ecol.* **36**, 17–27 (2021).
66. F.-S. Zhang *et al*., *Sci. Total Environ.* **724**, 138271 (2020).
67. M. Czarnecka, T. Kakareko, Ł. Jermacz, R. Pawlak, J. Kobak, *Sci. Total Environ.* **684**, 14–22 (2019).
68. E. Knop *et al*., *Nature* **548**, 206–209 (2017).
69. D. Sanders *et al*., *Sci. Total Environ.* **831**, 154893 (2022).
70. C. Elgert, J. Hopkins, A. Kaitala, U. Candolin, *Proc. Biol. Sci.* **287**, 20200806 (2020).
71. D. M. Dominoni, M. Quetting, J. Partecke, *PLOS ONE* **8**, e85069 (2013).
72. F. Altermatt, D. Ebert, *Biol. Lett.* **12**, 20160111 (2016).
73. A. K. Jägerbrand, C. A. Bouroussis, in *Midterm Meeting & Conference Commission Internationale de l'Eclairage* (Commission Internationale de l'Eclairage, 2021), pp. 1–10.
74. International Commission on Illumination, "Guide on the limitation of the effects of obtrusive light from outdoor lighting installations" (CIE 150:2017, ed. 2, CIE, 2017); https://cms.cnc-cie.ca/images/documents/PressRelease/CIE150_2017.pdf.
75. C. Azam *et al*., *Glob. Change Biol.* **21**, 4333–4341 (2015).
76. C. C. M. Kyba, A. Mohar, T. Posch, *Astron. Geophys.* **58**, 1.31–1.32 (2017).

ACKNOWLEDGMENTS

We thank E. Jägerbrand for valuable help with Fig. 1. **Funding:** The work by K.S. is funded by the research program Applied and Engineering Sciences (AES), which is financed by the Dutch Research Council (NWO). **Competing interests:** The authors declare no competing interests.

Submitted 14 February 2023; accepted 5 May 2023
10.1126/science.adg3173

REVIEW

Reducing nighttime light exposure in the urban environment to benefit human health and society

K. M. Zielinska-Dabkowska[1]*†, E. S. Schernhammer[2,3,4]†, J. P. Hanifin[5]†, G. C. Brainard[5]†

Nocturnal light pollution can have profound effects on humans and other organisms. Recent research indicates that nighttime outdoor lighting is increasing rapidly. Evidence from controlled laboratory studies demonstrates that nocturnal light exposure can strain the visual system, disrupt circadian physiology, suppress melatonin secretion, and impair sleep. There is a growing body of work pointing to adverse effects of outdoor lighting on human health, including the risk of chronic diseases, but this knowledge is in a more nascent stage. In this Review, we synthesize recent research on the context-specific factors and physiology relevant to nocturnal light exposure in relation to human health and society, identify critical areas for future research, and highlight recent policy steps and recommendations for mitigating light pollution in the urban environment.

Nighttime outdoor lighting provides many important functions in urban societies including aiding wayfinding, supporting feelings of safety and security, enabling outdoor activities, revealing historical and architectural heritage, and promoting economic development by bringing people together (*1*). At the same time, cycles of light and darkness are important physiological cues for key bodily functions. Exposure to too much light in the evening and nighttime has the potential to disrupt circadian physiology, suppress melatonin secretion, impair sleep, and stress the visual system (*2*).

With 4.4 billion humans currently living in cities, nighttime lighting is rising worldwide, especially in urban environments, and it is now visible even in remote locations. A recent study shows that the extent of electrical light at night has increased by almost 10% annually over the past 12 years—much more than previously thought (*3*). It is known that electrical light from local and more distant sources produces light pollution, i.e., outdoor light levels that differ from those occurring naturally. This results in over-illumination, urban sky-glow, light trespass, and glare, all of which have been linked to adverse effects on various living organisms, including humans (*2*, *4*).

The Commission Internationale de L'Eclairage (CIE, or International Commission on Illumination), an organization that sets standards for the lighting industry and lighting professionals, describes light pollution as "the sum total of all adverse effects of artificial light" (*5*). Effects on flora, fauna, human health, and society comprise this "sum total," so a more comprehensive definition may be warranted. Light emitted by any type of electrical light source has the potential to contribute to light pollution. The introduction of light-emitting diodes (LEDs) in the early 2000s continues to transform outdoor lighting, as traditional high-pressure sodium and metal halide lamps for street and area lighting are changed to LED light sources, because of their promised energy-saving potential and longevity. As with earlier light sources, the overuse of lighting, as well as poor luminaire design and placement, can elicit unwanted adverse effects.

"...nighttime light can disrupt circadian rhythms..."

In this Review, we synthesize recent research on the context-specific factors and physiology relevant to nocturnal light exposure and their impact on human health and society. We also identify critical areas for future research and highlight recent policy steps and recommendations for mitigating light pollution to benefit human health and well-being.

Nighttime light exposure and human health

Demographic shifts toward increasingly urban living mean that most humans are exposed to higher levels of light at night. Behaviorally, humans have considerably extended their daytime and consequently shortened their nighttime. We experience increased nighttime light exposure from lighting in the home before bedtime, from computers, mobile phones, and television screens, as well as from light entering the home from street lighting, security lighting, externally illuminated buildings, and brightly illuminated advertisements. Excessive light exposure in the evening and nighttime can be problematic as it strains the eyes, stresses the visual system, and can cause circadian, neurobehavioral, and neuroendocrine problems such as circadian desynchrony, sleep disruption, and suppression of melatonin secretion from the pineal gland (*2*, *6*–*8*).

Light stimuli are detected by two sen… systems in the eye (Fig. 1A). Classically, hu… light perception has been thought to be mediated exclusively by rod and cone photoreceptors. These form the entire photoreceptor layer of the retina and detect visible light to support the sensory capacity of vision and visual reflexes (Fig. 1A). In addition, there is a set of intrinsically photosensitive retinal ganglion cells (ipRGCs), which contain the photopigment melanopsin (*9*, *10*). The ipRGCs play roles in visual responses such as contrast detection, and, just as rods and cones, can contribute to physiological responses (*6*, *9*). Numerous human and animal laboratory studies have quantified the impact of light on circadian phase shifting, melatonin regulation, and pupillary light reflex, as well as sleep, alertness, and mood responses (*6*, *9*–*12*). Recently, a study showed that melatonin suppression and circadian phase shifting appear to have a strong cone input at the start of a 6.5-hour light exposure, but over most of the exposure, ipRGCs are the primary photoreceptors for circadian and neuroendocrine light detection (*13*).

Early research on human exposure to light at night came from individual-level studies of night workers. This work showed that nighttime light can disrupt circadian rhythms, thereby increasing breast cancer risk (*14*). Today, the effects of individual nighttime light exposure on the circadian system and physiology central to chronic disease and the immune system are well demonstrated (*15*–*17*). Indeed, in nighttime workers, exposure to light at night has been described not only as a likely cause of cancer, but also as a risk factor for cardiovascular disease, type 2 diabetes, hypertension, obesity, depression, and more (*8*). Additionally, evidence reveals more immediate adverse effects of night work, ranging from sleep problems to work errors and injuries.

Researchers have attempted to identify health impacts of ambient outdoor light pollution. Images of nighttime light emissions (the extent of sky glow and overlighting) in the outdoor environment can be obtained through satellite-based remote sensing (Fig. 2), and this has been used in some epidemiological studies. Most of these studies on light pollution from outside the home are either purely ecological or multilevel, when prospective cohorts with individual-level data are investigated but light exposure is used as an aggregate measure. These studies have yielded mixed findings, with some showing a higher risk of breast cancer, obesity, diabetes, cardiovascular diseases, and sleep problems or disorders (*15*, *17*–*19*). Support is also still mixed about the physiological effects of excess ambient nighttime light exposure on melatonin levels, perhaps in part due to inherent variability in individuals' sensitivity to light (*4*, *20*, *21*).

Unlike night worker studies, in which metrics capturing the intensity and duration of night work exposure tend to provide a more viable estimate reflective of circadian misalignment, light

[1]GUT LightLab Department of Architecture, Gdańsk University of Technology, 80-233 Gdańsk, Poland. [2]Department of Epidemiology, Center for Public Health, Medical University of Vienna, Vienna, 1090, Austria. [3]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. [4]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA. [5]Department of Neurology, Thomas Jefferson University, Philadelphia, PA 19107, USA.
*Corresponding author. Email: k.zielinska-dabkowska@pg.edu.pl
†These authors contributed equally to this work.

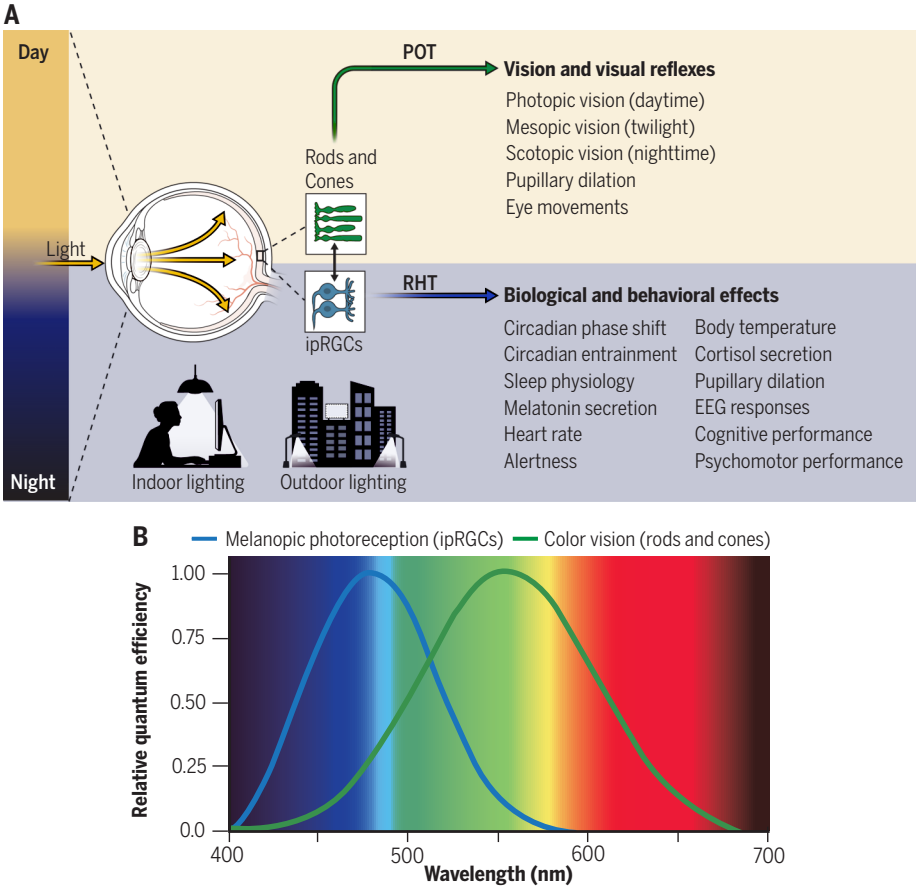


Fig. 1. The human eye has two distinct, interconnected sensory systems. (**A**) Simplified diagram illustrating that the human eye supports two distinct photoreceptive pathways. The primary optic tract (POT) delivers information about environmental light to regions of the brain involved in vision and visual reflexes. The retinohypothalamic tract (RHT) delivers information about light and darkness to nuclei in the brain that regulate circadian, neuroendocrine, and neurobehavioral effects of light. ipRGCs have the capacity to influence certain visual responses, and rods and cones have the ability to contribute to biological and behavioral responses. The list of effects is not comprehensive. (**B**). Graph illustrating the relative quantum efficiency of the photopic visual system for color vision (with a peak at ~555 nm) versus the melanopsin ipRGC system (with a peak at ~480 nm), which supports circadian, neuroendocrine, and neurobehavioral regulation in humans.

pollution studies using light exposure as an aggregate measure are currently fraught by their broad and imprecise measurements. For example, it has remained difficult to ascertain if the nighttime light pollution measured outdoors results in consistent lighting conditions in the bedrooms of exposed individuals, considering that individuals vary in their use of shutters and curtains to block light, additional indoor lighting exposure, and whether and when they sleep (*22*). It is challenging to obtain composite assessments of evening and nighttime interior lighting across a whole community, and only a few examples exist in which indoor nighttime light measurements have been attempted using, e.g., self-reported ambient light levels in private bedrooms while sleeping or portable light meters, linking this data with breast cancer and depressive symptoms (*23*, *24*).

Furthermore, some elements of light pollution cannot be quantified with satellite data owing to lack of resolution, failure to detect shorter wavelengths of light, complex three-dimensional structure of light sources, and cloud cover in the Northern Hemisphere (*25*). More recent studies of nighttime light pollution have made a greater distinction between indoor and outdoor lighting (*26*). They show that exposure to outdoor light at night with light rich in the blue-appearing spectrum is associated with increased breast, prostate, and colon cancer risk. Future empirical studies on the physiological and health effects of outdoor nighttime lighting should incorporate these considerations.

The direct impact of outdoor lighting has rarely been measured. One study with a small sample size examined car drivers, pedestrians, and people exposed to light trespass from streetlamps. Each group was exposed to solid-state light sources with differing spectral emissions. The spectrum of the light sources tested at recommended roadway exposure levels did not statistically affect salivary melatonin levels (*20*). In another study, scientists measured light exposure and wrist temperature among residents of an urban area and noted circadian effects on wrist temperature linked to the timing of environmental light exposure, although the contribution of outdoor light sources after dark was minimal (*27*). Glare from bright light sources such as unshielded LEDs is a growing public concern (*7*). Recently, a new approach for evaluating the discomfort glare of pedestrian-scale applications has been developed (*28*). A more in-depth study of temporal light modulation (flicker) at frequencies that may have adverse health effects is needed (*29*). There remain sparse direct, empirical measurements of the capacity of urban street, building, outdoor sport facility, and illuminated advertisement lighting to evoke detrimental visual, circadian, neuroendocrine, or neurobehavioral responses in healthy humans under naturalistic outdoor conditions.

The physiological effects of light exposure have been shown to be dependent on light spectrum, quantity, length of exposure, and the time of day (*6*, *9*, *15*). In addition, differences between individuals influence their susceptibility to developing disease when exposed (i.e., variation in "shift work tolerance" in night workers). Important determinants of such variability include age, gender, chronotype, and ethnicity (*15*). Although evidence is currently inconclusive, there is some suggestion that there are stronger effects of light on sleep in younger persons. An important recent study showed that a >50-fold variability exists between individuals regarding their sensitivity to evening light, suggesting that future studies should take interindividual differences into account (*30*).

Recently, other potentially relevant effects of light on human physiology have been described. These relate to daily oscillations of the gut microbiome, the relative abundance of gut microbes under normal conditions, and altered composition under modified dark-cycle conditions, including dim light at night (*31*). It is increasingly recognized that gut microbiota are involved in human physiology, affecting metabolism, neurodegeneration, and mental health, and interventions that modulate gut microbiota may mitigate the impact of light exposure on health outcomes associated with nighttime light (*31*). During the COVID-19 pandemic, studies showed that people became infected more often, more severely, and for longer periods of time when they were sleep deprived and/or worked at night or they resided in areas that had high levels of outdoor electrical light emissions at night (*32*, *33*).

Owing to the pronounced effects of nocturnal light on flora and fauna, indirect effects on human health, although even more difficult to measure, are practically inevitable (*34*). Urban light at night can attract and repel animals

and cause changes in their behavior, potentially providing more opportunities for vector-borne diseases to spread to humans, including through zoonoses. In addition, light at night has been implicated in the recently observed insect die-off, because of decreased reproduction and a fatal attraction to car headlights, streetlights, or other lighting (*35*). Insect decline has numerous downstream effects that are detrimental to human health, as insects control pest species, help decompose waste, and pollinate more than a third of our food (*35*).

Typically, light that supports human vision in daylight is quantified inside and outside of buildings in terms of photopic illuminance or lux. Information about light and darkness is also detected by photosensitive ipRGCs and transmitted to nuclei in the brain that regulate circadian, neuroendocrine, and neurobehavioral effects (Fig. 1A). Quantifying light that regulates human physiology, however, requires a different method of measurement.

Scientific consensus and an international standard, balloted by the CIE, selected the measurement of melanopic equivalent daytime illuminance (melanopic EDI) at eye level as a better means than photopic lux for quantifying light relative to the regulation of human physiology, and ultimately, health and well-being (*6*, *9*, *36*). Because this is an emergent field and the relative contribution of rods, cones, and ipRGCs remains unknown, the CIE recommends assessing and reporting all five α-opic photoreceptor values (*36*). Different spectral sensitivities of the photopic visual system versus the melanopic ipRGC photoreceptor have been demonstrated (Fig. 1B).

Social burden of light pollution

Understanding the drivers and consequences of light pollution is critical to moving society toward a humanity-centered (supporting the rights of all of humanity and the entire ecosystem) and sustainable lighting approach. Although there are numerous needs for nighttime lighting, some social and cultural misconceptions can contribute to poorly executed lighting, which, in turn, can lead to unnecessary light pollution.

Such misconceptions include the notion that increased electric light outdoors is seen as a sign of wealth and prosperity. Economic growth, for example, has been estimated through satellite imagery of Earth at night to record visible illumination and correlated to gross domestic product (GDP), even though most relevant activity actually occurs during the day, thereby reducing the accuracy of these estimates (*37*).

The fear of darkness and the associated perception that bright lighting provides safety and security have limited research support (*38*). In fact, unshielded, glary lighting can be momentarily blinding, making it more difficult to identify objects and potential threats, navigate the terrain, and adjust to low-contrast surrounding, consequently decreasing safety (*7*).

An absence of lighting, as well as poorly executed lighting, also can be a sign of social injustice in terms of wealth prospects and opportunities (*39*). In the rural populations of developing, low-income countries worldwide, there is no electricity in most schools and households, but without education it is difficult to escape poverty. Basic lighting relies on unsustainable sources of energy, e.g., wood, coal, or kerosene, and these can negatively affect wide-ranging aspects of children's health, such as respiratory problems or chemically related pneumonia from kerosene lamps, eye disease, or accidents from using candles while learning in bed. To avoid this, children often gather in publicly lit areas with streetlights to study and do homework. By contrast, in the Western world, the introduction of lower illuminance levels is seen as a sign of commitment to environmental protection and sustainability (*40*, *41*). In recent decades, affluent citizens (who generally live in safe neighborhoods) have enjoyed more natural darkness and considerably less light pollution from well-designed lighting on their own properties or in nearby natural park areas, compared to people in less affluent suburbs. Indeed, studies have emerged that highlight light pollution inequities, with more disparate exposure to light at night based on race and socioeconomic status (*42*). Moreover, in mixed-use urban developments that encompass two or more types of land uses (e.g., residential, commercial, retail), light trespass from outdoor lighting might include street lighting, media architecture, or LED advertisements shining through bedroom windows (*43*). This can disrupt sleep, requiring people to sleep in another room to escape light pollution, and in some cases, residents have even had to move to another apartment—an affordable option, however, only for a certain segment of the population.

"...there are better ways to manage light pollution for the benefit of public health and society in the urban environment..."

How a lack of night sky visibility might contribute to the lack of healthy urban living has been studied in the past (*44*). Urban residents have been disconnected from nature and the cosmos such as major constellations, planets, and the Moon, with the Milky Way being hidden from more than one-third of humanity (Fig. 2) (*45*). Other studies also indicated that stargazing can be a stress reliever, possibly reducing anxiety and depression (*46*).

Another concern is that light pollution has increasingly obscured the celestial culture of Indigenous communities, many of whom rely on visual celestial clues that are supported by stories passed down across generations (*47*). Some have even expressed concern that the eradication of starlore and the disconnection from the heavens as a result of light pollution could contribute to the decline of entire groups by impeding their history and culture (*47*).

Every citizen should have the right to access darkness and quality, responsible outdoor lighting for improved health and well-being, a better social life, and nocturnal sustainability (*48*). To help facilitate positive change, more empirical evidence is required to determine the amount, and also the quality of outdoor lighting necessary to ensure visibility, safety, and the confidence of end users (*49*).

Current actions and recommendations for managing and regulating light pollution

Countries such as the Czech Republic, France, Germany, South Korea, and Slovenia are taking regulatory and policy measures against light pollution by adopting laws, regulations, and action plans to protect human health, the environment, and biodiversity (*50*). Despite such encouraging initial steps, most countries have no regulatory infrastructure to monitor or regulate light emissions at night.

Furthermore, there are currently no recommendations for urban outdoor lighting standards that specifically address circadian, neuroendocrine, and neurobehavioral effects of light, nor are there properly developed frameworks for assessing the health effects of light pollution (*40*). A scientific consensus position published in 2014, which evolved into an internationally balloted CIE standard in 2018, provides guidance to the lighting and scientific communities on how best to quantify light exposure in an SI-compliant system for regulating human physiology and ultimately health (*6*, *36*). Based, in part, on an analysis of 19 published studies, a consensus view was published in 2022, with target lighting recommendations of melanopic EDI levels for indoor ocular light exposures measured at the eye, to best support physiology, sleep, and wakefulness in day-active healthy adults (*9*, *51*). Specific melanopic EDI values were given for daytime minimum amounts, as well as evening and nighttime maximum exposures (*9*). These last two elements are relevant for outdoor electrical lighting at night. Studies on the effects of outdoor nighttime light exposure on human health would be best served by characterizing the spectral power distribution (SPD) of the light and all five α-opic EDI values for rods, cones and ipRGCs (*6*, *9*, *36*). Although several metrics for quantifying outdoor lighting exist (e.g., photon flux, irradiance, luminance, photopic and scotopic illuminance, and sky brightness factor), to understand whether and how light pollution affects human neurophysiology, it is important to measure all five α-opic

EDI values. The published recommendations for evening and nighttime light exposures described above could be considered as preliminary guidance for limiting outdoor light trespass into residences or other indoor spaces (*9*). Notably, studies with nocturnal rodents show circadian and neuroendocrine sensitivities well below 1 melanopic lux, and there are indications that the same may be true for humans (*52*).

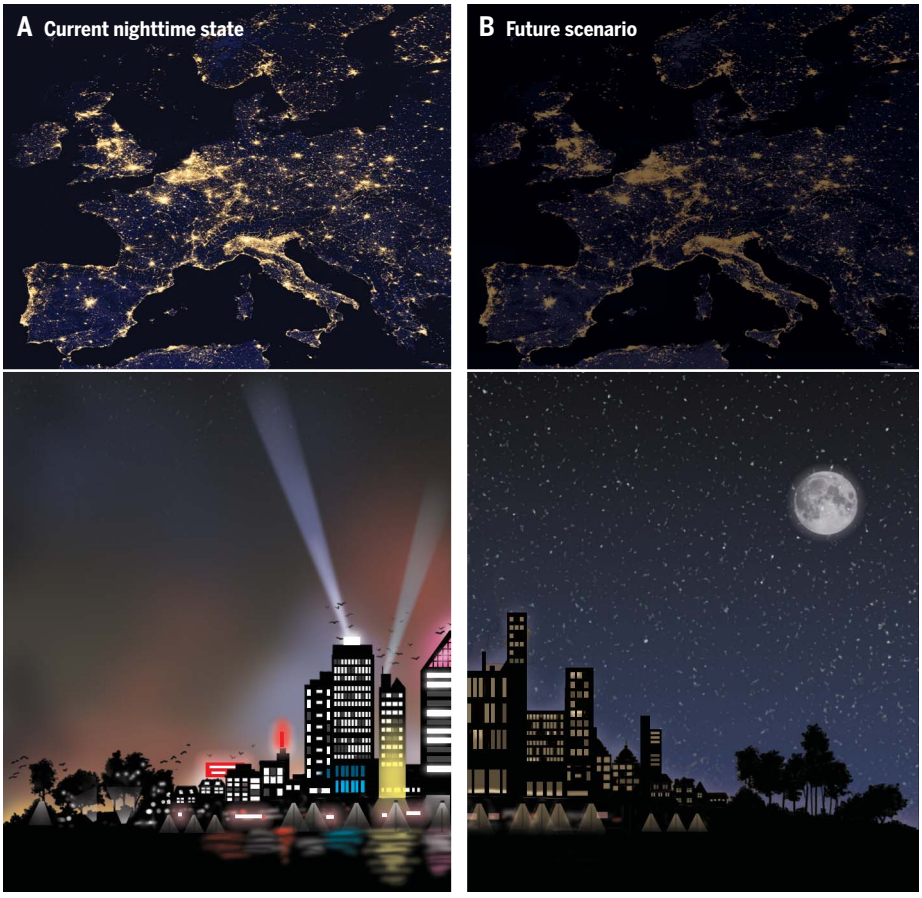
Rather than providing assistance in establishing plans or restrictions for electric light in terms of quantity, spectrum, timing, geometry, and the duration of exposure, current guidelines instead focus on energy efficiency. In fact, there are better ways to manage light pollution for the benefit of public health and society in the urban environment, including the driver-pressure-state-impact-response (DPSIR) framework, which is based on work of the European Environment Agency (Fig. 3). It provides a structure of the indicators needed to enable feedback for policy-makers on light pollution, public health, and the resulting impact of current or future political choices.

Recent changes in the SPD of street lighting sources have sparked an intense debate over the possible ill effects of LED lighting. The American Medical Association (AMA) produced two reports: one that details the impacts of inappropriate exposure to excessive light at night with a call for more research, and another, more recent document that recommends the use of roadway lighting with lower correlated color temperature lamps (<3000 K, which emit fewer short wavelengths of light) to minimize potential harmful effects to human health (*53*, *54*). The more recent AMA recommendation resulted in a strong reaction from the lighting industry and the US government, which disagreed with elements of that policy (*55*). Several initiatives, measures, and policies have subsequently evolved as a means to prevent the excessive use of electrical light at night, and many are still being enacted and developed. Data-driven design recommendations have been developed to match illumination levels to the needs of vehicle users, reducing extraneous light while maintaining a level of safety for pedestrians and residents (*56*).

Because of the energy crisis, many cities have begun to change their energy policies and introduce lighting curfews. Even if saving energy is the main driver behind the switch-off of lighting rather than the goal of minimizing the other impacts of exposure to light, these actions are a positive example for other governments to follow. CIE recommendations to limit light trespass on properties (e.g., windows) exist; however, these may not be effective enough. For maximum values of vertical illuminance including the sum of all area lighting installations in well-inhabited urban and rural settlements (medium district brightness, environmental lighting zone E3), there is an allowance of 10 lux (pre-curfew) and 2 lux (post-curfew) (*41*). For town and city centers including other commercial areas (high district brightness, lighting zone E4), there is an allowance of 25 lux (pre-curfew) and 5 lux (post-curfew) (*41*). Additionally, an evidence-based consensus should be developed on the maximum light levels allowed in outdoor urban residential areas for limiting light trespass.

With the many problems faced by the world's urban inhabitants, it is common for city authorities to wonder how to improve the quality of life and health of citizens. Human-generated light pollution, which affects people in the evening and at night, however, has often been omitted even in the forward-looking development strategies proposed by the World Health Organization, such as Healthy Cities (*57*). Therefore, proposed and implemented well-balanced responsible urban lighting strategies and masterplans that allow for the integration of dimming and/or the temporary switching off of street, area, and decorative lighting could improve the existing situation (*40*). The City of London's recently adopted lighting strategy is a step toward a new approach to urban lighting and nocturnal sustainability for the built environment. Prior to this, the brightness of decorative and advertisement illumination of buildings was not properly regulated in terms of the adverse impact of light trespass on the health and well-being of residents (*43*).

Contrary to widespread outdoor light pollution, which a single person cannot change, overlighting can easily be reduced and prevented in the home. Individuals can take action by either changing their light sources—e.g., to amber or warm white (with fewer short wavelengths light)—or by using only red light at night (with zero short wavelengths of light), or by turning lights off or dimming them in the evening using



CREDIT: K.HOLOSKI/*SCIENCE* BASED ON KAROLINA M. ZIELINSKA-DABKOWSKA AND SEREN DINCEL

Fig. 2. Exposure to excess light at night has the capacity to influence human health, other organisms, and ecological systems. (**A**) Image of Europe from space at the current nighttime state shows increased light pollution typical of many urban areas across the globe. Sky glow above cities caused by over-illumination affects not only humans but also the migration and behavior of various animal species. Excess nighttime lighting can have an indirect impact on animals and plants and may contribute to insect decline and the emergence of zoonoses. (**B**) A future scenario in which anthropogenic light pollution is reduced to improve human, animal, and environmental health. Protecting ecosystems in urban areas can reduce humanity's environmental footprint, improve quality of life, and provide a cultural connection to the night sky.

sensors and/or timers on their balconies, in gardens, and on façades. They can also become active members of various citizen groups, such as one advocating for a thorough investigation into the evidence of the impact of outdoor illumination (*58*). Recently, such citizens' actions resulted in an inquiry into the effects of electrical light on human health by the UK House of Lords (*59*).

In 2020, citizens, active members of the International Dark-Sky Association (IDA), and lighting professionals representing the Illuminating Engineering Society (IES) joined forces to protect the night from light pollution, and a consensus was reached between these stakeholders. This alliance created the Five Principles for Responsible Outdoor Lighting, which were adopted by the board of directors of both organizations (*60*). Another recent initiative, the Responsible Outdoor Lighting at Night (ROLAN) Manifesto, has further expanded and applied this knowledge (*61*). It provides 10 core principles for external illumination with easy-to-follow actions to minimize the impact of outdoor nighttime lighting on human health and well-being, and also to preserve darkness and humanity's access to the stars at night. It has been adopted by leading international professional lighting bodies such as the International Association of Lighting Designers (IALD) and the IES, which demonstrates that dark sky protection is starting to be accepted.

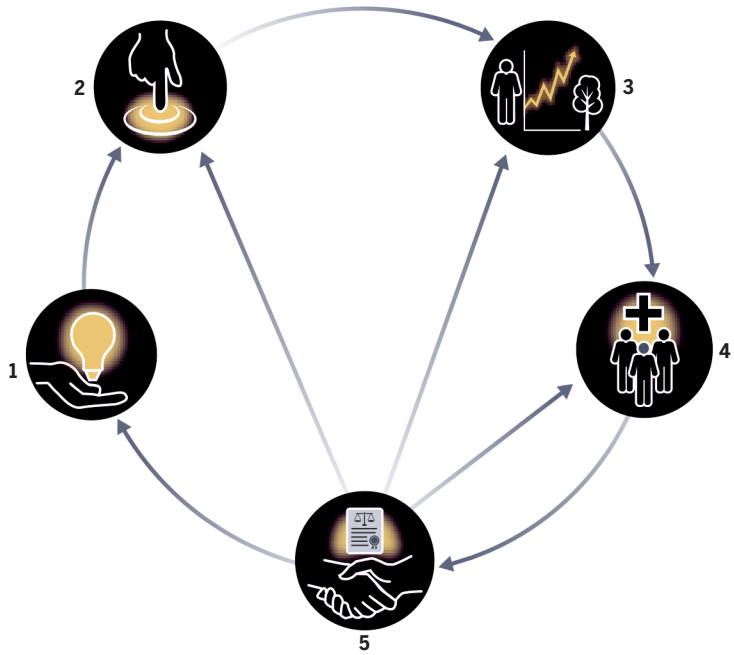
Effective reductions in light pollution ultimately require more comprehensive regulatory controls. Therefore, of utmost importance is that all concerned stakeholders understand the consequences of excess light exposure at night and work together to develop measures that standardize urban lighting.

Conclusion

Detailed scientific knowledge of how overexposure to light at night affects individuals on the physiological, population, and community levels is complex and incomplete. Further research is required to facilitate the prevention and management of light pollution and to develop best practices for safer and healthier nighttime outdoor illumination.

There is a critical need for field studies on the physiological impact of nighttime electrical outdoor light to validate the concepts discussed above. These studies should (i) achieve an improved characterization of the SPD of both indoor and outdoor light, and the melanopic EDI measured at eye level; (ii) identify the minimum levels of illuminance and luminance affecting human vision as well as other organisms and ecological systems; (iii) differentiate between the health consequences of excess nighttime light exposure, air pollution, and other environmental exposures; (iv) resolve factors that contribute to, or are a consequence of, individual variations in physiological responses to nighttime light such as sleep timing, the duration and timing of outdoor light exposure, and genetic variations; and (v) investigate health-related societal impacts of light pollution.

Furthermore, a collaboration between the research community and the lighting profession is essential so that relevant issues can be effectively identified to guide future research. Resultant findings can then be integrated into outdoor illumination projects, as well as into regulatory frameworks. In addition, it is essential to foster evidenced-based lighting design approaches and technological solutions including light sources, luminaire design, lighting controls, and measurement methods that support the health and well-being of humans and entire ecosystems, as well as Indigenous and cultural astronomy. It is hoped that this Review provides helpful guidance for lighting professionals and for addressing gaps in light pollution research, as well as for the organizations



1 ANTHROPOGENIC DRIVERS
Anthropogenic light pollution connected to human activities that influence environmental change and public health via:

- Over-illumination
- Urban sky glow
- Light trespass/light spill
- Glare

2 PRESSURES
Stressors that human activities (light pollution) place on the environment and public health

3 STATE
The state of the environment and socioeconomic status

The condition of the environment and public health and well-being

4 IMPACTS
Effects on public health and society

5 RESPONSES
Responses by society to the environmental/public health situation

- Light pollution laws, regulations and action plans
- Lighting impact assessments
- Outdoor lighting standards and guidelines
- Responsible urban lighting masterplans and lighting strategies
- Public inquiries/surveys
- Five Principles for Responsible Outdoor Lighting (IES/IDA)
- Responsible Outdoor Lighting at Night (ROLAN) Manifesto

Fig. 3. DPSIR framework for nighttime light exposure based on work by the European Environmental Agency, modified to benefit public health and society in the urban environment. This effective tool consists of five indicators that interreact with human activities in a closed loop of action, encouraging society or individuals to change their practices regarding the use of light at night. Appropriate responses (5) by society to anthropogenic drivers (1) such as light pollution can prevent or minimize their occurrence. Moreover, these behaviors can alleviate pressure (2) by reducing the extent of light pollution from excessive lighting, as these responses strive to understand the factors affecting this environmental change and impact on public health. By using appropriate tools such as public surveys that are based on expert interviews, testimony and evidence can be gathered along with adequate evidence of conditions so that the state (3) of the nighttime environment as well as public health and well-being can be partially restored to what it was before the use of electrical outdoor lighting. Lastly, impacts (4) in the form of public health and societal effects can be adapted to environmental conditions through the introduction of appropriate regulatory frameworks and the application of recommendations in the form of light pollution laws, lighting standards, guidelines, and other tools.

that set lighting standards. An approach that integrates the needs of all stakeholders opens the door to improve outdoor lighting to promote and support the long-term welfare of humanity.

REFERENCES AND NOTES

1. D. Casciani, *The Human and Social Dimension of Urban Lightscapes* (Springer Briefs in Applied Sciences and Technology, 2020).
2. R. M. Lunn *et al*., *Sci. Total Environ.* **607-608**, 1073–1084 (2017).
3. C. C. M. Kyba, Y. Ö. Altıntaş, C. E. Walker, M. Newhouse, *Science* **379**, 265–268 (2023).
4. M. Grubisic *et al*., *Sustainability* **11**, 6400 (2019).
5. CIE, E-ILV, Light pollution; https://cie.co.at/eilvterm/17-29-177 (accessed 20 February 2023).
6. R. J. Lucas *et al*., *Trends Neurosci.* **37**, 1–9 (2014).
7. S. Fotios, M. Kent, *Leukos* **17**, 338–358 (2021).
8. I. A. R. C. Monographs Vol, *Lancet Oncol.* **20**, 1058–1059 (2019).
9. T. M. Brown *et al*., *PLOS Biol.* **20**, e3001571 (2022).
10. J. Hannibal, A. T. Christiansen, S. Heegaard, J. Fahrenkrug, J. F. Kiilgaard, *J. Comp. Neurol.* **525**, 1934–1961 (2017).
11. D. C. Fernandez *et al*., *Cell* **175**, 71–84.e18 (2018).
12. L. S. Mure, *Front. Neurol.* **12**, 636330 (2021).
13. M. A. St Hilaire *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205301119 (2022).
14. E. S. Schernhammer *et al*., *J. Natl. Cancer Inst.* **93**, 1563–1568 (2001).
15. P. Razavi *et al*., *Cancer Epidemiol. Biomarkers Prev.* **28**, 1177–1186 (2019).
16. V. Jerigova, M. Zeman, M. Okuliarova, *Int. J. Mol. Sci.* **23**, 13722 (2022).
17. M. Cao, T. Xu, D. Yin, *J. Environ. Sci. (China)* **127**, 589–602 (2023).
18. M. Bożejko, I. Tarski, M. Małodobra-Mazur, *Environ. Res.* **218**, 115049 (2023).
19. W. R. Diver *et al*., *Int. J. Environ. Res. Public Health* **20**, 3658 (2023).
20. R. B. Gibbons, R. Bhagavathula, B. Warfield, G. C. Brainard, J. P. Hanifin, *Clocks Sleep* **4**, 633–657 (2022).
21. C. B. Swope *et al*., *Chronobiol. Int.* 10.1080/07420528.2023.2188091 (2023).
22. E. S. Schernhammer, K. L. Stone, *Chronobiol. Int.* **28**, 378–379, author reply 379–380 (2011).
23. S. Davis, D. K. Mirick, R. G. Stevens, *J. Natl. Cancer Inst.* **93**, 1557–1562 (2001).
24. K. Obayashi, K. Saeki, N. Kurumatani, *Am. J. Epidemiol.* **187**, 427–434 (2018).
25. S. Mander, F. Alam, R. Lovreglio, M. Ooi, *Sustain Cities Soc.* **92**, 104465 (2023).
26. A. Garcia-Saenz *et al*., *Epidemiology* **31**, 718–727 (2020).
27. D. Khodasevich *et al*., *Proc. Biol. Sci.* **288**, 20210721 (2021).
28. K. S. Hickcox, S. Fotios, B. Abboushi, N. J. Miller, *Light. Res. Technol.* **54**, 91–92 (2022).
29. S. Batra, C. S. Pandav, S. Ahuja, *J. Clin. Diagn. Res.* **13**, NE01–NE05 (2019).
30. A. J. K. Phillips *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12019–12024 (2019).
31. C. C. Lee *et al*., *EMBO Rep.* **23**, e52316 (2022).
32. L. Taylor *et al*., *iScience* **26**, 105877 (2023).
33. A. Argentiero, R. Cerqueti, M. Maggi, *Environ. Impact Assess. Rev.* **90**, 106602 (2021).
34. J. Falcón *et al*., *Front. Neurosci.* **14**, 602796 (2020).
35. A. C. Owens *et al*., *Biol. Conserv.* **241**, 108259 (2020).
36. CIE, *CIE System for Metrology of Optical Radiation for ipRGC-Influenced Responses to Light* (CIE, Vienna, Austria, 2018).
37. F. Bickenbach, E. Bode, P. Nunnenkamp, M. Söder, *Rev. World Econ.* **152**, 425–447 (2016).
38. R. Steinbach *et al*., *J. Epidemiol. Community Health* **69**, 1118–1124 (2015).
39. S. M. Nadybal, T. W. Collins, S. E. Grineski, *Environ. Res.* **189**, 109959 (2020).
40. K. M. Zielinska-Dabkowska, *Sustainability (Basel)* **14**, 14525 (2022).
41. International Commission on Illumination, *Guide on the Limitation of the Effects of Obtrusive Light from Outdoor Lighting Installations*, ed. 2; Technical Report CIE 150 (International Commission on Illumination, Vienna, Austria, 2017).
42. F. Crea, *Eur. Heart J.* **42**, 801–804 (2021).
43. K. M. Zielinska-Dabkowska, K. Xavia, *Sustainability* **11**, 3446 (2019).
44. T. Hartig, P. H. Kahn Jr., *Science* **352**, 938–940 (2016).
45. F. Falchi *et al*., *Sci. Adv.* **2**, e1600377 (2016).
46. R. Bell, K. N. Irvine, C. Wilson, S. L. Warber, *Eur. J. Ecopsychol.* **5**, 1–15 (2014).
47. A. Venkatesan, J. Lowenthal, P. Prem, M. Vidaurri, *Nat. Astron.* **4**, 1043–1048 (2020).
48. T. Stone, *Sci. Eng. Ethics* **24**, 607–628 (2018).
49. S. Fotios, R. Gibbons, *Light. Res. Technol.* **50**, 154–186 (2018).
50. C. Walker, P. Benvenuti, Dark and Quiet Skies II Working Group Reports, Zenodo (2022). https://doi.org/10.5281/zenodo.5874725.
51. T. M. Brown, *J. Pineal Res.* **69**, e12655 (2020).
52. T. J. Walbeek, E. M. Harrison, M. R. Gorman, G. L. Glickman, *Front. Neurol.* **12**, 625334 (2021).
53. American Medical Association, "Light pollution: Adverse health effects of nighttime lighting," CSAPH Report 4-A-12 (2012).
54. American Medical Association, "Human and Environmental Effects of Light Emitting Diode (LED) Community Lighting," CSAPH Report 2-A-16 (2016).
55. US Department of Energy, The Light Post—Official MSSLC e-Newsletter, July 2016; https://www.energy.gov/eere/ssl/articles/light-post-july-2016 (accessed 5 April 2023).
56. R. Gibbons *et al*., "Guidelines for the Implementation of Reduced Lighting on Roadways," FHWA-HRT-14-050 (2014); https://www.fhwa.dot.gov/publications/research/safety/14050/.
57. World Health Organization, "Healthy Cities"; https://www.who.int/southeastasia/activities/healthy-cities (accessed 6 May 2023).
58. K. M. Zielińska-Dabkowska, K. Xavia, K. Bobkowska, *Sustainability* **12**, 4997 (2020).
59. U. K. Parliament, "Call for Evidence"; https://committees.parliament.uk/call-for-evidence/3032 (accessed 10 February 2022).
60. International Dark-Sky Association, "Joining Forces to Protect the Night from Light Pollution"; https://www.darksky.org/joining-forces-to-protect-the-night-from-light-pollution (accessed 4 April 2023).
61. International Dark-Sky Association, "Responsible Outdoor Lighting at Night (ROLAN) manifesto for lighting"; https://www.darksky.org/responsible-outdoor-lighting-at-night-rolan-manifesto-for-lighting/ (accessed 20 February 2023).

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers of this paper for their helpful recommendations toward improving the manuscript. Additionally, the authors thank K. Pietrzak, B. Warfield. and S. Dincel for their contribution to the graphical content. **Funding:** K.M.Z.-D. was supported by The National Science Centre (NCN) grant 2020/04/X/NZ7/00307. E.S.S. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement no. 101053225. J.P.H. was supported, in part, by DOE grants DE-EE0008207 and DE-EE0009689; NAS Award HR 05-23 UNIT 905; NASA grant NNX15AC14G; the NSF ERC LESA Center; the Nova Institute, Biological Innovation and Optimization Systems, Toshiba Materials Science, and Seoul Semiconductor. G.C.B. was supported, in part, by DOE grants DE-EE0008207 and DE-EE0009689; NAS Award HR 05-23 UNIT 905; NASA grant NNX15AC14G; the NSF ERC LESA Center; and the Nova Institute, Toshiba Materials Science, and Seoul Semiconductor. **Author contributions:** Conceptualization: K.M.Z.-D., E.S.S., G.C.B., J.P.H. Methodology: K.M.Z.-D., E.S.S., G.C.B., J.P.H. Investigation: K.M.Z.-D., E.S.S., G.C.B., J.P.H. Visualization: K.M.Z.-D. (Figs. 1 to 3), G.C.B. (Fig. 1). Funding acquisition: K.M.Z.-D., E.S.S., G.C.B., J.P.H. Writing – original draft: K.M.Z.-D., E.S.S., G.C.B., J.P.H. Writing – review and editing: K.M.Z.-D., E.S.S., G.C.B., J.P.H. **Competing interests:** K.M.Z.-D. and E.S.S. declare no conflicts of interest relative to the scientific content of this manuscript. J.P.H. declares no conflicts of interest relative to the scientific content of this manuscript. He has been a paid consultant by Lutron, Inc. and McCullough Hill LLC. G.C.B. has no conflicts of interest relative to the scientific content of this manuscript. In the spirit of open disclosure, however, he reports having a current patent (USPTO 7678140 B2) related to the photoreceptor system for melatonin regulation. That intellectual property has been licensed by Litebook Company Ltd. He is a paid member of the scientific advisory board of PhotoPharmics Inc. He has been a paid consultant by Lutron, Inc. and McCullough Hill LLC. In addition, the Thomas Jefferson's Light Research Program (LRP) has received research grants and/or equipment donations from industry partners including Toshiba Materials, Seoul Semiconductor, BIOS, Robern, and the PhotoPharmics Company. The LRP has received philanthropic donations from The Philadelphia Chapter of the Illuminating Engineering Society, BIOS Inc., Robern, and Toshiba Materials for programmatic, research, and educational uses. **Data and materials availability:** All data are available in the main text.

Submitted 24 February 2023; accepted 9 May 2023
10.1126/science.adg5277

REVIEW

The increasing effects of light pollution on professional and amateur astronomy

Antonia M. Varela Perez[1,2]*

The starry sky has been a source of inspiration throughout human history. Astronomy has been a common element in all cultures and civilizations, being used to establish calendars, navigate and discover new lands, and drive numerous scientific and technical breakthroughs. This Review discusses how it is becoming increasingly difficult for professional and amateur astronomers to observe the night sky because of light pollution. Artificial light at night, radio interference, and the deployment of satellite constellations are all rapidly increasing and are having adverse impacts on astronomical observations, limiting scientific discoveries, cultural connections to the night sky, and opportunities presented by astrotourism. Potential mitigation strategies to preserve the night sky are discussed.

Since ancient times humans have illuminated their settlements using fire for safety or nighttime activities such as travel or trade. These were later replaced by longer-lasting oil lamps and wicks. Public electric lighting was developed in the late 19th century and soon spread around the world. These early forms of lighting (fire, then incandescent bulbs) emitted light with an approximately black body spectrum and low color temperature, consisting mostly of red light and very little blue content.

During the 20th century, street lighting adopted high-intensity discharge lamps and then metal vapor lamps, powered by electric discharge tubes containing metals such as mercury or sodium. These produce visible light with discrete emission lines, which contaminate observations of those lines (or others with similar wavelengths) in astronomical sources. In the 21st century, these technologies are being replaced by more efficient light-emitting diode (LED) lighting, which produces a broader spectrum with substantial blue content (*1*). These developments in lighting technology affect the amount of light pollution produced.

Increasing levels of light pollution

In recent decades the increasing global population, economic growth, and reduction of illumination costs due to the use of LEDs and the bluer light they emit are all contributing to rapidly increasing impacts on the natural night sky. The improvements in lighting efficiency have indirectly led to increased light pollution.

Outdoor lighting in cities alone consumes 19% of global electricity (predicted to reach 27% by 2040), costing 30 to 50% of a typical city's energy bill (*2*). Generation of that electricity produces greenhouse gas emissions, which contribute to climate change. Global lighting accounts for emissions equivalent to 1471 million tons of CO_2 per year, equivalent to 18% of total emissions in China or 27% in the US, the two countries with the highest greenhouse gas emissions (*3*). There is little public awareness of the link between lighting and CO_2 emissions.

Overillumination—the deployment of lighting that is unnecessary, brighter than required, or switched on at inappropriate times—has become widespread, affecting most towns and cities worldwide. This wasteful overuse of lighting causes unjustified energy consumption, needless economic expense, and excess emission of greenhouse gases.

83% of the world's population lives under light-polluted skies (*4*); 23% of the world's land surface between 75°N and 60°S is light-polluted (*4*), and this number is growing at 2% per year, i.e., doubling in about 35 years (*5*). By 2050 the world population is predicted to reach 9.6 billion people, of which 68% are expected to live in urban areas (*6*). Currently 733 million people lack electricity supply and another 2.4 billion do have access to clean cooking fuels and technologies (*7*). If current lighting practices are continued, the combination of population increase and economic development will lead to rapid increases in the levels of light pollution.

Sales of LEDs had a compound annual growth rate of more than 18% between 2017 and 2022 and are estimated to grow by another 15% between 2022 and 2027, reaching 141 million USD per year (*8*). Many types of LED lighting involved in this market expansion are not compatible with preserving the night sky because they have inappropriate color temperatures that emit large fractions of blue light, which is more strongly scattered in the atmosphere.

Light pollution is increasing worldwide at an estimated rate of 9.6% per year (*9*), reducing darkness everywhere, even at the most remote observation sites. The increasing levels of light pollution produced by the glow of artificial light at night (ALAN) is erasing the stars from our visible skies. This deterioration will leave a legacy with scientific, cultural, environmental, and aesthetic repercussions. Over-lighting also causes glare—the decrease in visual capacity or distortion of perception due to the presence of high luminance—and spectral light pollution, the emission of light at wavelengths that are useless for the intended purpose. Figure 1 shows an example of light pollution on the night sky.

Light pollution at astronomical observatories

Astronomical observations require dark skies and any sky glow can overwhelm the faint signal from astronomical objects, preventing their

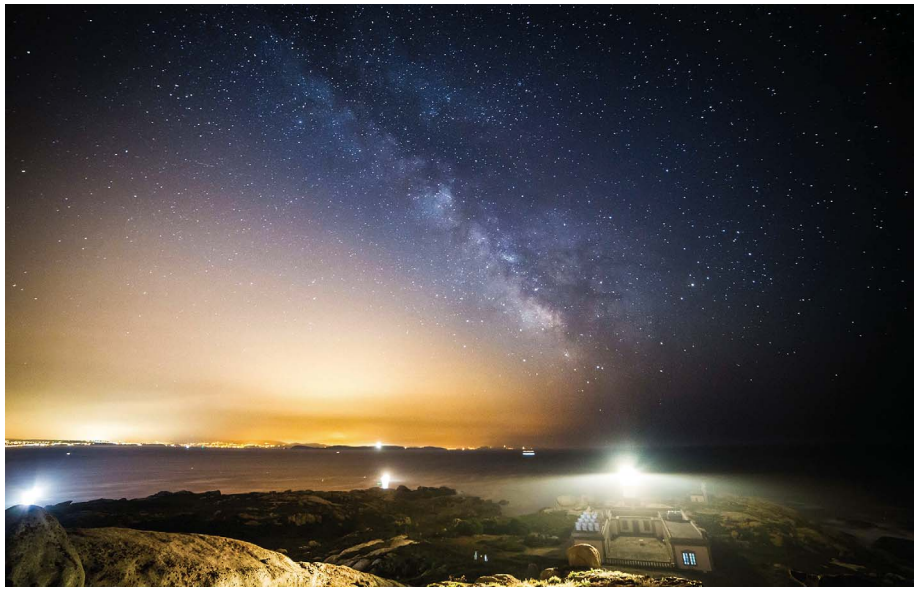


Fig. 1. Example of light pollution. The night sky photographed from the island of Sálvora in northern Spain, part of the Atlantic Islands of Galicia National Park and a certified Starlight Tourist Destination. The Milky Way extends diagonally across the image. The bright sky glow in the lower left is produced by light pollution, mainly from the use of white and blue LEDs and poorly shielded light fixtures from the towns located on the mainland coast of Rías Baixas, Galicia. Compare the number of stars visible in the lower left and upper right parts of the image.

[1]Instituto de Astrofísica de Canarias, 38205-La Laguna, Tenerife, Spain. [2]Starlight Foundation, 38205-La Laguna, Tenerife, Spain.
*Corresponding author. Email: avp@iac.es

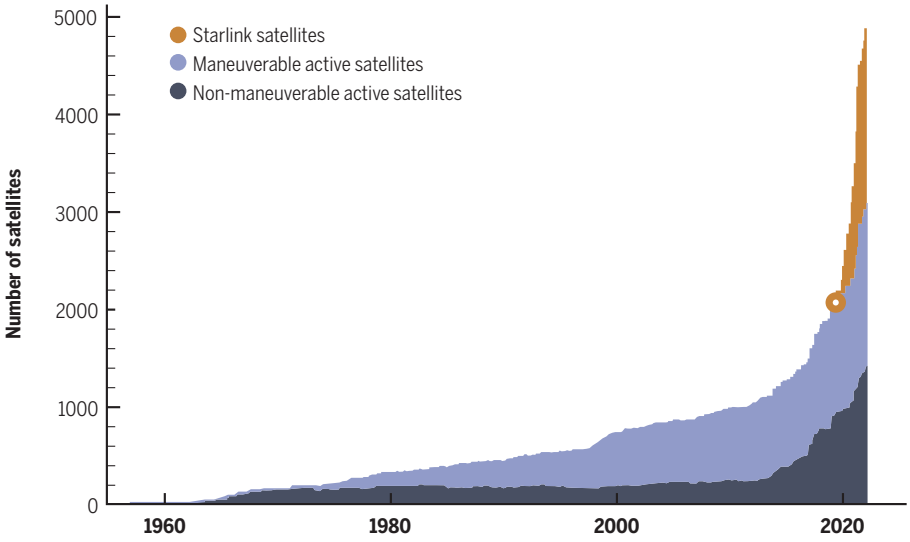


Fig. 2. Number of active satellites in orbit between 1957 and 2022. Data from (*18*). The number of maneuverable (blue) and nonmaneuverable (yellow) active satellites are plotted as a function of time. The pink circle indicates the beginning of the LEO communications constellation Starlink (pink). The three categories are additive. Between 2020 and 2022 the number of active satellites has grown by more than 50%.

detection. This problem becomes more acute for fainter astronomical objects or when observing with the naked eye. Because light pollution is generally worse in urban areas with high levels of economic development, dark skies are now mostly restricted to rural areas, often in more economically impoverished regions that are at risk of depopulation. In some of those places, access to dark skies is the basis for sustainable local economic development through star tourism, also known as astrotourism. Others have been chosen to host the large telescopes of professional observatories.

Even remote locations used for professional astronomical observatories or astrotourism no longer have pristine skies with natural levels of darkness. Light pollution is gradually encroaching everywhere as are radio signal interference and artificial changes to weather patterns caused by climate change and modifications to land use. Astronomical observations are also compromised by bright trails produced by increasing numbers of satellites orbiting Earth (discussed below). Two-thirds of major professional observatories are affected by light pollution at levels that exceed the target set by the International Astronomical Union (IAU) of 10% increase in radiance above the expected natural levels (*10*).

Ground-based astronomical observations continue to drive major, high-impact discoveries in astrophysics and fundamental physics. They are often essential to interpret observations from space-based telescopes. Ground-based optical telescopes can be built at a substantially larger size and roughly two orders of magnitude lower cost (per unit collecting area) than those launched into orbit. There are more than 40 ground-based optical telescopes with mirror diameters of 3 meters or larger, located in Australia, Chile, China, India, South Africa, Spain, Russia, and the United States, which constitute a substantial worldwide investment (*11*). The only space telescope of the same size is the James Webb Space Telescope (JWST), which took several decades to design, construct, and launch at a total cost of ~10 billion USD (*12*). The next generation of large ground-based optical telescopes currently under construction (the Giant Magellan Telescope, the Thirty-Meter Telescope, and the Extremely Large Telescope) will have apertures ranging from 3 to 6 times larger than JWST, are expected to be completed this decade, and have much smaller budgets (*12*, *13*).

Some types of astronomical data cannot be obtained from orbit, such as for planetary defense and space debris tracking. Arrays of light-collecting dishes similar in size to large radio antennas are deployed in remote locations (including in the United States, Namibia, Spain, and Argentina) to capture the faint flashes of blue Cherenkov radiation produced when high-energy gamma rays hit the top of Earth's atmosphere. These cannot be replaced with space-based telescopes.

Protecting sites for astronomy

Recognizing their importance for the local economy, some governments of regions that host astronomical observatories have introduced laws to limit light pollution, such as the 1988 Ley del Cielo (law of the skies) enacted by the Canary Islands to protect observatories on the islands of La Palma and Tenerife. Similar regulations have been introduced in other areas including Chile and Hawaii. However, the increasing impact of light pollution on observatories demonstrates that existing regulations are insufficient to halt the ongoing damage.

To do so, it is necessary to reduce, stop, and then reverse the growth of ALAN at its main sources (human population, agricultural, and industrial centers) within the next decade (*11, 12*). The IAU recommends that the total contribution of ALAN (quantifiable with sky glow measurement instruments and theoretical models) should be kept substantially below 10% of the natural dark sky level at an elevation of 45° in any azimuthal direction. Bluer photons are more strongly scattered in the atmosphere with the amount of scattering being inversely proportional to the fourth power of the wavelength. Therefore, reducing sky glow requires the reduction or removal of lighting that emits blue wavelengths: The blue light content (the amount of light emitted at wavelengths below 500 nm as a fraction of the total light emitted) should be as close as possible to zero (*11*). Recommended lighting sources should be quasi monochromatic and have their maximum radiant flux (in watts per nanometer) in the range of 570 to 605 nm (yellow-orange), similar to sodium vapor lamps (*11*). It is necessary for emissions to be at wavelengths longer than 555 nm—the astronomical *r*-band filter—which is already affected by emission lines as a result of oxygen and sodium that add to the natural night sky brightness.

Dark sky areas should be established around critical locations such as observatories and nature reserves, with zonal light output limits (similar to air quality zones). Close to these zones there should be no artificial light unless the need is clearly demonstrated. In this case, monochromatic illumination is strongly preferred (narrow-band amber), or when white light is required for safety LEDs, color temperatures of 1800 K should be used.

In addition to color, orientation and intensity also affect the amount of sky glow produced. Outside the area intended to be directly illuminated by the light source, the visible intensity should be as close as possible to zero even at low altitudes above the horizon (*14*). Light pollution could be further reduced by dimming lights, turning them off later in the night, or using motion sensors to increase brightness only when the lights are actually needed.

Wherever such regulations have been implemented, they have resulted in energy and economic savings as well as reductions in greenhouse gas emissions. Public administrations that host astronomical observatories are aware of the importance of protecting the night sky for science and to benefit the local economy.

The impact of satellite constellations

The deployment of large numbers of satellites in low Earth orbit (LEO) has had an unanticipated impact upon astronomy. Several communication providers have begun launching or are in advanced stages of planning networks

of communications satellites referred to as constellations or mega constellations. Tens of thousands of these satellites are anticipated, orbiting at low altitudes from 400 to 1200 km (*15*). When a prototype batch of 60 satellites was launched in May 2019, astronomers were surprised by how bright they appeared from the ground. Between 5 and 10% of satellites are present above astronomical sites at any given time, with a proportion of those being illuminated by the Sun in a dark sky (*16*). With plans for up to 400,000 satellites in such constellations by 2030, thousands will be visible from any location at any time.

Although these constellations have legitimate goals of providing communication access even in remote areas, they have damaging effects for the astronomical community. The use of LEO means that even small satellites are bright enough to be observable with the naked eye, especially at low elevations above the horizon and during twilight (close to dawn and dusk). Satellites are illuminated by the Sun, remaining in sunlight for longer than the ground, with the resulting brightness also depending on the season and latitude of the observer. Satellites in lower altitude orbits generally appear brighter from the ground, whereas those in higher altitude orbits are illuminated for longer periods before and after twilight. Therefore, in general, constellations in higher orbits are expected to be more damaging to astronomical observations.

Figure 2 shows the number of active satellites in orbit since 1957. The number has increased rapidly, from 2200 in May 2019 to 5000 in May 2022, with another 58,000 satellites predicted to be launched by 2030. Including inactive satellites, this increases the total number in orbit to ~8000 in 2023.

Satellite constellations affect optical and infrared astronomy as a result of their brightness and the number of spacecraft involved. At optical wavelengths and viewed from the ground, each spacecraft appears as a moving source with apparent visual magnitudes (as perceived by the human eye, smaller magnitudes indicate brighter objects) of about +3 to +4 mag during the deployment process and +6.5 to +9 mag in their final orbits. Several thousand satellites have already been deployed (Fig. 2) and hundreds of thousands more are proposed. The number of satellites visible above the horizon from an observatory is anticipated to be ~1600 immediately after sunset, decreasing to 1100 at the end of astronomical twilight (*16*). Of those satellites, 85% would be close to the horizon (below 30° elevation) but that still leaves more than 100 across the higher elevations commonly used for astronomy. That study assumed 26 thousand satellites (*16*); since then the number of proposed LEO spacecraft has increased by more than a factor of 10. Satellite mega constellations also produce a diffuse sky brightening due to the scattering of sunlight from satellite debris. Assuming 65,000 satellites, this contribution has been estimated as ~0.5% of the dark sky, and the sunlight scattering of the upcoming satellites from the mega constellation into debris would create an additional ~10% over the dark sky brightness (*17*). This limit (~10%) was adopted in 1979 by the International Astronomical Union for the light pollution level not to be exceeded at the sites of astronomical observatories (*10*).

There is currently no regulation of the impact on astronomy by satellite mega constellations. The astronomical community did not become aware of the issue until spacecraft were already being launched. The impacts are expected to be particularly adverse for observations at low elevations or during twilight and automated surveys searching for moving objects (such as potentially hazardous asteroids) (*16*). Survey telescopes with wide fields of view will be severely affected. Twilight images taken with short exposure times (1 s) are essentially unaffected by satellite trails, but those with medium (100 s) or long (1000 s) exposures are contaminated by satellite trails in up to 0.5 and 1% of the

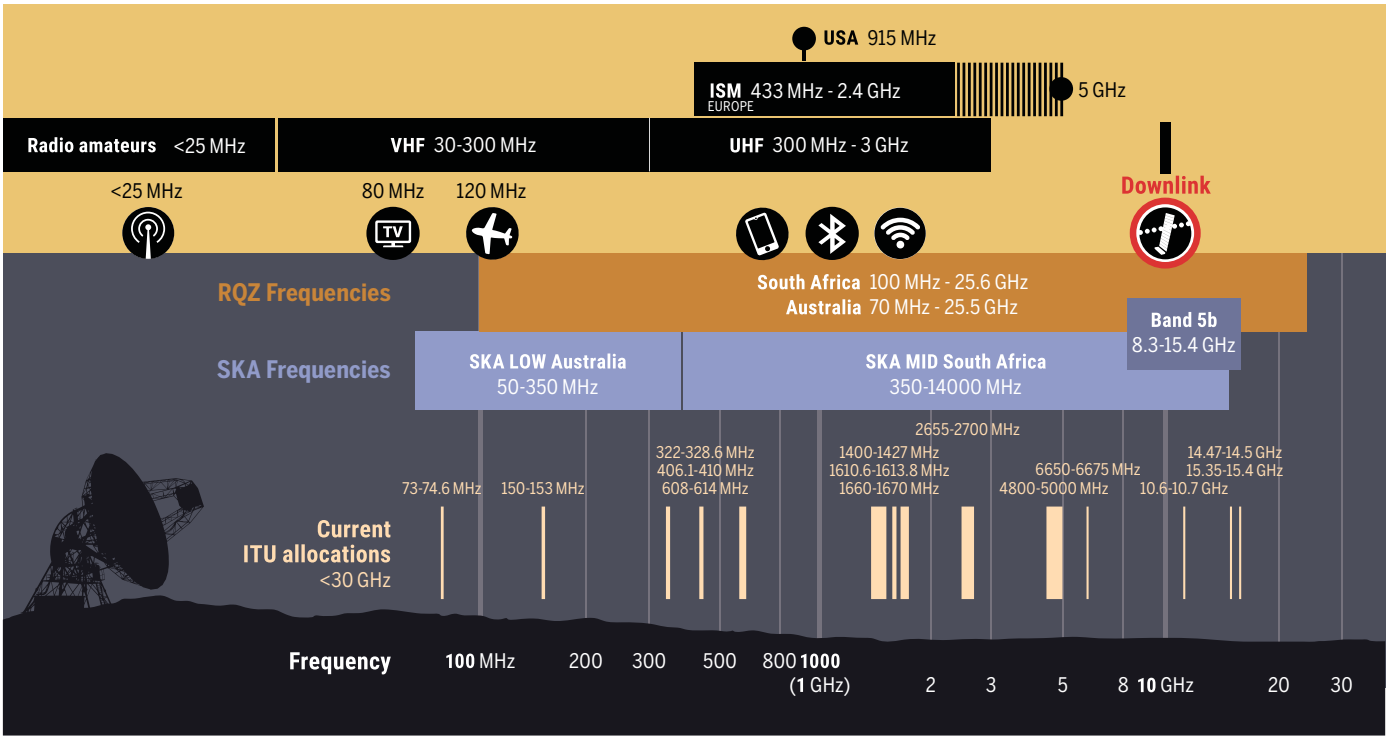


Fig. 3. Comparison of radio frequencies used for astronomy with sources of radio interference. Specific bands reserved for astronomy by the International Telecommunication Union (ITU) are shown up to 30 GHz, compared with observational ranges of the SKA and the radio quiet zones (RQZs) established by the host governments. Icons indicate the radio bands used for (from left to right) amateur radio, television, aviation, mobile telephony, bluetooth, wifi, 5b Band, and LEO satellites downlink. Black boxes indicate the frequency bands used for VHF (very high frequency) transmission, UHF (ultra high frequency) transmission, and ISM (industry, science, and medicine).

pixels, respectively (*16*). Up to 30% of wide-field exposures on a large telescope would be lost during the early evening pre-dawn hours, and nearly 50% of twilight exposures would be contaminated (*16*). The under-construction Vera C. Rubin Observatory, designed to perform wide-field surveys, is predicted to have up to 40% of its images be unusable due to satellite trails or saturation of the detectors (*16*). Even telescopes with smaller fields of view would have ~10% of images affected during twilight (*16*, *19*). Those studies considered the current satellite constellations so those rates will be higher if the anticipated numbers of spacecraft are launched.

Astronomical observations would greatly benefit if planned satellite constellations use as few spacecraft as possible (with the optimal number being zero), and to keep the satellite orbits low so that they enter the shadow of Earth soon after sunset. Astronomers and satellite operators have started to explore ways to mitigate their effects but there is an urgent need for national and international regulations. Several expert working groups (*11, 12*) have made more than 40 specific recommendations to various stakeholders (observatories, industry, astronomy community, science funding agencies, national and international policy-makers) for technical and political steps required to limit the impact on astronomy.

Radio interference impacts on radio astronomy

Radio astronomy involves observing the Universe at wavelengths that are also used by human-generated radio communications. The increasing bandwidth and transmission powers used by radio communications have led to increasing levels of radio frequency interference with astronomical observations. Satellite constellations also generate radio emissions. Some radio bands have been protected for astronomy by international agreements but transmitters always have some level of leakage into the protected bands, resulting in interference.

Radio quiet zones (RQZ) are regulatory zones around radio telescopes where human-generated radio emissions are controlled, established by local governments and the International Telecommunication Union (ITU) (*20*). These have been effective in reducing nearby radio emissions on the ground but do not protect against radio transmissions by spacecraft in orbit passing overhead. Until recently, satellite transmissions at frequencies between 10 and 80 GHz were generated mostly by satellites in geostationary Earth orbit (GEO), with lower contributions from LEO satellites. The deployment of LEO constellations will produce hundreds of fast-moving sources of bright radio interference, visible to radio telescopes at all times of day (*12*).

Figure 3 compares the radio frequency bands reserved for astronomy by ITU and those used by common sources of radio interference, both from ground-based technology and LEO satellites. For comparison, it also shows the observing ranges of the Square Kilometre Array (SKA), an international radio observatory currently under construction in Australia and South Africa, and the radio quiet zones that have been authorized at the telescope sites. This local ground-based protection of the observatory extends beyond the bands allocated by ITU for astronomy. However, it does not protect the observatory from radio interference produced by satellites. One exception is in Germany, where the government has an agreement with the Starlink satellite constellation to avoid the spacecraft transmitting while they pass over the Effelsberg radio telescope. Similar agreements will be necessary for other radio observatories.

Fig. 4. Photograph of Comet C/2020 F3 (NEOWISE) contaminated by satellite trails. The image was taken by Daniel López from Tenerife using amateur astronomer equipment. The comet is visible but so are numerous streaks generated by the passage of ~30 satellites through the field of view. Some of the streaks have gaps as a result of the multiple exposures that were stacked to produce the image.

Amateur astronomy and astrotourism

Approximately one million people are active in amateur astronomy, two orders of magnitude larger than the number of professional astronomers (*21*). ALAN and LEO satellite constellations will also affect amateur astronomy, particularly in the areas of professional-amateur scientific research programs, astrophotography, and astrotourism.

Amateur astronomers discover comets, search galaxies for supernovae, perform variable star and meteor tracking campaigns, and confirm candidate exoplanets (*12*). These activities are particularly vulnerable to light pollution because amateur astronomers do not have access to the economic and technological resources required to mitigate its effects. Increasing levels of light pollution seriously compromise these activities, which will become practically impossible in the next decade if current trends continue.

The trails of LEO satellites affect amateur astrophotography. Figure 4 shows an image of Comet C/2020 F3 (NEOWISE) taken using a digital reflex camera. The photographer obtained a series of 30-second exposure images, then stacked them to reach sufficient depth to see the comet. However, 90% of the exposures showed satellite trails that contaminated the image. Amateur astronomers use cameras and telescopes with wider fields of view than large professional telescopes and so are more likely to have artificial satellite trails in their images.

Degradation of the night sky affects wider society, as the night sky has inspired developments in philosophy, art, culture, and religion. Cultural interpretations of the stars have been developed worldwide through generations, expressed in legends, folktales, stories, and other traditional practices; many of these are at risk of extinction (*22*).

The night sky is a resource to be safeguarded not only for science but also for its value to culture, the environment, biodiversity, human health and quality of life, and as a driver of sustainable economy through astrotourism (*23*). Astrotourism combines night- and day-time sky observations, outreach, and leisure activities related to astronomy. It is intended to be a sustainable form of tourism and is suitable for territories that are less economically developed, which often means less light-polluted (*24*).

Areas where dark night skies are legally protected are known as dark sky oases (*11*). Similar to physical oases, these locations are distinguished by the abundance of a resource surrounded by

scarcity. Dark skies oases are intended to foster implementation of environmentally friendly and sustainable lighting solutions. There are currently more than 200,000 km^2 of territories (approximately the size of the UK) in more than 30 countries with protected skies (*11*, *25*), accredited by international institutions. These include areas designated as Biosphere Reserves, Natural or National Parks, Ramsar sites, World Heritage sites, etc. These areas are diversifying their economies and developing sustainable and responsible astrotourism. Astrotourism helps to disseminate an appreciation for astronomy and provides a tool against depopulation of undeveloped areas, providing jobs with technical and scientific skills (*26*).

Light pollution endangers the operation of infrastructure dedicated to astrotourism, such as tourist astronomical observatories, star parks, astronomical viewpoints, telescope farms, astronomy interpretation centers, and traveling planetariums. Many of these facilities receive hundreds of daily visits; collectively they provide employment opportunities to thousands of professionals. Over the last decade, hundreds of astrotourism companies have been established around the world. Many of their staff have backgrounds in amateur astronomy, who see their livelihood threatened by the growing levels of light pollution. Other astrotourism professionals come from the tourism sector (mainly from eco- and rural tourism) and the business sector, who regard astrotourism as a way to diversify their business (*24*). Therefore, light pollution affects not only amateur astronomy but also the business sector and local economic development. Demand for astrotourism certification and training has grown by more than 300% in the last five years (*26*), attracting tens of thousands of visitors and providing an economic return of over 100 million dollars in multiple territories. This growth is threatened by increasing light pollution.

Potential mitigation strategies

ALAN is regulated by local, regional, and national governments, some of whom have enacted legislation to control outdoor lighting specifically to reduce light pollution. Several international organizations work to protect the night sky for astronomy.

The United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) of the UN Office of Outer Space Affairs (UNOOSA) is the institution responsible for regulating human activities in space. The effect of spacecraft on astronomy falls within the COPOUS remit and requires protection from the uncontrolled growth of artificial sky glow. However the impact of satellites on astronomy is essentially unregulated except for the protection of narrow radio bands by the ITU.

To discuss the impact of ALAN, satellite constellations, and radio frequency interference on astronomy, two workshops on Dark and Quiet Skies for Science and Society were held in October 2020 and 2021. The purpose was to establish recommendations for protecting the dark and (radio-)quiet skies as well as exploration of how to implement them. The workshops produced two reports (*11, 12*) that aimed to identify the technical and political actions required by the stakeholders and partners who would need to collaborate to preserve a dark and quiet sky. Similarly, two Satellite Constellations Workshops were held in July 2020 and 2021, which made recommendations of how to mitigate the negative impacts of satellite constellations on astronomy and the night sky (*15*, *19*).

The IAU set up a Centre for Protection of the Dark and Quiet Sky from satellite constellation interference in 2022 (*27*). Its goal is to propose mitigation strategies, policy, and regulatory measures to be implemented by local and national governments.

Preserving the skies has been proposed for inclusion in the 2030 Agenda for Sustainable Development, a revision of the UN sustainable development goals (*28*). Light pollution is related to sustainable development goals in the areas of poverty, inequality, climate, environmental degradation, prosperity, peace, and justice. "Sky Quality and Access to Starlight" has been proposed as sustainable development goal 18 (SDG18) and a public petition has been set up to promote it (*28*). The SDG18 proposal was submitted in March 2022; more than 2000 individuals and institutions from over 25 countries have supported it.

We are all made of stardust, literally and culturally. Astronomy is the oldest science, helping to drive scientific and technological discoveries. Having access to dark skies is necessary for many livelihoods and cultures. The starry sky is an opportunity for development and empowerment for many indigenous and local communities in rural locations. It is our duty to protect and safeguard it for present and future generations.

REFERENCES AND NOTES

1. B. A. Bowers, *A History of Electric Light and Power* (The Science Museum, 1982).
2. F. Pardo-Bosch, A. Blanco, E. Sesé, F. Ezcurra, P. Pujadas, *Sustain Cities Soc.* **76**, 103454 (2022).
3. Green Paper. Policy Options to Accelerate the Global Transition to Advanced Lighting (UNEP, 2014); https://sdgs.un.org/sites/default/files/publications/1702Green_Paper_FINAL%20prepared%20for%20the%20Global%20Efficient%20Lighting%20Forum.pdf
4. F. Falchi *et al*., *Sci. Adv.* **2**, e1600377 (2016).
5. C. C. M. Kyba *et al*., *Sci. Adv.* **3**, e1701528 (2017).
6. 68% of the world population projected to live in urban areas by 2050 (Department of Economic and Social Affairs, United Nation News, 2018); https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html
7. Tracking SDG7: The Energy Progress Report (IEA, 2022); https://www.iea.org/reports/tracking-sdg7-the-energy-progress-report-2022
8. LED Market by Application, Product, and Geography -Forecast and Analysis 2023-2027 (TechnavioPlus, 2022); https://www.technavio.com/report/led-market-industry-size-analysis
9. C. C. M. Kyba, Y. Ö. Altıntaş, C. E. Walker, M. Newhouse, *Science* **379**, 265–268 (2023).
10. R. Cayrel, F. G. Smith, *Transactions of the International Astronomical Union* **18A**, 667 (1982).
11. Scientific Organizing Committee, Dark Sky Oases Working Group, Optical Astronomy Working Group, Bioenvironment Working Group, Satellite Constellation Working Group, Radio Astronomy Working Group, Dark & Quiet Skies for Science and Society (IAU, dqskies-book-29-12-20, 2020); https://www.iau.org/static/publications/dqskies-book-29-12-20.pdf.
12. C. Walker, P. Benvenuti, Dark & Quiet Skies for Science and Society II for Science and Society (Working Group Reports, 2021); https://zenodo.org/record/5874725#.ZGTzpnbMKUk.
13. G. T. van Belle, A. B. Meinel, M. P. Meinel, in *Proceedings of Ground-based Telescopes* (SPIE, 2004), pp. 563–570.
14. C. B. Luginbuhl, C. E. Walker, R. J. Wainscoat, *Phys. Today* **63**, 10 (2010).
15. C. Walker, J. Hall, Impact of Satellite Constellations on Optical Astronomy and Recommendations Toward Mitigations (NSF NOIRLab, AURA, AAS, 2020); https://aas.org/sites/default/files/2020-08/SATCON1-Report.pdf.
16. O. R. Hainaut, A. P. Williams, *Astron. Astrophys.* **636**, A121 (2020).
17. M. Kocifaj, F. Kundracik, J. C. Barrentine, S. Bará, *Mon. Not. R. Astron. Soc. Lett.* **504**, L40–L44 (2021).
18. J. C. McDowell, *Astrophys. J. Lett.* **892**, L36 (2020).
19. SATCON2 Scientific Organizing Committee, SATCON2 Working Group Reports (NSF NOIRLab, AURA, AAS, 2021); https://doi.org/10.5281/zenodo.5608820.
20. Characteristics of Radio Quiet Zones, (Report ITU-R RA.2259-1, International Telecommunication Union, 2021).
21. IAU Executive Committee Working Group for Professional-Amateur Relations in Astronomy (International Astronomical Union Annual Report, 2021); https://www.iau.org/static/science/scientific_bodies/working_groups/330/wg-330-annual-report-2021-2022.pdf
22. UNESCO-WHC, Mab Urban Ecology Programme of UNESCO, IAU, UNWTO, OTPC-IAC, CIE, Fundación Starlight, C. Marín, Starlight Reserve Concept (2009); https://en.fundacionstarlight.org/docs/files/89_concept-st-reserve-english.pdf
23. Fundación Starlight, C. Marín, J. Jafari, StarLight: Declaration in Defence of the Night Sky and the Right to Starlight (Institute of Astrophysics of the Canary Islands, 2017); https://fundacionstarlight.org/docs/files/78_declaracion-sobre-la-defensa-del-cielo-nocturno-y-el-dereho-a-la-luz-de-las-estrellas-ingles.pdf
24. Fundación Starlight, The Astro-Tourism by the Starlight Foundation (Starlight Foundation, 2023); https://en.fundacionstarlight.org/contenido/110-astrotourism.html
25. D. Welch, World list of dark sky places (Dark Sky Advisory Group of the International Union for Conservation of Nature, 2023); http://darkskyparks.org/dsag/2021-02-28_DSAG_word_list.htm
26. A.M. Varela, Turismo de estrellas La preservación del cielo nocturno como motor de desarrollo sostenible socio-económico local, *Ambient@* **132** (2022).
27. IAU Centre for the Protection of the Dark and Quiet Sky from Satellite Constellation Interference: CPS (IAU, 2023); https://cps.iau.org
28. BPW Spain, Fundación Starlight, SDG 18: Sky Quality and Access to Starlight (2023); https://ods18.org/en/.

ACKNOWLEDGMENTS

My greatest gratitude goes to all the colleagues who have constituted the working groups of the Dark & Quiet Skies for Science and Society Workshops for their valuable contribution to the results presented in this article and to the institutions that have promoted them. My deepest thanks to the referees for their detailed review and fruitful comments. My thanks also to J. Sánchez-Almeida for his detailed review of the article and his helpful comments. I thank G. Pérez for his assistance in preparing Fig. 3.

Submitted 2 February 2023; accepted 26 April 2023
10.1126/science.adg0269

EDITORIAL

Commit to the Global South

Vulnerabilities of the world's least developed countries to health, economic, and environmental crises were recently in the spotlight at a United Nations (UN) conference in Qatar. In March, the Doha Programme of Action was announced as a call for the developed world to renew commitments to support low- and middle-income countries in overcoming major challenges. UN Secretary General António Guterres aptly stated there are "no more excuses." Such a commitment requires Global North-South and South-South partnerships that harness science and technology and that also empower the South to sustain progress. As a scientist from the Global South who can attest to science's transformative role in this region of the world, individuals and organizations across sectors and society can play a strong role in supporting such a commitment.

Growing up in apartheid South Africa in the 1970s, I lived through huge racial differences in socioeconomic status and living conditions across the country. Blacks in South Africa and people in the Global South in general struggled to meet their basic needs for food, housing, and health care, and access to education, water, and sanitation. I saw how countries in the Global North prospered through quality education and an emphasis on science, technology, innovation, and engineering. Starting a career in epidemiology in the early 1990s, I focused my research on HIV/AIDS, a rigorous scientific and social justice challenge. This disease unmasked huge differences within and between countries regarding those who bear the burden of infection. A common thread was that the most vulnerable were (and still are) the most poor, marginalized, stigmatized, and discriminated against populations. However, global solidarity in seeking scientific solutions and social activism catalyzed drastic reductions in the costs of antiretroviral therapy. Establishment of the US President's Emergency Plan for AIDS Relief and the Global Fund to Fight AIDS, Tuberculosis and Malaria reflected commitment of the Global North to tackle the pandemic in the Global South. Africa alone, home to about 70% of the global burden of HIV infection, now has over 20 million AIDS patients on this treatment, with a resulting 20-year increase in the life expectancy of those individuals.

> "...the world must do more than lament the growing inequalities and the looming threats..."

Although collaborations and partnerships in scientific endeavors across borders are important in the short to medium term, sustaining progress requires commitment to long-term investments in building human resource capacity at regional and national levels. Given that 27% of the population in low- and middle-income countries are under 14 years of age, it is time to sow the seeds for career paths across the Global South's scientific enterprise that are accessible to all.

Earlier this year, I was elected to serve as the president of The World Academy of Sciences (TWAS). Established 40 years ago as the Third World Academy of Sciences, it has envisioned sustainable development in vulnerable regions of the world through scientific solutions in research, education, policy, and diplomacy—endeavors that long preceded the UN Sustainable Development Goal (SDG) framework. Over the past four decades, TWAS and other global organizations such as the Abdus Salam International Centre for Theoretical Physics, the Organization for Women in Science for the Developing World, the InterAcademy Partnership, and the International Science Council have strengthened the science base in developing countries by supporting training and capacity building through strong South-South and North-South collaborations. For example, there are currently more than 700 TWAS fellowship students working toward graduate degrees locally in the Global South. But much remains to be done to fill this education and training pipeline with students. And long-term investment in infrastructure that supports a growing regional pool of talent also is needed.

In a time of unprecedented social, political, and economic upheavals and deepening inequalities within and between countries, organizations across science sectors must more actively support improving the quality of life globally, especially in the Global South. As member states gather for the mid-term review of SDG progress later this year, the world must do more than lament the growing inequalities and the looming threats that face everyone. It is important to consider not only who does the science but also who receives the fruits of science if the Global South is to leapfrog into an equitable present and future.

–Quarraisha Abdool Karim

PHOTO: RAJESH JANTILAL

Quarraisha Abdool Karim is president of The World Academy of Sciences, Trieste, Italy; associate scientific director of the Centre for the AIDS Programme of Research in South Africa, Durban, South Africa; and professor of Clinical Epidemiology at the Mailman School of Public Health, Columbia University, New York, NY, USA. quarraisha.abdoolkarim@caprisa.org

10.1126/science.adj1621

NEWS

> “I think this species will rapidly go extinct, because the demand is so intense and the price is so high.”
>
> **Gerontologist Malene Hansen,** in *Sixth Tone*, a China-based magazine, about reports that smugglers in China are illegally capturing wild long-tailed macaques to meet a large demand for the monkeys in research.

IN BRIEF

Edited by **Jeffrey Brainard**

Brain specialists propose banning checks in high school hockey games like this one in Biddeford, Maine.

SPORTS MEDICINE

Concussion experts call for rule changes

An expert panel of researchers and clinicians this week called for an end to virtually all checking in youth ice hockey and backed other changes for preventing and treating sports-related concussions (SRCs). A systematic review overseen by the international group found that banning checking (a collision by a defensive player to dislodge the puck) reduced the rate of concussions in child and adolescent games by 58%; requiring mouthguards dropped the rate by 28% at all levels of play. The authors of the consensus statement, published in the *British Journal of Sports Medicine*, also conclude that the standard prescription of strict rest and bans on screen time for many days after an SRC may not be beneficial. The group recommends that easy physical activity such as walking begin immediately. As for an especially controversial question in professional sports—the extent to which SRCs cause later neurological disease such as dementia and Parkinson’s disease—the authors, some of them consultants to or employees of professional sports leagues, say research has not proved a causal relationship. They call for further studies.

Diabetes drug deters Long Covid

BIOMEDICINE | The inexpensive and widely used diabetes drug metformin decreased the risk of developing Long Covid in overweight and obese outpatients who took it while acutely ill with SARS-CoV-2 infection, researchers reported last week in *The Lancet Infectious Diseases*. The risk was 41% lower for the metformin group than for a comparison group receiving a placebo. The study presents the first evidence from a randomized clinical trial that a medication taken by patients who have acute symptoms such as cough and fever can reduce the risk of Long Covid, an ill-defined, chronic condition with symptoms that can include extreme fatigue, breathlessness, and brain fog. About 1100 U.S. trial participants were given placebo or metformin within 7 days of becoming ill. In the following 180 to 300 days, 10.4% in the placebo group reported being diagnosed with Long Covid versus 6.3% in the metformin group.

Global earthquake model released

SEISMOLOGY | A nonprofit scientific group this week released a set of global models and maps describing earthquake hazards and risks to help inform building codes and tailor other precautions in areas most vulnerable to damage. The tools from the Global Earthquake Model Foundation, updated for the first time since 2018, are based on 30 seismic models that together provide about twice the resolution of prior efforts. The team updated and expanded data about inflation, population growth, and building practices, all of which elevate the potential for loss in earthquake-prone regions. The risk models suggest the greatest potential for structural damage is in India; economic loss, Japan; displaced people, Pakistan; and fatalities, China.

Lab safety doesn’t hurt output

WORKPLACE | Prioritizing lab safety does not hamper research productivity, a study of practices at the University of California, Los Angeles (UCLA) shows. In recent decades, laboratory disasters have killed or grievously injured young scientists at several

institutions, but some researchers say tighter precautions would interfere with getting work done in the lab. To test this viewpoint, researchers examined the productivity of 600 chemistry labs at UCLA, measured by numbers of scientific publications, before and after a fatality from a 2008 fire led the institution to systematically tighten inspections and rules. A few labs increased safety precautions by using safer alternatives to hazardous chemicals. Among all 600 labs, there was no change in productivity after the researchers controlled for other factors, they reported this month in a National Bureau of Economic Research working paper. Lab-safety specialists say many universities have room to improve.

Antacid for the ocean

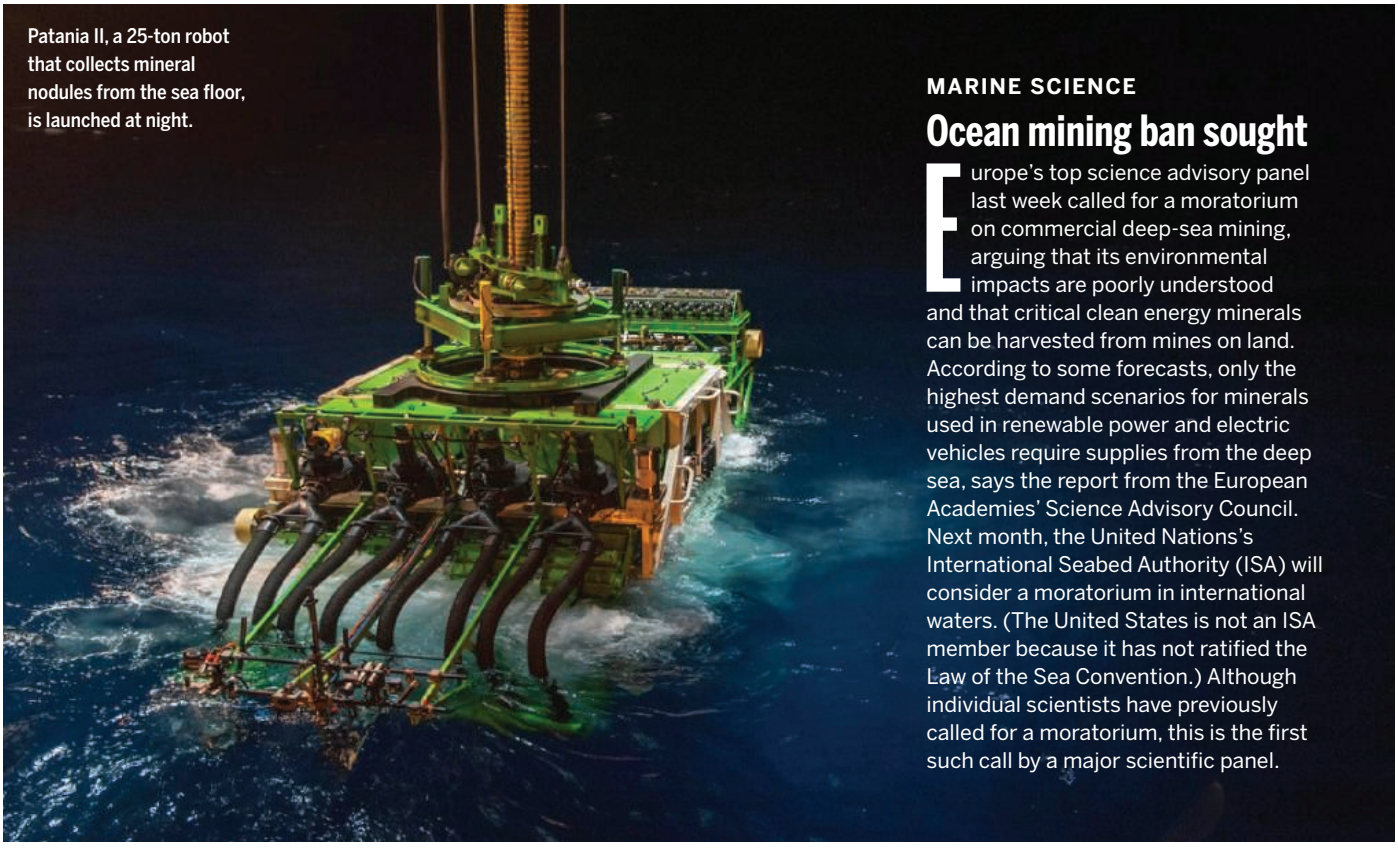
GEOENGINEERING | A nonprofit group that supports research on manipulating ocean chemistry to remove atmospheric carbon dioxide announced last week its first research grants, $23 million for nine studies. The ocean already naturally absorbs 30% of annual carbon emissions and contains 50 times more carbon than the atmosphere. The water's natural capacity to store carbon can be boosted by pulverizing and adding alkaline rocks such as basalt or limestone. Alkaline minerals in the rocks convert dissolved carbon dioxide into long-lasting carbonate and bicarbonate ions, allowing the ocean to absorb more atmospheric carbon dioxide. The Carbon to Sea Initiative, founded by Meta's former chief technology officer, will support projects to examine the efficacy, costs, and environmental effects of ocean alkalinization schemes.

Egypt criticizes 'Afrocentrism'

ARCHAEOLOGY | A decades-old Dutch archaeological mission in Egypt has been suspended after authorities there revoked the National Museum of Antiquities's permit to excavate at Saqqara, a necropolis on the outskirts of Cairo. The move was a reaction to the museum's new exhibit "Kemet: Egypt in Hip Hop, Jazz, Soul & Funk," a look at how ancient Egyptian and Nubian culture inspired modern-day Black musicians. In an email to museum Director Wim Weijland last week, an official at Egypt's Ministry of Tourism and Antiquities accused the museum of "falsifying history" by including an Afrocentric approach to Egypt's past and for depicting Nefertiti with dark skin. The decision follows recent statements by prominent Egyptian archaeologists criticizing Netflix for casting a Black actor to portray Cleopatra in a recent docudrama. In a statement on its website, the museum said its social media has been bombarded with racist comments. "We cordially invite everyone to visit the exhibition and form their own opinions," the museum writes.

Humans entered Asia early

PALEOANTHROPOLOGY | Scientists have dated skull and shin bones from a modern human, found in a cave in northern Laos, to between 68,000 and 86,000 years ago, before the main migration of our species out of Africa that gave rise to all living non-African populations. Genetics-based models agree the main migration occurred sometime between 50,000 and 60,000 years ago. The new study, published this week in *Nature Communications*, bolsters evidence that modern humans sojourned out of Africa more than once and traveled vast distances—although these precocious migrants likely didn't contribute many, if any, genes to modern populations. Previous claims to have found human remains in Southeast Asia from before 60,000 years ago have been met with skepticism. But experts say their confidence in the dating of the Laotian cave is bolstered by the authors' use of a combination of state-of-the-art methods.



Patania II, a 25-ton robot that collects mineral nodules from the sea floor, is launched at night.

PHOTO: DEME

MARINE SCIENCE

Ocean mining ban sought

Europe's top science advisory panel last week called for a moratorium on commercial deep-sea mining, arguing that its environmental impacts are poorly understood and that critical clean energy minerals can be harvested from mines on land. According to some forecasts, only the highest demand scenarios for minerals used in renewable power and electric vehicles require supplies from the deep sea, says the report from the European Academies' Science Advisory Council. Next month, the United Nations's International Seabed Authority (ISA) will consider a moratorium in international waters. (The United States is not an ISA member because it has not ratified the Law of the Sea Convention.) Although individual scientists have previously called for a moratorium, this is the first such call by a major scientific panel.

IN DEPTH

ORIGIN OF LIFE

How the molecules of life became one-handed

"Breakthrough" experiments suggest magnetic materials skewed early biomolecules

By **Robert F. Service**

In 1848, French chemist Louis Pasteur discovered that some molecules essential for life exist in mirror image forms, much like our left and right hands. Today, we know biology chooses just one of these "chiral" forms: DNA, RNA, and their building blocks are all right-handed, whereas amino acids and proteins are all left-handed. Pasteur, who saw hints of this selectivity, or "homochirality," thought magnetic fields might somehow explain it, but its origin has remained one of biology's great mysteries. Now, it turns out Pasteur may have been onto something.

In three new papers, researchers suggest magnetic minerals common on early Earth could have caused key biomolecules to accumulate on their surface in just one mirror image form, setting off a positive feedback that continued to favor the same form. "It's a real breakthrough," says Jack Szostak, an origin of life chemist at the University of Chicago who was not involved with the new work. "Homochirality is essential to get biology started, and this is a possible—and I would say very likely—solution."

Chemical reactions are typically unbiased, yielding equal amounts of right- and left-handed molecules. But life requires selectivity: Only right-handed DNA, for example, has the correct twist to interact properly with other chiral molecules. To get life, "you've got to break the mirror, or you can't pull it off," says Gerald Joyce, an origin of life chemist and president of the Salk Institute for Biological Studies.

Over the past century, researchers have proposed various mechanisms for skewing the first biomolecules, including cosmic rays and polarized light. Both can cause an initial bias favoring either right- or left-handed molecules, but they don't directly explain how this initial bias was amplified to create the large reservoirs of chiral molecules likely needed to make the first cells. An explanation that creates an initial bias is a good start, but "not sufficient," says Dimitar Sasselov, a physicist at Harvard University and a leader of the new work.

Hints of another option date to 1999, when researchers led by Ron Naaman, a chemical physicist at the Weizmann Institute of Science, found that the electrons in opposite chiral forms of a molecule have contrasting patterns of spin, a magnetic property. Later experiments revealed that the spin differences can cause chiral molecules to interact differently with magnetic materials, in which electron spins are aligned to create magnetic forces. For example, Naaman and his colleagues found that left-handed peptides (short amino acid chains) might bind to a magnetic surface while right-handed ones are repelled.

An RNA-making molecule crystallizes on magnetite, which can bias the process toward a single chiral form.

But this finding, too, did not explain how the initial bias could be amplified.

A glimmer of an amplification mechanism emerged in 2009. Researchers led by Matthew Powner and John Sutherland at the University of Manchester were studying possible origins of RNA, which many researchers think was a central player in the origin of life. They were intrigued by a molecule called ribo-aminooxazoline (RAO), which they discovered could react to form two of RNA's nucleotide building blocks. RAO is among a rare class of crystals that enforce a single chirality: Once a crystal starts to grow from either right- or left-handed versions of the molecule, only molecules with the same chirality can bind to the structure. Such crystals, if they started with an initial bias, could have caused chiral RAO to build up.

Now, Sasselov and his colleagues have put these two pieces together. They wondered whether magnetic surfaces might favor a single RAO chiral form. To find out, they turned to magnetite, a magnetic mineral that is common in Earth's crust. They applied a strong external magnetic field, aligning electron spins in the magnetite and strengthening its magnetism. When they exposed the magnetite surface to a

solution containing an equal mix of right- and left-handed RAO molecules, 60% of those that settled on top were of a single handedness. This created a crystalline seed that caused additional like-handed RAOs to bind, eventually forming pure single-handed RAO crystals, the researchers reported last week in *Science Advances*. When they flipped the field's orientation and repeated the experiment, crystals with the opposite handedness took shape. "It's a really cool effect and a way to break the symmetry," says Powner, now at University College London.

One concern is that the applied magnetic field was some 6500 times stronger than Earth's own field, cautions Noémie Globus, a physicist at the University of California, Santa Cruz, whose own work supports cosmic rays as the source of life's chiral bias. "It requires conditions that are quite unrealistic," she says.

But previous reports show that magnetite subjected only to Earth's natural magnetic field can still cause an initial, though smaller, bias toward one form of a chiral molecule. And Sasselov and his colleagues reported in a 13 April arXiv preprint that when pure chiral RAO crystals were placed on top of magnetite, the alignment of the electron spins in the crystals forced more and more electron spins in the underlying magnetic material to align, creating a positive feedback. "It's self-enhancing and increases the persistence of the bias" toward one molecular form, says team member Furkan Ozturk, a Harvard Ph.D. student.

The chiral RAO in turn imposes its handedness on the RNA building blocks it generates, and Sasselov's team has now shown that the effects cascade to other biological molecules. In a report accepted last week in *The Journal of Chemical Physics* they show that once an excess of chiral RNA is formed, known chemical reactions could pass on this chiral bias, templating amino acids and proteins with the opposite handedness and ultimately fostering other chiral molecules essential to cell metabolism. "There is no solution out there that solves all the steps out there that this does," Szostak says.

The quest that began with Pasteur isn't quite over, though. One loose end, Sasselov acknowledges, is that RAO has only been shown to lead to the synthesis of two of RNA's four nucleotides, cytosine and uracil. It isn't known to produce the other two, adenine and guanine, although Sasselov says there's a "big push" to search for RAO reactions that could do it. If they can, the mystery of biological handedness might be another step closer to being solved. ■

PHOTO: DAVID J. PHILLIP/ASSOCIATED PRESS

POLICY

DOE lab would probe climate effects on poor communities

New national lab would go to a historically Black institution

By **Jeffrey Mervis**

President Joe Biden has asked Congress for $35 million to begin planning a new Department of Energy (DOE) laboratory that would study the impact of climate change on poor communities. The initiative is buried in the agency's fiscal year 2024 budget, which the president submitted this spring, and *Science* has learned that DOE officials did not request the money.

If funded, the new national lab would join the 17 basic science, energy, and weapons labs now run by DOE. Apart from the topic, its most striking feature would be its location. It would be built at one of the country's 104 historically Black colleges and universities (HBCUs), which were founded to provide opportunities for Black people barred from attending most U.S. institutions of higher education because of their race.

"Establishing a national lab at an HBCU is a top priority for the Department of Energy and for the Biden administration as a whole," says a senior White House official who spoke with *Science* on the condition of anonymity. DOE and several other agencies now run competitive programs that give small grants to HBCUs and other institutions serving large numbers of minority students. But the new lab would be a much larger and longer lasting enterprise.

DOE's network of national laboratories has taken on monumental technical challenges over more than 8 decades, from developing the atomic bomb to pursuing laser fusion as a future energy source. With annual budgets that can top $2 billion, the labs are typically operated by a private contractor under renewable contracts from the government.

The new lab would tackle problems facing "disadvantaged communities that are marginalized, underserved, and overburdened by pollution," the White House official said. That effort would also further the administration's Justice40 initiative, which seeks to funnel "40 percent of the overall benefits of certain Federal investments" to those distressed communities.

Current global climate models don't provide enough spatial resolution to guide local policymakers who are preparing for impacts such as excessive heat and flooding, says Harvard University physicist Cherry Murray. "Coming up with community-scale climate models is a huge challenge," says Murray, a former head of DOE's basic science shop, which manages 10 of the national labs. "Nobody knows how to do that now." The varied environments within cities make fine-scale modeling even more difficult.

Murray worries that the challenge may be too big for a lab that would be starting from scratch and run by an institution that historically has received relatively little federal funding. "Building up the necessary computing capacity is going to be very expensive," she notes. "And then you also need to have the administrative expertise to manage a national lab." She would prefer to see DOE make it easier for HBCU faculty to tap into the extensive facilities and large scientific teams already in place at an existing national lab.

Flooding from Tropical Storm Harvey displaced many Houston residents in 2017.

The president's budget request doesn't flesh out how the $35 million would be spent. "We realize it's not sufficient to build a full lab," the White House official says. "But it's meant to start the planning process and figure out the best opportunities going forward."

Should Congress approve the requested funding, the next step would likely be a series of community workshops on what such a lab could look like. DOE would then solicit proposals from HBCUs it deems eligible to build and operate the lab. ■

solution containing an equal mix of right- and left-handed RAO molecules, 60% of those that settled on top were of a single handedness. This created a crystalline seed that caused additional like-handed RAOs to bind, eventually forming pure single-handed RAO crystals, the researchers reported last week in *Science Advances*. When they flipped the field's orientation and repeated the experiment, crystals with the opposite handedness took shape. "It's a really cool effect and a way to break the symmetry," says Powner, now at University College London.

One concern is that the applied magnetic field was some 6500 times stronger than Earth's own field, cautions Noémie Globus, a physicist at the University of California, Santa Cruz, whose own work supports cosmic rays as the source of life's chiral bias. "It requires conditions that are quite unrealistic," she says.

But previous reports show that magnetite subjected only to Earth's natural magnetic field can still cause an initial, though smaller, bias toward one form of a chiral molecule. And Sasselov and his colleagues reported in a 13 April arXiv preprint that when pure chiral RAO crystals were placed on top of magnetite, the alignment of the electron spins in the crystals forced more and more electron spins in the underlying magnetic material to align, creating a positive feedback. "It's self-enhancing and increases the persistence of the bias" toward one molecular form, says team member Furkan Ozturk, a Harvard Ph.D. student.

The chiral RAO in turn imposes its handedness on the RNA building blocks it generates, and Sasselov's team has now shown that the effects cascade to other biological molecules. In a report accepted last week in *The Journal of Chemical Physics* they show that once an excess of chiral RNA is formed, known chemical reactions could pass on this chiral bias, templating amino acids and proteins with the opposite handedness and ultimately fostering other chiral molecules essential to cell metabolism. "There is no solution out there that solves all the steps out there that this does," Szostak says.

The quest that began with Pasteur isn't quite over, though. One loose end, Sasselov acknowledges, is that RAO has only been shown to lead to the synthesis of two of RNA's four nucleotides, cytosine and uracil. It isn't known to produce the other two, adenine and guanine, although Sasselov says there's a "big push" to search for RAO reactions that could do it. If they can, the mystery of biological handedness might be another step closer to being solved. ■

POLICY

DOE lab would probe climate effects on poor communities

New national lab would go to a historically Black institution

By **Jeffrey Mervis**

President Joe Biden has asked Congress for $35 million to begin planning a new Department of Energy (DOE) laboratory that would study the impact of climate change on poor communities. The initiative is buried in the agency's fiscal year 2024 budget, which the president submitted this spring, and *Science* has learned that DOE officials did not request the money.

If funded, the new national lab would join the 17 basic science, energy, and weapons labs now run by DOE. Apart from the topic, its most striking feature would be its location. It would be built at one of the country's 104 historically Black colleges and universities (HBCUs), which were founded to provide opportunities for Black people barred from attending most U.S. institutions of higher education because of their race.

"Establishing a national lab at an HBCU is a top priority for the Department of Energy and for the Biden administration as a whole," says a senior White House official who spoke with *Science* on the condition of anonymity. DOE and several other agencies now run competitive programs that give small grants to HBCUs and other institutions serving large numbers of minority students. But the new lab would be a much larger and longer lasting enterprise.

DOE's network of national laboratories has taken on monumental technical challenges over more than 8 decades, from developing the atomic bomb to pursuing laser fusion as a future energy source. With annual budgets that can top $2 billion, the labs are typically operated by a private contractor under renewable contracts from the government.

The new lab would tackle problems facing "disadvantaged communities that are marginalized, underserved, and overburdened by pollution," the White House official said. That effort would also further the administration's Justice40 initiative, which seeks to funnel "40 percent of the overall benefits of certain Federal investments" to those distressed communities.

Current global climate models don't provide enough spatial resolution to guide local policymakers who are preparing for impacts such as excessive heat and flooding, says Harvard University physicist Cherry Murray. "Coming up with community-scale climate models is a huge challenge," says Murray, a former head of DOE's basic science shop, which manages 10 of the national labs. "Nobody knows how to do that now." The varied environments within cities make fine-scale modeling even more difficult.

Murray worries that the challenge may be too big for a lab that would be starting from scratch and run by an institution that historically has received relatively little federal funding. "Building up the necessary computing capacity is going to be very expensive," she notes. "And then you also need to have the administrative expertise to manage a national lab." She would prefer to see DOE make it easier for HBCU faculty to tap into the extensive facilities and large scientific teams already in place at an existing national lab.

Flooding from Tropical Storm Harvey displaced many Houston residents in 2017.

PHOTO: DAVID J. PHILLIP/ASSOCIATED PRESS

The president's budget request doesn't flesh out how the $35 million would be spent. "We realize it's not sufficient to build a full lab," the White House official says. "But it's meant to start the planning process and figure out the best opportunities going forward."

Should Congress approve the requested funding, the next step would likely be a series of community workshops on what such a lab could look like. DOE would then solicit proposals from HBCUs it deems eligible to build and operate the lab. ■

After 20 years in prison, Kathleen Folbigg enjoys freedom again thanks to genetic evidence that her four children likely died from natural causes.

GENETICS

How a geneticist led an effort to free a convicted serial murderer

"Science has been heard," says Carola Vinuesa after mother jailed for killing her four kids was pardoned and released

By **John Travis**

Carola Vinuesa woke up early on 5 June in London, around 3 a.m., when colleagues from Australia called the clinical geneticist with big news: Kathleen Folbigg, who on only circumstantial evidence was famously convicted of killing her four young children and jailed 2 decades ago, had just been pardoned by New South Wales and set free.

Her freedom is in large part due to the efforts of Vinuesa and other scientists who had amassed genetic evidence they said likely explained the children's deaths. The pardon "was so exciting, so beautiful. I was so happy, for Kathleen first of all, but for science [too]," says Vinuesa, who now runs a lab at the Francis Crick Institute. "It's a day to celebrate that science has been heard and has made a difference. And not just to this case, I think." She told *Science* that the Australian Academy of Science played a crucial role as an independent science adviser to the new inquiry into the Folbigg verdict, an arrangement Vinuesa hopes will become a model for how legal systems address complex science.

Folbigg was convicted after her two sons and two daughters, ranging in age from 19 days to 18 months, died mysteriously at home. The original guilty verdict rested primarily on the apparent unlikelihood of four children in one family dying naturally and ambiguous writing in Folbigg's diary that the prosecution suggested betrayed a guilty conscience.

A Spaniard who conducted research in Australia for years before recently moving to London, Vinuesa was drawn into the Folbigg case in 2018 when an inquiry into the verdict was launched and she was contacted by a former student who had become a lawyer and then joined the incarcerated woman's legal team. Vinuesa, a mother herself, read the medical files of the dead children and saw signs of underlying illnesses in all four, such as respiratory infections, a major cause of child mortality. And just 1 month earlier, she and colleagues had identified a genetic mutation that appeared to explain four mysterious infant deaths in a family in Macedonia.

Noting that more than one-third of sudden deaths in children can be explained by genetic conditions, Vinuesa says she quickly became convinced a miscarriage of justice had happened in Folbigg's case. "The idea that four deaths in a family is too rare [to be natural]? It isn't," she says. "It's obviously very, very unfortunate but these things happen, particularly when there is a genetic condition." Multiple high-profile murder cases have recently spotlighted the pitfalls of relying on statistics to indicate guilt (*Science*, 20 January, p. 232).

Vinuesa enlisted a colleague, Todor Arsov, to visit Folbigg in prison and get a DNA sample. When the two examined the sequence, they found a mutation in a gene, *CALM2*, encoding a protein called calmodulin whose misfunction had already been implicated in heart arrythmias and disorders and even deaths in infancy. If Folbigg's children had inherited the mutation, it could have caused their deaths, especially if the kids also had an infection or other stress.

In 2019, Vinuesa's team tested newborn blood samples from the daughters and learned that they had indeed inherited the mutation. But it still wasn't clear at the time whether Folbigg's specific *CALM2* mutation was harmless or pathogenic. In the parlance of genetics, it was a variant of unknown significance.

Vinuesa reached out to cardiac experts, including Peter Schwartz of the Italian Auxological Institute, who had long studied heart issues arising from mutations in the genes for calmodulins. He realized that a patient registry he oversaw had a family that mirrored the Folbigg situation: a healthy mother with a similar mutation in a gene for another calmodulin, who had two kids that suffered heart attacks, with one dying. "It really brought the [Folbigg] mutation over the line. You could classify it as likely pathogenic," Vinuesa recalls.

Carola Vinuesa testified at two hearings on Kathleen Folbigg's case.

But other Australian researchers testified in 2019 at the first inquiry that they didn't think the data were definitive, and Folbigg's verdict was upheld. Vinuesa says that was a time of frustration and anger for her and others working on Folbigg's behalf. "We're used to having science that some people might believe more than others or that is criticized ... but when [your] work hinges on the freedom and

public image of someone, [disagreement is] much more difficult to deal with," she says. "You stop sleeping at night sometimes."

Vinuesa and her allies continued the fight. They amassed further scientific evidence, including studies from three different labs suggesting the *CALM2* mutation was pathogenic and data showing that both sons had mutations in another gene that can cause lethal epilepsy in mice. They compiled it all in a 2020 peer-reviewed publication. Its authors and many other scientists joined a petition for yet another Folbigg inquiry, and the Australian Academy of Science backed the plea. "The academy was very courageous. Anna-Maria Arabia [its CEO] persuaded the academy this was a cause worth endorsing and that was key," Vinuesa says.

After a second inquiry was launched in 2022, the tide turned. Vinuesa and Schwartz testified at the new hearing, as well as protein scientist Michael Toft Overgaard and geneticist Mette Nyegaard, both of Denmark's Aalborg University, whose initial cell studies of the *CALM2* mutation suggested it could cause heart arrhythmias. The state attorneys involved in the new hearing last month told the New South Wales attorney general there was now reasonable doubt that Folbigg had killed her children. He agreed and last week's pardon was the result. A former judge in charge of the inquiry is still scheduled to release a final report, which could bring an official exoneration of Folbigg and perhaps even compensation.

Vinuesa believes science can now play a greater role in explaining sudden deaths that seem suspicious. The databases recording normal human genetic variation have grown tremendously. And high-throughput methods to test the effects of individual mutations in cells have been developed, she notes. Gene variants can now be officially scored, using criteria endorsed by the American College of Medical Genetics and Genomics, as "likely pathogenic" in a death (a 90% chance it was the cause) or "pathogenic" (a 99% chance).

"People have to be retrained to know how to interpret whole genome data," she says. "For judges and barristers, particularly if they haven't had any scientific training, it's very difficult. There have to be mechanisms put in place to help."

Other women accused of killing or inflicting harm on a child have sought Vinuesa's help as a result of the Folbigg case. She says she's not eager to get back into court, but in some cases may look for genetic diagnoses that may have been missed. "There needs to be more collaboration between the scientific world and the legal profession," she says. Folbigg would no doubt agree. ■

HEALTH INEQUITIES

Segregation was already hurting Black health 100 years ago

Research links structural racism of 1900s U.S. society to striking disparities in childhood mortality

By **Rodrigo Pérez Ortega**

From the choice of schools to safety to access to green spaces and healthy food, the neighborhood where a child is raised can play a determining role in their future health. And because structural racism can systematically silo nonwhite people in certain neighborhoods, those local factors shape the health of millions of people of color in the United States. Now, census data link Black children's neighborhoods and mortality rates in the early 20th century, exposing segregation's devastating impact on health more than 100 years ago.

The study shows segregation drove racial health disparities "not just today, but [also] in the past," says New York University community psychologist Adolfo Cuevas, who was not involved in the work. John Parman, an economist at the College of William & Mary, says the new results are striking because they document the impacts even before the makings of the Jim Crow era in the late 19th century, which legalized and enforced racial segregation and is known to have exacerbated health inequities.

A growing body of evidence has shown that, today, neighborhoods with majority nonwhite residents tend to have poorer health—the result of many accumulated social and environmental inequalities such as systematic overcrowding, higher noise levels due to industrial projects, and exposure to toxic hazards. But how early such residential segregation began to affect health was not clear, says J'Mag Karbeah, a health services researcher at the University of Minnesota (UM) who led the new study.

So Karbeah and J. David Hacker, a demographic historian at UM, set out to correlate early segregation with child mortality, a proxy for the health of the entire population. "If you don't have a healthy young population, you won't have healthy working-age adults, [and] you will not have healthy seniors," Karbeah says. "It's really predictive of the quality of your society in the next 40 or 50 years." The researchers used census data from 1900 and 1910 that were recently processed by the Minnesota Population Center at UM. The lists, which together cover about 168 million people, include information on literacy, race, and whether the individual lived in a rural or urban area. Census takers also asked each surveyed woman who had ever been married how many children she had given birth to and how many were still alive.

From the data, Karbeah and Hacker reconstructed the number of children born in the 5 years before each census to arrive at a sample of nearly 4.7 million Black and white children. Focusing on the South because 90% of the Black population resided there at the time, they compared the mortality rates for Black and white children. They also calculated the spatial distribution of houses headed by Black or white people as a measure of segregation.

The largest mortality gap was in Savannah, Georgia, in 1910, where Black children were 3.2 times more likely to die than their white counterparts, with almost half dying before age 5. To tease out the influence of segregation, the team controlled for socioeconomic and other variables such as literacy, occupation, and unemployment. They found that in 1910, neighborhood segregation as much as doubled the mortality gap between Black and white children in cities, the team reported last week in *Population, Space and Place*.

Although the researchers couldn't explain exactly how segregation affects child mortality, children are extremely vulnerable to environmental pollutants as well as to poor sanitation, Karbeah says, all of which tend to go hand in hand with housing segregation. "These are the populations that have been most impacted by these inequities within the neighborhood environment," she says.

Parman notes that a century ago, access to modern medical interventions, clean water and sanitation was not as different for Black and white residents as it was in later years. That makes the mortality gulf between Black and white children all the more striking. "That speaks to a broader set of mechanisms through which segregation might impact health," he says. ■

public image of someone, [disagreement is] much more difficult to deal with," she says. "You stop sleeping at night sometimes."

Vinuesa and her allies continued the fight. They amassed further scientific evidence, including studies from three different labs suggesting the *CALM2* mutation was pathogenic and data showing that both sons had mutations in another gene that can cause lethal epilepsy in mice. They compiled it all in a 2020 peer-reviewed publication. Its authors and many other scientists joined a petition for yet another Folbigg inquiry, and the Australian Academy of Science backed the plea. "The academy was very courageous. Anna-Maria Arabia [its CEO] persuaded the academy this was a cause worth endorsing and that was key," Vinuesa says.

After a second inquiry was launched in 2022, the tide turned. Vinuesa and Schwartz testified at the new hearing, as well as protein scientist Michael Toft Overgaard and geneticist Mette Nyegaard, both of Denmark's Aalborg University, whose initial cell studies of the *CALM2* mutation suggested it could cause heart arrhythmias. The state attorneys involved in the new hearing last month told the New South Wales attorney general there was now reasonable doubt that Folbigg had killed her children. He agreed and last week's pardon was the result. A former judge in charge of the inquiry is still scheduled to release a final report, which could bring an official exoneration of Folbigg and perhaps even compensation.

Vinuesa believes science can now play a greater role in explaining sudden deaths that seem suspicious. The databases recording normal human genetic variation have grown tremendously. And high-throughput methods to test the effects of individual mutations in cells have been developed, she notes. Gene variants can now be officially scored, using criteria endorsed by the American College of Medical Genetics and Genomics, as "likely pathogenic" in a death (a 90% chance it was the cause) or "pathogenic" (a 99% chance).

"People have to be retrained to know how to interpret whole genome data," she says. "For judges and barristers, particularly if they haven't had any scientific training, it's very difficult. There have to be mechanisms put in place to help."

Other women accused of killing or inflicting harm on a child have sought Vinuesa's help as a result of the Folbigg case. She says she's not eager to get back into court, but in some cases may look for genetic diagnoses that may have been missed. "There needs to be more collaboration between the scientific world and the legal profession," she says. Folbigg would no doubt agree. ■

HEALTH INEQUITIES

Segregation was already hurting Black health 100 years ago

Research links structural racism of 1900s U.S. society to striking disparities in childhood mortality

By **Rodrigo Pérez Ortega**

From the choice of schools to safety to access to green spaces and healthy food, the neighborhood where a child is raised can play a determining role in their future health. And because structural racism can systematically silo nonwhite people in certain neighborhoods, those local factors shape the health of millions of people of color in the United States. Now, census data link Black children's neighborhoods and mortality rates in the early 20th century, exposing segregation's devastating impact on health more than 100 years ago.

The study shows segregation drove racial health disparities "not just today, but [also] in the past," says New York University community psychologist Adolfo Cuevas, who was not involved in the work. John Parman, an economist at the College of William & Mary, says the new results are striking because they document the impacts even before the makings of the Jim Crow era in the late 19th century, which legalized and enforced racial segregation and is known to have exacerbated health inequities.

A growing body of evidence has shown that, today, neighborhoods with majority nonwhite residents tend to have poorer health—the result of many accumulated social and environmental inequalities such as systematic overcrowding, higher noise levels due to industrial projects, and exposure to toxic hazards. But how early such residential segregation began to affect health was not clear, says J'Mag Karbeah, a health services researcher at the University of Minnesota (UM) who led the new study.

So Karbeah and J. David Hacker, a demographic historian at UM, set out to correlate early segregation with child mortality, a proxy for the health of the entire population. "If you don't have a healthy young population, you won't have healthy working-age adults, [and] you will not have healthy seniors," Karbeah says. "It's really predictive of the quality of your society in the next 40 or 50 years." The researchers used census data from 1900 and 1910 that were recently processed by the Minnesota Population Center at UM. The lists, which together cover about 168 million people, include information on literacy, race, and whether the individual lived in a rural or urban area. Census takers also asked each surveyed woman who had ever been married how many children she had given birth to and how many were still alive.

From the data, Karbeah and Hacker reconstructed the number of children born in the 5 years before each census to arrive at a sample of nearly 4.7 million Black and white children. Focusing on the South because 90% of the Black population resided there at the time, they compared the mortality rates for Black and white children. They also calculated the spatial distribution of houses headed by Black or white people as a measure of segregation.

The largest mortality gap was in Savannah, Georgia, in 1910, where Black children were 3.2 times more likely to die than their white counterparts, with almost half dying before age 5. To tease out the influence of segregation, the team controlled for socioeconomic and other variables such as literacy, occupation, and unemployment. They found that in 1910, neighborhood segregation as much as doubled the mortality gap between Black and white children in cities, the team reported last week in *Population, Space and Place*.

Although the researchers couldn't explain exactly how segregation affects child mortality, children are extremely vulnerable to environmental pollutants as well as to poor sanitation, Karbeah says, all of which tend to go hand in hand with housing segregation. "These are the populations that have been most impacted by these inequities within the neighborhood environment," she says.

Parman notes that a century ago, access to modern medical interventions, clean water and sanitation was not as different for Black and white residents as it was in later years. That makes the mortality gulf between Black and white children all the more striking. "That speaks to a broader set of mechanisms through which segregation might impact health," he says. ■

BIOMEDICINE

NIH's data policy for foreign partners stirs outcry

Some scientists worry reporting requirement is burdensome—and even insulting

By **Jocelyn Kaiser**

Many U.S. biomedical researchers and their collaborators abroad are reacting with alarm to a new National Institutes of Health requirement that foreign groups send their NIH partners copies of their lab notebooks and other raw data at least every few months. It's "crazy," says Brazilian researcher Mauro Teixeira, who gets NIH funding through a U.S. university for a mosquito-borne disease study.

The 19 May policy change, sparked by concerns about U.S.-funded virology research in China, could impose a costly new paperwork burden on all NIH grants with foreign partners, from lab studies to clinical trials, scientists and a biomedical research advocacy group say. Currently, collaborators only share results used in publications and keep underlying data such as lab notebooks within their institutions.

"It could be a voluminous amount of data and information that's being transferred" for reasons that "are very unclear," says Heather Pierce, an attorney for the Association of American Medical Colleges, which plans to raise concerns about the policy.

Several global health researchers in the United States worry the mandate will damage long-standing relationships across the globe. The policy "further reinforces the mounting perception that the USA regards foreign scientists as problems and not as key collaborators in improving global health," says immunologist John Moore of Weill Cornell Medicine.

Indeed, Teixeira, who is at the Federal University of Minas Gerais, is offended as well as annoyed by the coming data reporting burden. "I'm going to need many more people just to send things, and the U.S. government will need more people just to look at it in a professional way," he says. The policy makes him wonder why he's even participating in the NIH-funded study, which examines whether releasing mosquitoes infected with bacteria that reduce virus transmission will slash rates of dengue and other diseases in Belo Horizonte, Brazil. "You're not trusting me to do the work. So why should you hire me anyway?"

A spokesperson for NIH, which estimates fewer than 2% of the agency's primary awards will be affected, explains that the rule "empowers" its grantees to obtain their foreign partners' data "without having to worry that they will not be able to access materials when needed." The new requirement, which takes effect on 1 October, should not be a burden for "well-run collaborations," the spokesperson says, because they will "already have in place processes by which there is seamless information flow."

Brazilian scientists studying mosquito-borne diseases as part of an NIH grant must now regularly submit raw data to U.S. partners.

NIH says the policy change resulted from a recent audit by the Department of Health and Human Services (HHS) that found problems with NIH's oversight of an award to the nonprofit EcoHealth Alliance. A subaward from EcoHealth to the Wuhan Institute of Virology (WIV) funded bat virus studies that some scientists and politicians have claimed, without direct evidence, sparked the COVID-19 pandemic.

Under orders from then-President Donald Trump, NIH terminated the grant in April 2020. It reinstated the award 3 months later but suspended the funding until EcoHealth met conditions, such as getting lab records from WIV, that the nonprofit said were impossible. Since then the award has been restarted with no China activities and new oversight restrictions.

In a January report, HHS's inspector general chided EcoHealth and NIH for failing to adequately monitor the WIV subaward and found that the China group did not turn over needed "scientific documentation" for the bat virus studies. To avoid similar standoffs, the watchdog advised NIH to step up reporting requirements for all grant subrecipients abroad.

NIH agreed and will now require the foreign grant partners to "provide copies of all lab notebooks, all data, and all documentation that supports the research outcomes as described in the progress report" no less often than every 3 months. The agency already requires that subaward recipients retain records related to the project and be ready to turn them over for audit purposes. But that is different from submitting all data to the main grant holder on a regular basis, Pierce says.

Yale University's Amy Bei runs a malaria vaccine development project with colleagues in Senegal that collects lab and survey data and malaria test results from participants. "The burden of regularly transmitting all records at regular intervals will be incredibly taxing for our partners," she says. And it's unclear what the main U.S. grantee is supposed to do with the information, scientists say.

Yale's Albert Ko, who collaborates with Teixeira on the mosquito release study, adds that for another project he'll have to demand raw data from trainee scientists in 21 countries. He wonders whether NIH will even review the data and instead is "generating an enormous amount of work for its awardees which will be rarely acted upon."

Pierce says the new mandate might make sense with "higher risk" projects in certain countries, but not as a blanket policy. NIH is taking comments on the policy change until 26 June and says the agency may update it "if determined necessary." ■

Floodwaters submerged at least 1200 square kilometers in Ukraine after a major dam collapsed.

PHOTO: ASSOCIATED PRESS

ENVIRONMENTAL SCIENCE

After Ukrainian dam breach, war hampers study of ecological toll

Mines and active combat make fieldwork too risky

By **Erik Stokstad**

As emergency managers finish evacuating people in southern Ukraine from disastrous flooding caused by the destruction of a major dam last week, conservation scientists fear for the region's ecosystems. The 6 June breach of the Kakhovka hydroelectric dam released a torrent that killed at least 13 people. It has also left downstream nature reserves underwater and caused vast mudflats to emerge from an emptying reservoir.

"This dam breach is especially dire since it broadsides three ecosystems simultaneously," says David Lytle, an expert on river flooding and biodiversity at Oregon State University. The 2155-square-kilometer reservoir, home to numerous species of fish and other animals, is rapidly drying out. The Dnipro River below the dam has experienced a flood far beyond normal proportions, eroding topsoil in some places and burying others in silt. The river's floodplains are submerged in water up to 5.6 meters deep. Newts that live in the river have washed ashore in Odessa, 150 kilometers across the Black Sea from the river's mouth.

The ecological disruptions are likely to be extensive, but it's too soon to know exactly how they will play out in the long term, says Oleksandra Shumilova, a freshwater ecologist at the Leibniz Institute of Freshwater Ecology and Inland Fisheries who has studied the Ukraine war's effects on water infrastructure. Much of the affected area is in a war zone, making it too risky for researchers to venture into the field. "For sure there will be a lot of [research] projects, but they are at the planning stage at the moment," Shumilova says.

Russian troops seized the dam soon after invading Ukraine last year. On the night it collapsed, Ukrainian and Russian soldiers on both sides of the Dnipro heard explosions in its power plant. The reservoir's 19 cubic kilometers of water—more than the major dams of the Colorado River now store—began to spill through the breach. Both sides blame the other for the catastrophe.

The flooding inundated 80 settlements in Ukraine-controlled territory, Oleksandr Krasnolutskiy, Ukraine's first deputy minister of environmental protection and natural resources, said last week. That's only about 40% of the flooded zone, he noted; the remainder is now controlled by Russia. On 7 June, the Institute of Geological Sciences of the National Academy of Sciences of Ukraine estimated that floodwaters covered 1200 square kilometers. (As *Science* went to press, the institute had not updated its estimate, based on data from a European Space Agency satellite, because of cloudy weather.)

The reservoir, Ukraine's second largest, was a key source of drinking water and helped irrigate farms that produced about 80% of Ukraine's fruits and vegetables, mostly in areas currently controlled by Russia. Downstream agriculture could also suffer; Krasnolutskiy said the flooding has washed topsoil from tens of thousands of hectares of farmland.

Ecosystems also will be hit hard, the Ukrainian Nature Conservation Group (UNCG) said in a 7 June report. The draining of the reservoir has destroyed the spawning grounds of fish, including pikeperch, that supported a commercial harvest of as much as 2600 tons a year. The receding waters are expected to allow foxes and other predators to reach bird habitat on what used to be islands, which could jeopardize nesting yellow bitterns (*Ardeola ralloides*) and other rare species. Important wetlands along the reservoir's Velyki and Mali Kuchuhury archipelagos could dry out. Ecologists fear invasive plants, such as ragweed and goldenrod, will colonize newly exposed sediment.

Downstream, the flooding could wipe out rare floodplain species, such as the sandy blind mole-rat (*Spalax arenarius*), and hammer populations of Nordmann's birch mouse (*Sicista loriger*) and other endemic species. "There is a risk we will lose these species forever," Krasnolutskiy said. Biologists with UNCG also expect large losses of endemic plants and important species such as giant oaks. Some low-lying places could remain flooded for up to 2 months, according to a division of the World Data Center in Ukraine. Long-term inundation—or even simply higher levels of groundwater—could be particularly damaging to steppe ecosystems, which have dry-adapted plants such as the Dnipro birch, Shumilova says.

Contamination is a concern, too. Sediment flowing from the reservoir is likely laced with heavy metals and other toxic chemicals from decades of industrial pollution. About 150 tons of machine oil leaked from the damaged hydroelectric plant, but environmental officials did not find excessive oil concentrations in water samples collected at three locations downstream, around Kherson.

Iryna Dronova, a landscape ecologist at the University of California, Berkeley, has been using satellite data to study how the war has changed Ukraine's farmland. She says remote sensing will also help chart the impacts of the flood after it recedes. But, "It is critical that assessment and monitoring efforts do not stop at the 'immediate' time frame around the disaster." And some of the data will require ground-truthing, she adds.

Safety risks are complicating fieldwork. The flood has washed Russian land mines into unexpected places. "It's very dangerous … because mines can be everywhere," Krasnolutskiy says. "We will be able to do research only when the Ukrainian army liberates [Russian-held] territories," adds environmentalist Oleksii Vasyliuk, who leads UNCG.

Much of the damage is still hidden under the floodwaters, and more change is coming. On 11 June, the reservoir was still 9 meters deep at Nikopol, about 120 kilometers upstream from the dam, with the level falling at about 4 centimeters per hour. Once the reservoir empties—which could occur within days—the river's lowermost reaches will return to a narrower channel unseen since the dam was built in the 1950s. ■

NATURAL BLUE

A food dye from a South American fruit has become a test case for the ethical development of natural resources

By **Kai Kupferschmidt**, *in Polines, Colombia; Photography by* **Cristina Abad Angel**

The road ends in Luz del Mundo. A few kilometers' walk from here, just across the Chigorodó River, lies Polines, a scattering of open-sided houses on the edge of the forest. It's a settlement of Indigenous Emberá people, about 80,000 of whom live in communities across Colombia and Panama. On a recent hot day, an Emberá man named Alirio Niaza scaled a large tree. Standing on one of the lower branches, he wielded a long pole with shears at the end. He carefully positioned the tip against a tiny branch, then pulled a cord that closed the blades. With a thud, a green, fist-size fruit landed in the grass.

For generations the Emberá have used these fruits to paint their skin. They scrape out the flesh of the unripe fruit and squeeze the pulp in a rolled leaf to extract a milky white juice. Mixed with charcoal, the liquid turns inky black and can be used to trace patterns on the skin. The sap reacts with the skin to form dark blue lines that appear within hours and stay visible for weeks.

"We use this for certain rituals," says Gabriel Bailarin, who lives in Polines and is Emberá. A newborn child may be painted head to toe to ensure good health, he says. *Jaibanas*, the traditional healers, draw mountains, snakes, and other patterns, each with a different meaning, says Nataly Domicó, an Emberá woman who has studied Emberá traditional medicine. "The patterns are like a language on the body."

Hector Bailarin, an Emberá man from Polines, Colombia, climbs a jagua tree to collect its unripe fruit.

Bluish veins in the pulp of the jagua fruit contain the makings of a dye used for body painting, which is an important part of Emberá culture.

Found in tropical forests across South America, the tree has many names. It is called quepera by the Emberá and jagua by most others in Colombia. Scientists know it as *Genipa americana*. The ripe fruit is sometimes eaten or squeezed for juice. But unlike its relative the coffee plant, it has gotten little attention as a potential crop—until now. After nearly 20 years of research to identify the chemical origins of the blue color and convert it into a stable dye, a jagua-based colorant is likely to be approved for use in the United States and other countries soon. It will enter a billion-dollar market for natural dyes that is hungry for a vivid, stable blue to color candies, cereals, and other foodstuffs. It will also become a test case for the responsible development of a natural resource long-used by Indigenous communities.

The man behind the project, Colombian entrepreneur Nicolás Cock Duque, says his business will help conserve biodiversity and improve people's livelihoods in Colombia, breaking with the long history of exploitation of impoverished communities. It's a novel effort that's being closely watched in Colombia. But observers point to a host of uncertainties: Can the fruit be sourced in a sustainable, affordable way from these rainforest trees? Can jagua blue succeed in the market? And if so, what benefits will local communities, including the Emberá, actually reap?

IN 2005, Cock Duque was working in the Chocó region in Colombia's northwest, where the largely undeveloped rainforest is one of the most biodiverse on Earth. Many people here subsist by mining gold and platinum. Cock Duque, who studied environmental policy at George Washington University, had founded an organization to certify traditional, less destructive mining operations with the idea of demanding a premium price for certified gold and platinum. The profit would finance restoration of the forests and provide extra income to the local population.

One day at the market in Quibdó, the region's capital, Cock Duque came across a stall where an Emberá woman was selling a fruit he'd never seen before. "I asked her what it was and she said: 'It's jagua. We use it to paint our bodies.'" She cut it open for him. "Maybe 30 seconds after she opened it, these blue veins appeared in the white flesh." These veins, he realized, might be mined, too.

A couple of years earlier, a German researcher had told Cock Duque to look out for a stable, natural blue, something that has long eluded food companies. The industry has for decades largely relied on two synthetic blues: brilliant blue, also called blue No. 1, usually made from petroleum; and blue No. 2, derived from synthetic indigo, which is used to color Viagra pills, among other things. But as consumer preference has shifted to all things natural, the demand for a nonsynthetic blue has grown, says Maria Buchweitz, a chemist at the University of Hamburg. Blue is desirable not just for its own sake, but because a bright blue dye can be mixed with yellow to produce a bright green, which is also in demand.

But bright blue colorants are rare in nature. "It's really complicated for plants to make blue," says Beverley Glover, a botanist at the University of Cambridge. In order for a molecule to appear blue to humans, it needs to absorb red light, the lowest energy part of the visible spectrum. "And it turns out in order to do that you need to make really complicated molecules, so they tend to be bigger, they have more side chains."

Only one class of plant pigments has achieved a true blue: anthocyanins. The bright blue cornflower, for instance, assembles a large complex of six anthocyanins and six copigments arranged like spokes on a wheel around two metal

Color source

A company based in Medellín, Colombia, is working to commercialize a blue food colorant derived from jagua fruit, long used by the Indigenous Emberá people in the Chocó region. Elsewhere, the company has encouraged non-Indigenous farmers to cultivate the fruit.

CREDITS: (PHOTO) CRISTINA ABAD ANGEL; (GRAPHIC) D. AN-PHAM/*SCIENCE*

After jagua juice has been squeezed from the pulp, charcoal is added. The black liquid is then used to trace lines on the skin that later turn blue.

ions. A few years ago, Buchweitz tried to develop the cornflower blue as a colorant for the food industry, but it didn't pan out. "It's too complicated and the food industry needs something simple that it can just add to different foods."

THAT DAY AT the Quibdó market, Cock Duque bought a dozen jagua fruits and brought them back to Medellín, thinking he might be able to turn the blue into a product. First, his company needed to figure out the chemistry. A young researcher on the team, Sandra Zapata, who grew up in the Chocó and has known the fruit since she was a child, took on the project in 2007 and later turned it into her Ph.D. thesis with Fernando Echeverri, a chemist specializing in natural products at the University of Antioquia.

Previous research suggested a compound called genipin produced the blue color by reacting with amino acids. (This is what happens in the skin when the juice is used for body painting.) Zapata worked on standardizing the process and making it cheaper. Glycine, the cheapest and simplest amino acid available, turned out to work well. In 2010, she and others received a patent in the United States for a "blue colorant derived from *Genipa americana* fruit." It spells out a simple process: mixing raw fruit juice with glycine and heating it.

Standardizing the color, however, was harder. Comparing fruits from different regions and even from the same tree, Zapata found that genipin concentrations varied widely, from 1% to 14%. "When I realized that I said: I need to change my strategy, I will work as an engineer." She developed a method for blending different batches to create a more consistent color.

Identifying the molecule produced by the reaction of genipin and glycine was another huge effort. After years of work, Zapata and Echeverri determined that the main compound is a polymer, a long chain of the same molecular building blocks repeated dozens of times. They received a U.S. patent for this and related compounds in 2016. Ironically, although the goal was to create a natural colorant, it could be patented only because the polymer itself is not found in nature and requires human intervention to create—such as painting it on skin or mixing it with glycine.

Cock Duque's company, EcoFlora Cares, has now spent more than a decade characterizing the dye and doing toxicology tests; it is now waiting for the U.S. Food and Drug Administration (FDA) to approve jagua blue as a food colorant. It has also applied for approval from the European Food Safety Authority and an intergovernmental body known as the Codex Alimentarius Commission, which would open the way for use in more than 100 other countries. "This has been an incredibly long process but we are hopefully nearing the end," Cock Duque says.

"It's very personal," Zapata says. The project consumed her life and tinged milestone memories blue. She remembers clearly the first time she got a nice blue from mixing the jagua fruit juice with glycine. "I was so happy," she says. Ten minutes later she got a phone call that her mom had died suddenly. Her son was born around the same time, and grew up eating jagua-colored ice cream and other treats she was working to develop.

Zapata's driving force was always what the project could mean for many people in the Chocó and Colombia in general, she says. "I realized that this project could really change the history of many communities here." That has not happened yet, she acknowledges. And even if the approvals come soon, there will be hurdles ahead that could jeopardize that goal. "So, this is like a very, very dangerous phase."

MOST OF ECOFLORA'S work in recent years has focused on areas outside the Chocó. In the hills around the small town of El Prodigio, about 3 hours' drive from Medellín, non-Indigenous farmers have long raised cattle and planted coffee, cocoa, and citrus fruits. Recently some of them have added the jagua. "In the beginning I thought it was crazy to plant jagua," says Libardo Diaz, a farmer who previously knew the tree only as one to chop down for wood. He is part of a program by EcoFlora that encourages farmers to plant about 1 hectare of jagua trees, providing seeds and technical assistance and promising to buy the fruits if and when the colorant enters the market.

"These are the first commercial jagua plantations in the world," says EcoFlora's Sergio Arango Arcila. "When we started a few years ago, I had a lot of questions," he says, emphasizing "lot": how close to plant the trees? How long until they bear fruit? "Now we have some answers," he says. The trees on Diaz's land were planted a little more than 3 years ago and should bear fruit next year. "We are all waiting for it," the farmer says. Keeping communities engaged in the yearslong process has been difficult given there are no profits to share yet, Arango Arcila says.

The plantations mark a departure from Cock Duque's original vision: using wild jagua trees in the Chocó as the sole source of the blue colorant. The company's patent application declared the fruits were obtained "through agreements with the ethnic communities" and that local communities would share in the profits. But collecting fruits from the wild trees and transporting them to the plant in Medellín turned out to be logistically challenging, Cock Duque says. "Scaling that up to the volumes that we will need is not economically feasible."

If production relies on plantations, that erases one benefit of the original scheme: creating a reason for communities to conserve these trees. And Indigenous communities could lose out on profits—which highlights the biggest question about jagua blue: Who benefits if it is successful? In the past, the exploitation of biodiversity in the Global South was "an Indiana Jones situation," says Maria Julia Oliva, an expert on genetic resources policy at the Union for Ethical Biotrade. "People would fly in, get something, go out—and maybe they would do something fantastic like develop a new cancer medicine. But there would be no recognition of the communities or the countries where the plant came from, the traditional knowledge on which some of that research would have been based, and no benefits flowing back to local development or biodiversity protection."

In recent years, Indiana Jones has been reined in. The Convention on Biological Diversity, the international conservation treaty enacted in 1993, established the principle that individual states have sovereign rights over the genetic resources found on their territory and that any benefits from their use should be shared with that country. The Nagoya Protocol, adopted in 2010, laid down how this sharing was to be conducted and enforced.

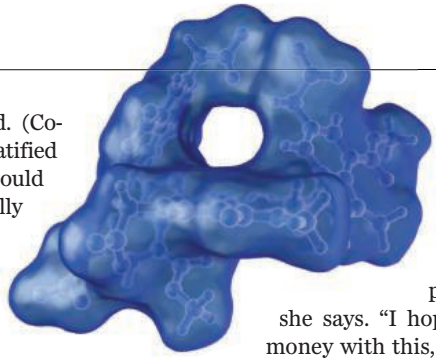
But in the early 2000s, when EcoFlora started its quest to commercialize the jagua blue, there was no template, Cock Duque says. So the company negotiated with local communities in the Chocó to get their consent to proceed and then negotiated and signed a deal with the Colombian government that allows the company to exploit the jagua in exchange for royalty payments to the government, which is supposed to ensure that any benefits are shared in a fair and equitable way.

Because it was the first of its kind in Colombia, the deal has been closely watched, says Oscar Lizarazo, a legal scholar who studies benefit sharing at the National University of Colombia, Bogotá. "This is probably the most important test case for this in Colombia." Lizarazo says the company has done far better than most by adhering to the spirit of the Nagoya Protocol even before it was adopted. (Colombia still has not ratified it.) But he says they could still do better, "especially by further recognizing Indigenous peoples."

EcoFlora has kept its original vision alive in Polines, buying some fruits the community harvests from wild trees and helping locals plant trees, too. "This can be an important source of income for us in the future," says Bailarin, whose father recently planted his first jagua trees. But at the moment there is no cooperation with other Emberá communities. And although EcoFlora has agreed to pay royalties both to the Colombian government and the University of Antioquia, there is no royalty agreement with the Emberá, which Colombian rules do not require. (One of the few such arrangements in the world was signed in 2019, when South Africa's rooibos tea industry agreed to pay significant royalties to the San and Khoikhoi people, who have traditionally used it to make tea and herbal medicines.)

Cock Duque says he hopes to include more Emberá communities in the future. But he argues that his company's obligations to the Emberá are limited. He says his discovery of jagua at the market in Quibdó was a moment of serendipity, and its use as a natural food colorant has little to do with the traditional knowledge of body painting.

That's not how Domicó sees it, though. Emberá traditional knowledge has clearly played an important role, she says. "I hope when EcoFlora makes money with this, they make sure that benefits also reach the Emberá."



A compound in jagua fruit called genipin reacts with glycine to form a long, stable polymer (shown here) that creates a coveted blue.

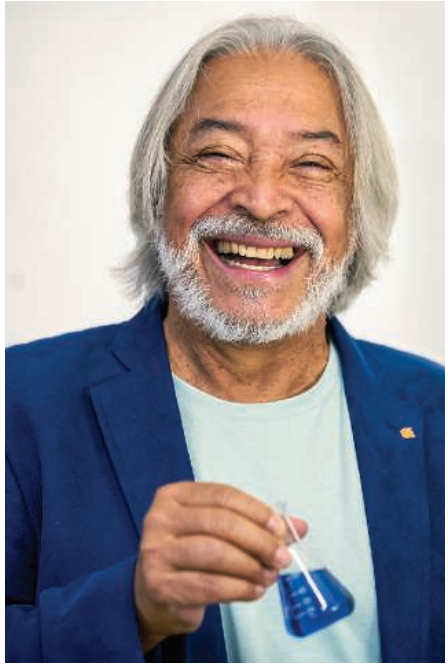
EVEN IF ECOFLORA can successfully grow and harvest the fruits, it faces another obstacle: competition. Although no natural blue food colorant was approved in the U.S. when EcoFlora submitted its application, two others have since been greenlit. FDA approved a crude extract from algae, called spirulina, in 2014 and an extract from the flowers of the butterfly pea in 2021. Both have disadvantages, Buchweitz says: Extracting the blue compound from the algae is a laborious process, and the anthocyanins from the butterfly pea aren't as stable, she says. "In terms of the color stability and the reproducibility of the color, this genipin-based blue seems to me the most likely to be successful in the future."

Even there EcoFlora has competition. Months before the patent for its blue colorant was granted, a company called Wild Flavors in Kentucky patented a blue dye made by mixing jagua juice with juice from watermelons or other fruits. Agribusiness giant ADM bought the company in 2014 for $3 billion and has been selling the colorant as "huito blue" and sourcing the fruit from Peru. (The company does not need FDA approval because fruit juices are exempt.) And a genipin-derived blue from gardenia fruits that has long been used in parts of Asia could soon enter the U.S. and international markets, too.

After more than 15 years of work, the fate of jagua blue remains uncertain. If the product fails it would be a disaster for many people, Zapata says. "It would be crazy, crazy, crazy for the company, the communities, the investors," she says.

But Lizarazo argues the project could end up having a positive impact even then. For one, it has probably helped foster a closer relationship between academia and industry in Colombia. And more important, it helps set a standard for how to take access and benefit sharing seriously when using natural resources, he says. "It has been an opportunity for the company, the communities, and the government to learn how to put these rules into practice."

It also shows how much still needs to be done, Oliva says. Even if jagua blue succeeds and has the benefits its developers hope it will, that's only one small victory, she says. "What we really need is for these types of practices to become mainstream." ■



Fernando Echeverri spent years working to uncover the chemistry behind the blue dye made from jagua fruit.

IMAGES: (TOP TO BOTTOM) A. FISHER/*SCIENCE*; CRISTINA ABAD ANGEL

PERSPECTIVES

ECOLOGY

Potential for recovery of declining reef sharks

Data on shark populations in coral reefs raise concern and hope for recovery

By **David S. Shiffman**[1,2]

Sharks and their relatives are some of the most threatened vertebrates on Earth, with approximately one-third estimated or assessed as threatened with extinction (*1*). This is a major problem because as predators that help keep the food web in balance, these animals play a variety of vitally important ecological roles (*2*) and in doing so help to keep healthy many ecosystems that humans depend on. Coral reefs provide homes for countless fish species that are vital for fisheries and are therefore an especially important ecosystem for humans—and one where the decline of shark populations seems to be especially acute (*3*). On page 1155 of this issue, Simpfendorfer *et al.* (*4*) report the results of a species-level and reef-level analysis of common resident reef sharks across the world. They show startling declines of once-common reef shark species but also signs of hope that these populations can recover with the right protection.

The study by Simpfendorfer *et al.* is the result of a worldwide collaboration called the Global FinPrint project. The data analyzed include more than 20,000 hours of standardized underwater video taken at nearly 400 reefs in 67 countries and territories around the world—that is nearly 3 years of raw video. The baited remote underwater video stations (BRUVSs) used by FinPrint are a simple but powerful tool. They are essentially underwater camera traps that consist of a small quantity of bait suspended in front of a camera. In addition to being good at documenting the presence and absence and the behavior of different marine organisms (*5*), they also generate high-definition images and video of marine life that are tailor-made for public education about what lives in the threatened habitats off our coastlines.

The results of Simpfendorfer *et al.* reveal declines of 60 to 73% of once-abundant coral reef shark species at reefs around the world. This adds to a large and growing volume of similarly alarming conclusions about the global conservation status of sharks and their relatives. The global conservation status of sharks and rays is worse than a decade ago (*6*) and is even more concerning for some groups of sharks (*7*). Sharks caught as bycatch in global tuna fisheries are declining in population even as those same tuna are rebounding (*8*).

However, the findings of Simpfendorfer *et al.* include signs of hope and a clear path forward. Their results show that although shark populations in many reefs had declined, some healthy reef shark populations remained. The reefs with healthier shark populations had some important similarities: They tended to be in the waters of high-income countries with stronger natural resource management regulations, participatory natural resource management (where citizens have the right to petition the government about changes in natural resource management policy), and resources for enforcing the rules. Unfortunately, such countries are relatively rare, and lower-income countries tend to have fewer resources for sustainable management and enforcement. These observations show that conservation problems involve solving human problems as well as those associated with ecology; a country that lacks the resources to feed its people is less able to sustainably manage and protect its biodiversity.

Science-based, well-enforced marine protected areas—in which harmful fishing practices are restricted or banned—also tended to have healthier reefs. However, Simpfendorfer *et al.* suggest that some highly touted shark conservation solutions were enacted in places where there were not many threats to the shark population to begin with and advise caution in interpreting the success of those solutions. For example, the British Virgin Islands shark sanctuary bans all commercial shark fishing in its territorial waters, but between 1950 and the 2014 establishment of the sanctuary, only 3 tons of shark were fished from those waters (*9*, *10*), suggesting that there was not much of a shark fishery to ban. Another sanctuary was established in the Bahamas in 2011 but decades after the most common shark fishing gear was already banned, suggesting that the country's relatively high shark population is most likely due to the older, less-hyped regulation (*11*).

The most unexpected result of the study by Simpfendorfer *et al.* is that a decline or complete loss of shark species in one reef was not always associated with similar changes

[1]Arizona State University New College of Interdisciplinary Arts and Sciences, Glendale, AZ, USA. [2]Consortium for Science Policy Outcomes, Arizona State University Washington, Washington, DC, USA. Email: david.shiffman@gmail.com

Numbers of reef sharks, such as this Caribbean reef shark (*Carcharhinus perezi*) in the Bahamas, have fallen overall, but some healthy populations remain.

in nearby reefs. They found that one reef can be overfished so badly that a once-common reef shark species is totally gone, but another reef a short distance away can have healthy populations of that same species. Strong, effective management (including but not limited to no-fishing-allowed marine protected areas) on one reef protected local species, even while their populations on neighboring reefs faced collapse. The presence of these possible future "source" populations—that is, healthy populations that can eventually help repopulate nearby areas—gives hope that if the threats that led to population decline are resolved, then these important and threatened animals may recover.

The study by Simpfendorfer *et al.* also demonstrates the growing importance of global collaboration. Global problems require huge multidisciplinary teams because scientists or laboratories working by themselves simply cannot generate or analyze data on this scale. In many ways, FinPrint has been a model for international collaboration. Such studies have documented how the decline of sharks leads to increasing abundance of mesopredatory fishes such as moray eels (*12*), what aspects of marine protected area design are most effective for sharks (*13*), and much more—and there is more to learn.

The problem is clear—animals that provide ecosystem services that are vital for human food security and livelihoods are disappearing at an alarming rate, overwhelmingly owing to bad management practices that allow unsustainable overfishing of these ecologically important and biologically vulnerable creatures. The loss of sharks and the ecosystem services they provide represents an ecological disaster that can cause substantial harm to humans. Action must be taken to prevent further population declines and allow rebuilding of depleted populations before it is too late. ■

REFERENCES AND NOTES

1. N. K. Dulvy *et al.*, *Curr. Biol.* **31**, 4773 (2021).
2. M. R. Heithaus, A. Frid, A. J. Wirsing, B. Worm, *Trends Ecol. Evol.* **23**, 202 (2008).
3. M. A. MacNeil *et al.*, *Nature* **583**, 801 (2020).
4. C. A. Simpfendorfer *et al.*, *Science* **380**, 1155 (2023).
5. C. S. Sherman, M. R. Heupel, S. K. Moore, A. Chin, C. A. Simpfendorfer, *Mar. Ecol. Prog. Ser.* **641**, 145 (2020).
6. N. K. Dulvy *et al.*, *eLife* **3**, e00590 (2014).
7. N. Pacoureau *et al.*, *Nature* **589**, 567 (2021).
8. M. J. Juan-Jordá *et al.*, *Science* **378**, eabj0211 (2022).
9. C. A. Ward-Paige, *Mar. Policy* **82**, 87 (2017).
10. L. N. Davidson, M. A. Krawchuk, N. K. Dulvy, *Fish Fish.* **17**, 438 (2016).
11. D. S. Shiffman, N. Hammerschlag, *Anim. Conserv.* **19**, 401 (2016).
12. G. M. Clementi *et al.*, *iScience* **24**, 102097 (2021).
13. J. S. Goetze *et al.*, *Glob. Change Biol.* **27**, 3432 (2021).

10.1126/science.adi5759

NEUROSCIENCE

Epigenetic changes in astrocytes make sense

Serotonin induces gene expression changes in astrocytes to regulate olfactory behavior

By **Flora Vasile**[1,2] *and* **Nathalie Rouach**[1]

The ever-changing nature of the world requires the brain to constantly adapt to its environment to optimize behavioral output at different spatial and temporal scales. Although short-term synaptic and circuit plasticity materializes as changes in synaptic transmission, longer-term plasticity can be supported by transcriptional and epigenetic modifications. This process has been well described in neurons, but whether and how non-neuronal cells also undergo such modifications to promote flexible neuronal circuits underlying behavior is unknown. On page 1146 of this issue, Sardar *et al.* (*1*) report a newly identified role for the monoamine neurotransmitter serotonin in inducing epigenetic histone modifications in astrocytes that alter inhibitory neurotransmitter signaling and ultimately sculpt olfactory processing and behavior in mice.

"...a newly identified role for the monoamine neurotransmitter serotonin in inducing epigenetic histone modifications in astrocytes..."

Astrocytes are an integral part of brain circuits: They undergo intracellular calcium fluctuations in response to physiological sensory-driven neuronal activity (*2*) that elicit the calcium-dependent release of neuroactive molecules that affect neural activity, plasticity, and behavior (*3*). It is becoming evident that the influence of astrocytes on neurons is multifold (*4*) and is specific to developmental stages, brain regions, activity regimes, and pathophysiological states. Hence, the regulation of neuronal activity by astrocytes is remarkably specific yet flexible. Transcriptional modifications have recently been identified as a mechanism for long-lasting adaptations in the modulatory program of astrocytes. Building on the first astroglial transcriptome database (*5*), studies have since characterized the development-, region-, and activity-dependent properties of the astrocyte transcriptome. Additionally, specific transcription factors that are involved in supervising astrocyte gene expression and ultimately neuronal functions have been revealed. Yet the intricacies of neuronal activity–dependent transcriptomic plasticity in astrocytes and the mechanisms underlying downstream neuronal circuit and behavioral alterations have remained obscure.

Sardar *et al.* identify a mechanism through which neuronal activity induces transcriptomic changes in astrocytes, which in turn drive epigenetic changes and regulate olfactory sensory processing in mice. They show that neuronal activity, artificially evoked using chemogenetics or physiologically triggered through olfactory stimulation, induces widespread changes in gene expression in astrocytes in a brain region–dependent manner. The expression of immediate early genes—a set of genes that are rapidly and transiently expressed in response to various stimuli—were specific to astrocytes and included transcription factors such as SOX9. The evoked neuronal activity increases binding of SOX9 to the promoter of the solute carrier family 22 member 3 (*Slc22a3*) gene. This gene encodes the monoamine transporter organic cation transporter 3 (OCT3), which contributes to the uptake of serotonin. SOX9-mediated up-regulation of *Slc22a3* expression was specific to the olfactory bulb; deleting *Sox9* did not affect *Slc22a3* expression in the cortex or hippocampus. This up-regulation in OCT3 expression increases levels of astroglial serotonin, which enters the nucleus to drive histone serotonylation (see the figure). Because SOX9 regulates the expression of various genes, it would be interesting to further investigate the downstream effects of SOX9-mediated gene regulation.

What is the role of such histone modification? Using conditional deletion of *Slc22a3*, Sardar *et al.* show that *Slc22a3* up-regulation

[1]Center for Interdisciplinary Research in Biology, College de France, CNRS, INSERM, Labex Memolife, Université PSL, Paris, France. [2]Champalimaud Neuroscience Programme, Champalimaud Foundation, Lisbon, Portugal.
Email: nathalie.rouach@college-de-france.fr

Numbers of reef sharks, such as this Caribbean reef shark (*Carcharhinus perezi*) in the Bahamas, have fallen overall, but some healthy populations remain.

in nearby reefs. They found that one reef can be overfished so badly that a once-common reef shark species is totally gone, but another reef a short distance away can have healthy populations of that same species. Strong, effective management (including but not limited to no-fishing-allowed marine protected areas) on one reef protected local species, even while their populations on neighboring reefs faced collapse. The presence of these possible future "source" populations—that is, healthy populations that can eventually help repopulate nearby areas—gives hope that if the threats that led to population decline are resolved, then these important and threatened animals may recover.

The study by Simpfendorfer *et al.* also demonstrates the growing importance of global collaboration. Global problems require huge multidisciplinary teams because scientists or laboratories working by themselves simply cannot generate or analyze data on this scale. In many ways, FinPrint has been a model for international collaboration. Such studies have documented how the decline of sharks leads to increasing abundance of mesopredatory fishes such as moray eels (*12*), what aspects of marine protected area design are most effective for sharks (*13*), and much more—and there is more to learn.

The problem is clear—animals that provide ecosystem services that are vital for human food security and livelihoods are disappearing at an alarming rate, overwhelmingly owing to bad management practices that allow unsustainable overfishing of these ecologically important and biologically vulnerable creatures. The loss of sharks and the ecosystem services they provide represents an ecological disaster that can cause substantial harm to humans. Action must be taken to prevent further population declines and allow rebuilding of depleted populations before it is too late. ■

REFERENCES AND NOTES

1. N. K. Dulvy *et al.*, *Curr. Biol.* **31**, 4773 (2021).
2. M. R. Heithaus, A. Frid, A. J. Wirsing, B. Worm, *Trends Ecol. Evol.* **23**, 202 (2008).
3. M. A. MacNeil *et al.*, *Nature* **583**, 801 (2020).
4. C. A. Simpfendorfer *et al.*, *Science* **380**, 1155 (2023).
5. C. S. Sherman, M. R. Heupel, S. K. Moore, A. Chin, C. A. Simpfendorfer, *Mar. Ecol. Prog. Ser.* **641**, 145 (2020).
6. N. K. Dulvy *et al.*, *eLife* **3**, e00590 (2014).
7. N. Pacoureau *et al.*, *Nature* **589**, 567 (2021).
8. M. J. Juan-Jordá *et al.*, *Science* **378**, eabj0211 (2022).
9. C. A. Ward-Paige, *Mar. Policy* **82**, 87 (2017).
10. L. N. Davidson, M. A. Krawchuk, N. K. Dulvy, *Fish Fish.* **17**, 438 (2016).
11. D. S. Shiffman, N. Hammerschlag, *Anim. Conserv.* **19**, 401 (2016).
12. G. M. Clementi *et al.*, *iScience* **24**, 102097 (2021).
13. J. S. Goetze *et al.*, *Glob. Change Biol.* **27**, 3432 (2021).

10.1126/science.adi5759

NEUROSCIENCE

Epigenetic changes in astrocytes make sense

Serotonin induces gene expression changes in astrocytes to regulate olfactory behavior

By **Flora Vasile**[1,2] *and* **Nathalie Rouach**[1]

The ever-changing nature of the world requires the brain to constantly adapt to its environment to optimize behavioral output at different spatial and temporal scales. Although short-term synaptic and circuit plasticity materializes as changes in synaptic transmission, longer-term plasticity can be supported by transcriptional and epigenetic modifications. This process has been well described in neurons, but whether and how non-neuronal cells also undergo such modifications to promote flexible neuronal circuits underlying behavior is unknown. On page 1146 of this issue, Sardar *et al.* (*1*) report a newly identified role for the monoamine neurotransmitter serotonin in inducing epigenetic histone modifications in astrocytes that alter inhibitory neurotransmitter signaling and ultimately sculpt olfactory processing and behavior in mice.

"...a newly identified role for the monoamine neurotransmitter serotonin in inducing epigenetic histone modifications in astrocytes..."

Astrocytes are an integral part of brain circuits: They undergo intracellular calcium fluctuations in response to physiological sensory-driven neuronal activity (*2*) that elicit the calcium-dependent release of neuroactive molecules that affect neural activity, plasticity, and behavior (*3*). It is becoming evident that the influence of astrocytes on neurons is multifold (*4*) and is specific to developmental stages, brain regions, activity regimes, and pathophysiological states. Hence, the regulation of neuronal activity by astrocytes is remarkably specific yet flexible. Transcriptional modifications have recently been identified as a mechanism for long-lasting adaptations in the modulatory program of astrocytes. Building on the first astroglial transcriptome database (*5*), studies have since characterized the development-, region-, and activity-dependent properties of the astrocyte transcriptome. Additionally, specific transcription factors that are involved in supervising astrocyte gene expression and ultimately neuronal functions have been revealed. Yet the intricacies of neuronal activity–dependent transcriptomic plasticity in astrocytes and the mechanisms underlying downstream neuronal circuit and behavioral alterations have remained obscure.

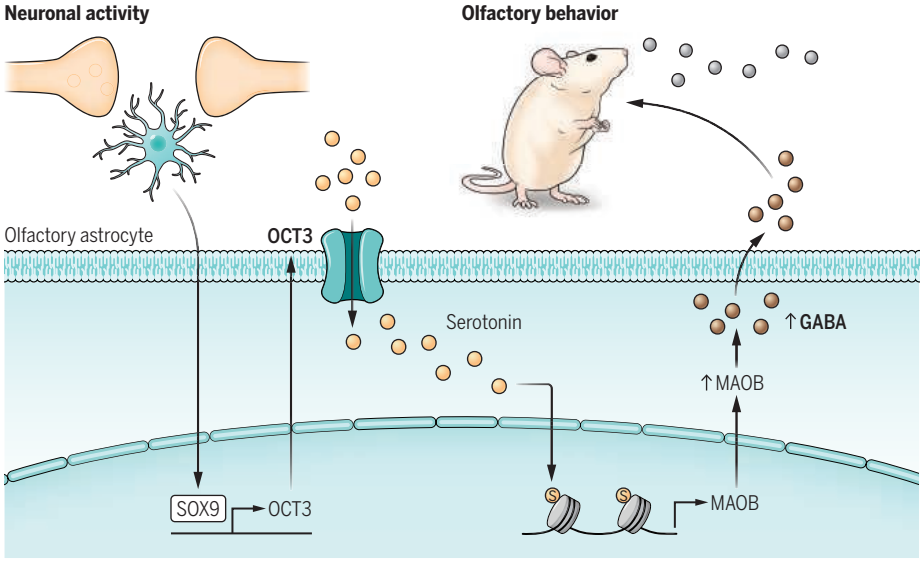
Sardar *et al.* identify a mechanism through which neuronal activity induces transcriptomic changes in astrocytes, which in turn drive epigenetic changes and regulate olfactory sensory processing in mice. They show that neuronal activity, artificially evoked using chemogenetics or physiologically triggered through olfactory stimulation, induces widespread changes in gene expression in astrocytes in a brain region–dependent manner. The expression of immediate early genes—a set of genes that are rapidly and transiently expressed in response to various stimuli—were specific to astrocytes and included transcription factors such as SOX9. The evoked neuronal activity increases binding of SOX9 to the promoter of the solute carrier family 22 member 3 (*Slc22a3*) gene. This gene encodes the monoamine transporter organic cation transporter 3 (OCT3), which contributes to the uptake of serotonin. SOX9-mediated up-regulation of *Slc22a3* expression was specific to the olfactory bulb; deleting *Sox9* did not affect *Slc22a3* expression in the cortex or hippocampus. This up-regulation in OCT3 expression increases levels of astroglial serotonin, which enters the nucleus to drive histone serotonylation (see the figure). Because SOX9 regulates the expression of various genes, it would be interesting to further investigate the downstream effects of SOX9-mediated gene regulation.

What is the role of such histone modification? Using conditional deletion of *Slc22a3*, Sardar *et al.* show that *Slc22a3* up-regulation

[1]Center for Interdisciplinary Research in Biology, College de France, CNRS, INSERM, Labex Memolife, Université PSL, Paris, France. [2]Champalimaud Neuroscience Programme, Champalimaud Foundation, Lisbon, Portugal.
Email: nathalie.rouach@college-de-france.fr

Epigenetic changes in astrocytes contribute to olfaction

Neuronal activity alters gene expression in astrocytes, including up-regulating the expression of the transcription factor SOX9, which drives the expression of the monoamine transporter organic cation transporter 3 (OCT3). This increases the uptake of serotonin released from neurons and histone serotonylation (S). This epigenetic modification up-regulates the expression of monoamine oxidase B (MAOB), a γ-aminobutyric acid (GABA) biosynthesis enzyme. GABA is then released by astrocytes to inhibit synaptic transmission and regulate olfactory behavior.



induced by histone serotonylation increases the expression of monoamine oxidase B (MAOB), a biosynthetic enzyme of the inhibitory neurotransmitter γ-aminobutyric acid (GABA). Increased MAOB expression was associated with enhanced levels of GABA and its release by astrocytes, leading to increases in inhibitory synaptic transmission. Behaviorally, the GABA-mediated olfactory sensory processing allows responses to scents and ultimately contributes to odor detection and discrimination in mice. This translated to an enhancement of the ability of mice to detect new scents present at low concentration and to distinguish between structurally similar odorants.

Notably, altering the expression of OCT3 modifies the astrocyte phenotype at multiple levels, with changes in transcriptome, calcium signaling, morphology, and GABA release. This raises the questions of whether and how these different effects individually or synergistically contribute to controlling sensory processing. Astrocyte serotoninergic signaling is implicated in numerous physiological and pathological processes, including circadian rhythms, mood regulation, pain, stress, anxiety, and depression (*6*). Identifying the extent to which this transcriptomic pathway is involved in the underlying regulatory and dysregulatory mechanisms of these various processes would be of particular interest. Histone serotonylation was initially identified in serotonin-producing neurons and was shown to be permissive for gene expression (*7*). However, the role of these gene expression changes in physiological processes is unknown. The regulation of olfaction by astrocytes undergoing this same histone modification after serotonin uptake through transporters (*8*) thus expands the scale of action of this epigenetic modification.

What may be the advantage of such activity-dependent epigenetic regulation in several cell types? Considering that an astrocyte contacts hundreds of thousands of synapses (*9*), downstream effects of histone serotonylation could be broadcast to a large neuronal population. This is particularly suitable for the circuit-wide plasticity that is needed to discriminate scents. Could other neuromodulators, such as catecholamines and histamine, also modulate changes that adaptively affect synaptic and circuit functions? Both the potential diversity in histone monoaminylation and regional specificities highlighted in the study of Sardar *et al.* could support the tailored regulation of gene expression according to local activity-dependent needs. ■

REFERENCES AND NOTES

1. D. Sardar *et al.*, *Science* **380**, eade0027 (2023).
2. N. Bazargani, D. Attwell, *Nat. Neurosci.* **19**, 182 (2016).
3. S. Guerra-Gomes, N. Sousa, L. Pinto, J. F. Oliveira, *Front. Cell. Neurosci.* **11**, 427 (2018).
4. G. Dallérac, J. Zapata, N. Rouach, *Nat. Rev. Neurosci.* **19**, 729 (2018).
5. J. D. Cahoy *et al.*, *J. Neurosci.* **28**, 264 (2008).
6. X. Zhou *et al.*, *Front. Mol. Neurosci.* **12**, 136 (2019).
7. L. A. Farrelly *et al.*, *Nature* **567**, 535 (2019).
8. M. Inazu *et al.*, *Neurochem. Int.* **39**, 39 (2001).
9. E. A. Bushong, M. E. Martone, Y. Z. Jones, M. H. Ellisman, *J. Neurosci.* **22**, 183 (2002).

10.1126/science.adi5765

CHEMISTRY

A big breakthrough for beryllium

A stable organometallic compound with a Be-Be bond has been synthesized

By **Jason L. Dutton**

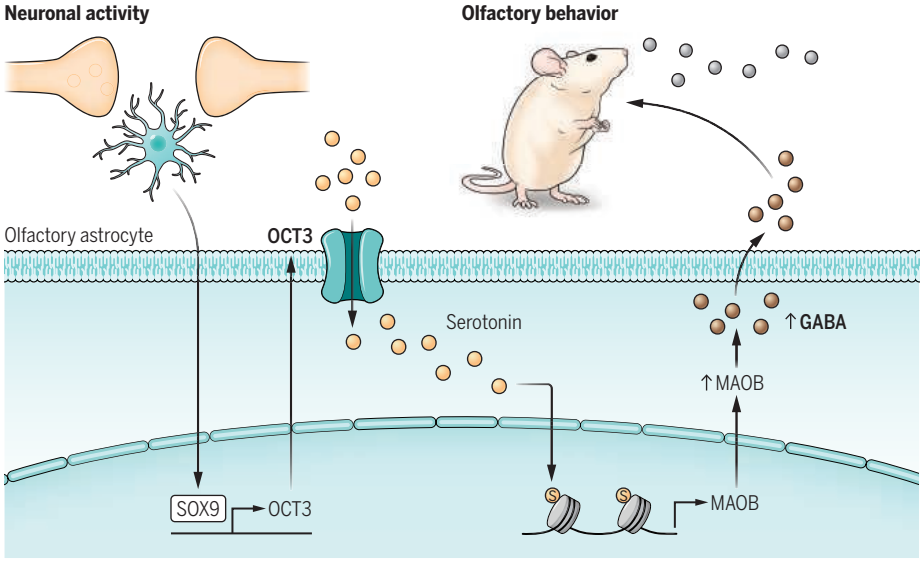
The investigation of organometallic compounds containing unsupported homoatomic metal-metal (M-M) bonds has been an area of major interest for decades. These compounds feature distinct, otherwise inaccessible bonding, such as M-M quadruple and quintuple bonds, and fascinating reactivity, including mimicking the reactions of C-C multiple bonds. Ultimately, the main driver for research in this area is to push the boundaries of bonding for a given element and, in doing so, rewrite the textbooks. On page 1147 of this issue, Boronski *et al.* (*1*) report the isolation of diberyllocene as the latest entrant in the field. Diberyllocene is a stable Be-Be bonded compound and is a relatively accessible source of nucleophilic beryllium, which has the potential to unlock the reactivity of organoberyllium with a vast array of new substrates.

The most conspicuous appearance of beryllium is as a constituent in the mineral beryl, of which emeralds are one variety. As an element, it is used as an x-ray window—being a very light metal (atomic number 4), beryllium is largely transparent to x-rays—and it is also used as a metal or alloy in applications demanding lightweight strength. From a chemistry perspective, beryllium is the most electronegative element in the s-block of the periodic table; because of this, combined with being a light second-row element, it is expected that beryllium should display the richest chemistry of the s-block, with capabilities for covalent bonding not found in the heavier metals. However, the chemistry of beryllium is dominated by Be^{2+} or molecular compounds in the 2+ oxidation state through loss of the 2s electrons. A good example of the opportunity that beryllium offers is the recent discovery of a Be-N

Department of Biochemistry and Chemistry, La Trobe University, Melbourne, VIC, Australia.
Email: j.dutton@latrobe.edu.au

Epigenetic changes in astrocytes contribute to olfaction

Neuronal activity alters gene expression in astrocytes, including up-regulating the expression of the transcription factor SOX9, which drives the expression of the monoamine transporter organic cation transporter 3 (OCT3). This increases the uptake of serotonin released from neurons and histone serotonylation (S). This epigenetic modification up-regulates the expression of monoamine oxidase B (MAOB), a γ-aminobutyric acid (GABA) biosynthesis enzyme. GABA is then released by astrocytes to inhibit synaptic transmission and regulate olfactory behavior.



induced by histone serotonylation increases the expression of monoamine oxidase B (MAOB), a biosynthetic enzyme of the inhibitory neurotransmitter γ-aminobutyric acid (GABA). Increased MAOB expression was associated with enhanced levels of GABA and its release by astrocytes, leading to increases in inhibitory synaptic transmission. Behaviorally, the GABA-mediated olfactory sensory processing allows responses to scents and ultimately contributes to odor detection and discrimination in mice. This translated to an enhancement of the ability of mice to detect new scents present at low concentration and to distinguish between structurally similar odorants.

Notably, altering the expression of OCT3 modifies the astrocyte phenotype at multiple levels, with changes in transcriptome, calcium signaling, morphology, and GABA release. This raises the questions of whether and how these different effects individually or synergistically contribute to controlling sensory processing. Astrocyte serotoninergic signaling is implicated in numerous physiological and pathological processes, including circadian rhythms, mood regulation, pain, stress, anxiety, and depression (*6*). Identifying the extent to which this transcriptomic pathway is involved in the underlying regulatory and dysregulatory mechanisms of these various processes would be of particular interest. Histone serotonylation was initially identified in serotonin-producing neurons and was shown to be permissive for gene expression (*7*). However, the role of these gene expression changes in physiological processes is unknown. The regulation of olfaction by astrocytes undergoing this same histone modification after serotonin uptake through transporters (*8*) thus expands the scale of action of this epigenetic modification.

What may be the advantage of such activity-dependent epigenetic regulation in several cell types? Considering that an astrocyte contacts hundreds of thousands of synapses (*9*), downstream effects of histone serotonylation could be broadcast to a large neuronal population. This is particularly suitable for the circuit-wide plasticity that is needed to discriminate scents. Could other neuromodulators, such as catecholamines and histamine, also modulate changes that adaptively affect synaptic and circuit functions? Both the potential diversity in histone monoaminylation and regional specificities highlighted in the study of Sardar *et al.* could support the tailored regulation of gene expression according to local activity-dependent needs. ■

REFERENCES AND NOTES

1. D. Sardar *et al.*, *Science* **380**, eade0027 (2023).
2. N. Bazargani, D. Attwell, *Nat. Neurosci.* **19**, 182 (2016).
3. S. Guerra-Gomes, N. Sousa, L. Pinto, J. F. Oliveira, *Front. Cell. Neurosci.* **11**, 427 (2018).
4. G. Dallérac, J. Zapata, N. Rouach, *Nat. Rev. Neurosci.* **19**, 729 (2018).
5. J. D. Cahoy *et al.*, *J. Neurosci.* **28**, 264 (2008).
6. X. Zhou *et al.*, *Front. Mol. Neurosci.* **12**, 136 (2019).
7. L. A. Farrelly *et al.*, *Nature* **567**, 535 (2019).
8. M. Inazu *et al.*, *Neurochem. Int.* **39**, 39 (2001).
9. E. A. Bushong, M. E. Martone, Y. Z. Jones, M. H. Ellisman, *J. Neurosci.* **22**, 183 (2002).

10.1126/science.adi5765

CHEMISTRY

A big breakthrough for beryllium

A stable organometallic compound with a Be-Be bond has been synthesized

By **Jason L. Dutton**

The investigation of organometallic compounds containing unsupported homoatomic metal-metal (M-M) bonds has been an area of major interest for decades. These compounds feature distinct, otherwise inaccessible bonding, such as M-M quadruple and quintuple bonds, and fascinating reactivity, including mimicking the reactions of C-C multiple bonds. Ultimately, the main driver for research in this area is to push the boundaries of bonding for a given element and, in doing so, rewrite the textbooks. On page 1147 of this issue, Boronski *et al.* (*1*) report the isolation of diberyllocene as the latest entrant in the field. Diberyllocene is a stable Be-Be bonded compound and is a relatively accessible source of nucleophilic beryllium, which has the potential to unlock the reactivity of organoberyllium with a vast array of new substrates.

The most conspicuous appearance of beryllium is as a constituent in the mineral beryl, of which emeralds are one variety. As an element, it is used as an x-ray window—being a very light metal (atomic number 4), beryllium is largely transparent to x-rays—and it is also used as a metal or alloy in applications demanding lightweight strength. From a chemistry perspective, beryllium is the most electronegative element in the s-block of the periodic table; because of this, combined with being a light second-row element, it is expected that beryllium should display the richest chemistry of the s-block, with capabilities for covalent bonding not found in the heavier metals. However, the chemistry of beryllium is dominated by Be^{2+} or molecular compounds in the 2+ oxidation state through loss of the 2s electrons. A good example of the opportunity that beryllium offers is the recent discovery of a Be-N

Department of Biochemistry and Chemistry, La Trobe University, Melbourne, VIC, Australia.
Email: j.dutton@latrobe.edu.au

double bond (*2*). Pi-bonding is a common feature in nitrogen-containing compounds, but only by using beryllium can such a bond be accessed between nitrogen and an s-block element. Despite these features, beryllium is one of the least-explored elements in the entire periodic table. This is because of safety concerns surrounding the toxicity of beryllium. The primary risk factor is the inhalation of beryllium-containing dust, for which even very small quantities can result in both an acute and a chronic immune response, leading to potentially fatal scarring of the lungs. However, there is an emerging recognition that beryllium chemistry can be done safely in academic settings with proper precautions and controls (*3*).

To avoid safety concerns, advances in beryllium chemistry are often preceded by theoretical prediction using computational chemistry. Diberyllocene was first suggested to be a stable molecule and an interesting synthetic target in a 2005 theoretical study (*4*). The synthesis of diberyllocene that is reported by Boronski *et al.* takes advantage of one of the other great breakthroughs in s-block chemistry—the Jones reagent, which is a stable Mg(I) compound (*5*). In addition to being a stable group 2 complex with an M-M bond, the Jones reagent has found widespread use as a powerful, selective, and soluble reducing agent. Notably, a previous attempt to generate a Be-Be complex by making an analog of the Jones reagent was undone by a side reaction of the ligand (*6*). Boronski *et al.* showed that reduction of beryllocene with the Jones reagent in toluene followed by purification by sublimation resulted in 85% yield of diberyllocene. Crystallographic characterization revealed that diberyllocene has a similar structure to dizincocene (*7*). This is not unexpected because zinc is one of the most similar elements to beryllium in terms of valence and electronegativity, with the main difference being shorter bonds in diberyllocene owing to the smaller atomic radius of beryllium.

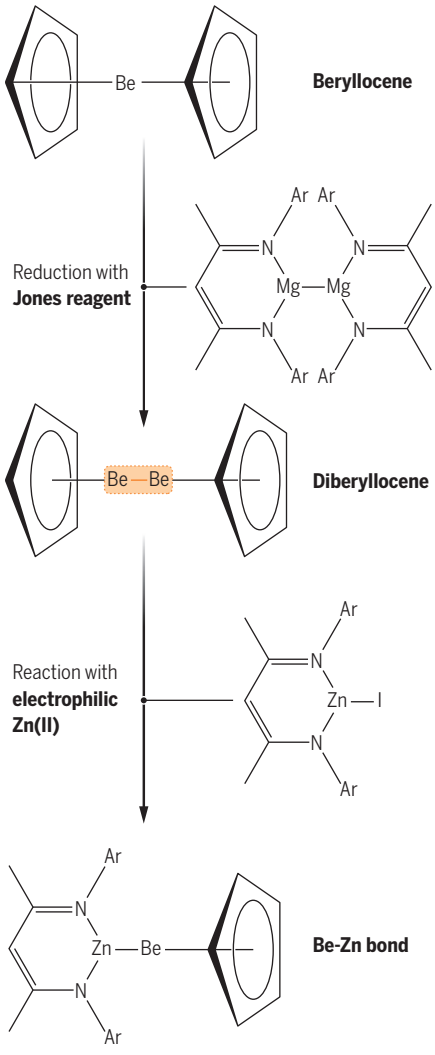
There are often questions asked in the field about whether an M-M compound is actually a covalent bond and not a bridging hydride, where rather than an M-M bond, the metals are held together by hydrogen atoms between them. There has been an instance where a purported homoatomic M-M bond in dicobaltocene (*8*) was later revealed to be a bridging tetrahydride (*9*). Diberyllocene is structurally similar to dicobaltocene, and bridging hydrides are a well-established feature in organoberyllium chemistry (*10*, *11*). The detection of metal-hydrides by x-ray crystallography is often hampered by the short M-H bonds and the high electron density of the metal, which can obscure hydrogen atoms, but this is not an issue for diberyllocene because beryllium has few electrons. Indeed, the x-ray structure for diberyllocene obtained by Boronski *et al.* shows no anomalous electron density where hydrides would be located, and the C-H hydrogens are clearly apparent in the electron density map. Nuclear magnetic resonance (NMR) and infrared (IR) spectroscopy also indicate an absence of potential Be-H.

It can be challenging to describe the bonding in M-M complexes as well as to conclusively determine the oxidation state of the metals. The Ga-Ga triple bond is an example of a system where there has been extensive discussion on the correct description of the M-M bond, with opinions ranging from there being a single bond to there obviously being a triple bond, and others settling in between (*12*). Regarding ambiguity in assigning oxidation states, a relevant example is the report of a Be(0) compound, cyclic alkyl(amino)carbene–bound cAAC-Be-cAAC (*13*). The assignment of the oxidation state of beryllium as 0 has been questioned, with it being suggested that beryllium should be considered in the +2 oxidation state with reduced cyclic alkyl amino carbene ligands each carrying an additional electron, but the matter has remained unresolved (*14*, *15*). The analyses by Boronski *et al.* establish that diberyllocene shows a clear Be-Be single bond and occupation of the 2s orbital of each beryllium with one electron through density functional theory (DFT) calculations. The assignment of the +1 oxidation state for beryllium is also uncontroversial because the cyclopentadienyl (Cp) ligands bound to the beryllium atoms cannot accept more electrons. ^{9}Be NMR spectroscopy experimentally supports the assignment, with diberyllocene featuring an unusually upfield chemical shift of –28 parts per million, which is indicative of very electron-rich beryllium.

Preliminary reactivity studies with diberyllocene with L-Al-I and L-Zn-I complexes gave L-Al-Be-Cp and L-Zn-Be-Cp (see the figure). These reactions demonstrate that diberyllocene can act as a nucleophilic equivalent of Be^{+}, in contrast to virtually all other sources of beryllium, which are electrophilic in nature. This will open the door to a huge range of potential bond-forming reactions between beryllium and electrophilic substrates to continue to explore the chemistry of this relatively unmapped element. ■

Synthesis and reactivity of diberyllocene

Reduction of Be(II) beryllocene using the Jones reagent furnishes Be(I) diberyllocene. This can be used to react with various electrophilic substrates, such as a Zn(II) complex to give a Be-Zn compound.



GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. J. T. Boronski, A. E. Crumpton, L. L. Wales, S. Aldridge, *Science* **380**, 1147 (2023).
2. G. Wang *et al.*, *Angew. Chem. Int. Ed.* **60**, 9407 (2021).
3. M. R. Buchner, M. Müller, *ACS Chem. Health Saf.* **30**, 36 (2023).
4. Y. Xie, H. F. Schaefer III, E. D. Jemmis, *Chem. Phys. Lett.* **402**, 414 (2005).
5. S. P. Green, C. Jones, A. Stasch, *Science* **318**, 1754 (2007).
6. M. Arrowsmith *et al.*, *Inorg. Chem.* **51**, 13408 (2012).
7. I. Resa, E. Carmona, E. Gutierrez-Puebla, A. Monge, *Science* **305**, 1136 (2004).
8. J. J. Schneider, R. Goddard, S. Werner, C. Krüger, *Angew. Chem. Int. Ed.* **30**, 1124 (1991).
9. J. L. Kersten *et al.*, *Angew. Chem. Int. Ed.* **31**, 1341 (1992).
10. M. Arrowsmith *et al.*, *Angew. Chem. Int. Ed.* **51**, 2098 (2012).
11. T. J. Hadlington, T. Szilvási, *Nat. Commun.* **13**, 461 (2022).
12. R. Dagani, *Chem. Eng. News* **76**, 31 (1998).
13. M. Arrowsmith *et al.*, *Nat. Chem.* **8**, 890 (2016).
14. M. Gimferrer *et al.*, *Chem. Sci.* **13**, 6583 (2022).
15. M. Gimferrer *et al.*, *Chem. Sci.* **14**, 384 (2023).

ACKNOWLEDGMENTS

J.L.D. is supported by the Australian Research Council (FT16010007 and DP20010013).

10.1126/science.adi5762

SOCIAL SCIENCE

AI and the transformation of social science research

Careful bias management and data fidelity are key

By **Igor Grossmann[1,2], Matthew Feinberg[3], Dawn C. Parker[2,4], Nicholas A. Christakis[5], Philip E. Tetlock[6], William A. Cunningham[7,8,9]**

Advances in artificial intelligence (AI), particularly large language models (LLMs), are substantially affecting social science research. These transformer-based machine-learning models pretrained on vast amounts of text data are increasingly capable of simulating human-like responses and behaviors (*1*, *2*), offering opportunities to test theories and hypotheses about human behavior at great scale and speed. This presents urgent challenges: How can social science research practices be adapted, even reinvented, to harness the power of foundational AI? And how can this be done while ensuring transparent and replicable research?

Social sciences rely on a range of methods, including questionnaires, behavioral tests, mixed-method analyses of semi-structured responses, agent-based modeling (ABM), observational studies, and experiments. The common goal is to obtain a generalized representation of characteristics of individuals, groups, cultures, and their dynamics (*2*). With the advent of advanced AI systems, the landscape of data collection in social sciences may shift. LLMs take advantage of deep learning to capture complex relationships within language. Such language literacy capabilities in processing, generating, and interacting with human language in a contextually aware and semantically accurate fashion (*1*) represent a major shift from previous AI approaches, which often struggled with such nuanced aspects of language as irony, metaphor, or emotional tone. With proper conditioning (*3*), LLMs can more accurately simulate human behavioral responses in social science research.

LLMs may supplant human participants for data collection. For example, LLMs have already demonstrated their ability to generate realistic survey responses concerning consumer behavior (*2*). Although opinions on the feasibility of this application vary, at a minimum, studies that use simulated participants could be used to generate hypotheses that could then be confirmed in human populations (*3*, *4*). The success of this approach depends on algorithmic fidelity of the trained data (*3*), transparency in model training, prompt engineering, and benchmark selection.

Why is this scenario plausible? Pretrained on massive datasets, advanced AI models can represent a vast array of human experiences and perspectives, possibly giving them a higher degree of freedom to generate diverse responses than that of conventional human participant methods, which can help to reduce generalizability concerns in research (*2*). LLMs can also generate responses across a wider range of parameters than human participants because of pragmatic concerns of limited attention span, response bias, or habituation among humans, providing a less biased view of underlying latent dimensions. This makes them especially useful in high-risk projects for which traditional data collection is impractical, allowing for the testing of interventions in simulated populations before real-world implementation.

LLMs could be used as surrogates in other ways. They have the potential to enhance policy analysis by reproducing the views of different theoretical or ideological schools of thought. For example, LLMs could be trained to capture nuances of complex debates, such as concerning the stability and reliability of nuclear deterrence in the face of human and technical factors (*5*). LLMs could be trained to capture varied perspectives, including evaluating "what-if" scenarios that nearly occurred, such as the Cuban Missile Crisis in 1962, and providing assessments of how plausible these scenarios were. Once LLMs can pass the Ideological Turing Test—meaning that they can accurately represent opposing viewpoints in a way indistinguishable from real humans—researchers can use them to generate future scenarios. Future LLMs, appropriately trained (*3*), may thus out-perform humans on analytic tasks such as synthesizing clashing views to generate superior forecasts and policy prescriptions.

AI could also fill the role of a "confederate" (controlled experimental partner) in social interaction research involving individuals or groups (*6*), potentially as components to agent-based simulations. An LLM-ABM hybrid could use LLM to derive empirically based rules of social decision-making or behavior to simulate social interactions of individuals with specific characteristics and beliefs (*4*). This approach could explore how agents with these particular characteristics influence subsequent interaction with humans, informing broader social science questions such as how misinformation spreads throughout social networks (*7*).

Such investigations raise questions about the limits of LLMs as human cognition and decision models. Can we "nudge" an LLM by asking it to assess the quality of a news item before sharing, replicating research with humans (*7*)? If so, could we use the integrated LLM-ABM model to identify interventions that would reduce the spread of misinformation through social networks? Generally, if LLM-ABMs can provide new insights on how human agents choose to share information, cooperate and compete in social dilemmas, and conform with social norms, they can provide valuable insights into both the underlying mechanisms governing human behavior and social dynamics (*8*) with higher fidelity than has been possible with previous human decision models.

Incorporating LLMs into ABMs introduces new challenges because of their differing operational principles. Whereas LLMs generate and interpret language according to statistical patterns derived from vast linguistic data, traditional ABMs operate on the basis of predefined formal rules (*9*) that can be generated by using real-world linguistic and other qualitative data. New ABM design will be needed to take advantage of LLMs' capability to simulate performance on questionnaires, behavior in ill-defined situations, or open-ended responses (*2*). By creating realistic initial populations for ABMs, LLMs can model subjects' latent cognitive or affective states, surpassing traditional researchers' capacity and opening doors for future theory generation.

LLMs' potential future benefits include creating samples as diverse as the cultural products (*2*, *3*) on which the models were trained, offering a more accurate

[1]Department of Psychology, University of Waterloo, Waterloo, ON, Canada. [2]Waterloo Institute for Complexity and Innovation, University of Waterloo, Waterloo, ON, Canada. [3]Rotman School of Management, University of Toronto, Toronto, ON, Canada. [4]School of Planning, University of Waterloo, Waterloo, ON, Canada. [5]Yale Institute for Network Science, Yale University, New Haven, CT, USA. [6]Wharton School of Business, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Psychology, University of Toronto, Toronto, ON, Canada. [8]Vector Institute, Toronto, ON, Canada. [9]Schwartz Reisman Institute for Technology and Society, University of Toronto, Toronto, ON, Canada. Email: igrossma@uwaterloo.ca

portrayal of human behavior and social dynamics than those from conventional methods that rely on typically less heterogenous and representative convenience samples (*2*). Because of their population-scale calibration data, LLMs could help address common challenges in social science research that can lead to biased models, including generalizability and self-selection concerns (*2*).

Effective AI-assisted research will depend on the AI being able to accurately mirror the perspectives of diverse demographic groups. Pretrained models from linguistic cultural products are known to capture sociocultural biases present in society (*2*, *10*). When biases are recognized, a key question is their provenance: Do they correctly reflect the populations, or are they artifacts of model construction (*11*)? Model construction bias may result from incorrect or invalid choices throughout the design and development pipeline (for example, choosing constructs that are differentially valid across demographic groups, curating datasets that lack diversity or that encode biases of certain human annotators, or selecting models that fail to capture specific patterns pertinent to minorities) or because of existing societal disparities (*2*).

The scientist-humanist dilemma emerges as a key issue: Although scientists aim to study "pure" LLMs with embedded sociocultural biases to simulate human behavior and trace its cultural evolution (*2*), ethical constraints require engineers to protect LLMs from these very biases. Already, LLM engineers have been fine-tuning pretrained models for the world that "should be" (*12*) rather than the world that is, and such efforts to mitigate biases in AI training (*2*, *13*) may thus undermine the validity of AI-assisted social science research. The proprietary "black box" nature of LLM training challenges the ability of researchers to evaluate underlying mechanisms and replicate findings. To address this, advocating for open-source LLMs, access to pretrained but not fine-tuned models for scientific research, and transparent methodologies (such as BLOOM, Cerebras-GPT, or LLaMA) are essential for ensuring reliable and credible AI-driven research (*2*).

Overall, researchers will need to establish guidelines for the ethical use of LLMs in research, addressing concerns related to data privacy, algorithmic fairness [versus monoculture (*2*)], environmental costs (*2*, *13*), and the potential misuse of LLM-generated findings. Pragmatic concerns with data quality, fairness, and equity of access to the powerful AI systems will be substantial.

In deciding whether to use LLMs to approximate human behavior, researchers must first validate language-mediated (latent) constructs (*2*). They can treat LLM-generated responses as a "sample" of nonhuman participants and systematically vary prompts, akin to presenting random stimuli in traditional experiments. A crucial consideration in using LLMs for research is the trade-off between external and internal validity. Future LLMs, trained on diverse cultural content, will offer greater external validity by simulating human-like responses and generalizing to real-world scenarios. However, their opaque nature will limit their internal validity. Conversely, laboratory-grown natural-language processing models, built on smaller controlled datasets, will provide stronger internal validity at the expense of reduced reliability and generalizability because the limited training data may hinder their ability to perform consistently and broadly across different contexts. Researchers should carefully choose between these approaches according to their priorities.

"...large language models rely on 'shadows' of human experiences described in cultural products."

Researchers must also consider the context of their study. High-risk situations that involve violence or situations that are plainly infeasible with large numbers of human participants may be more suitable for LLMs. For example, LLMs might be used to explore human dynamics of space travel, or create predator and victim prototypes for studies of online sexual predators, an ethically fraught realm because of the potential trauma to human participants.

As AI reshapes the landscape of social science (*14*), researchers will diversify their expertise, embracing new roles such as model bias hunters, AI-data validators, or human-AI interactionist. In this context, maintaining conceptual clarity (*2*), understanding foundations of measurement (*2*), and adhering to ethically grounded practical wisdom (*15*) for selecting an AI-assisted design that fits one's research question will be essential. With the democratization of AI-assisted data collection, the importance of early-stage social science training and supporting quantitative methods (such as computation and statistics) is crucial, calling for revision of social science education programs.

Just as the prisoners in the allegory of Plato's Cave observe shadows on a wall and believe them to represent reality, LLMs rely on "shadows" of human experiences described in cultural products. These shadows offer a limited view of the true nature of the phenomena they represent because folk psychology (*2*) captured in cultural products may not always reflect the mechanisms that govern human behavior—a limitation essential for social scientists to acknowledge. Examining the limitations and biases of LLMs also puts a mirror to common practices in many fields, be it bias in representation, sampling methods, or methodological individualism (*2*).

Despite these obstacles, LLMs allow social scientists to break from traditional research methods and approach their work in innovative ways. LLM models will likely vitiate online crowdworking platforms, which are the dominant source of human participant data in many social science fields, for the simple reasons of on-par performance of simple tasks, and because open-ended responses from LLM-guided bots will become indistinguishable from human participants, calling for new methods for human data verification. Social scientists must be prepared to adapt to the uncertainty (*15*) that comes with evolving technology while being mindful of the limitations of ongoing research practices. Only by maintaining transparency and replicability (*2*) can we ensure that AI-assisted social science research truly contributes to our understanding of human experience. ■

REFERENCES AND NOTES

1. S. Bubeck *et al*., *arXiv*:2303.12712 [cs.CL] (2023).
2. Extended documentation of LLMs abilities, ethical challenges, and methodological concerns, along with foundational social science principles, is available on Open Science Framework (https://osf.io/h4e2a).
3. L. P. Argyle *et al*., *Polit. Anal*. 10.1017/pan.2023.2 (2023).
4. J. S. Park *et al*., *arXiv*:2304.03442 [cs.HC] (2023).
5. P. E. Tetlock, C. B. McGuire, G. Mitchell, *Annu. Rev. Psychol*. **42**, 239 (1991).
6. H. Shirado, N. A. Christakis, *Nature* **545**, 370 (2017).
7. G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, D. G. Rand, *Psychol. Sci*. **31**, 770 (2020).
8. M. Galesic, H. Olsson, J. Dalege, T. van der Does, D. L. Stein, *J. R. Soc. Interface* **18**, 20200857 (2021).
9. P. Antosz, S. Bharwani, M. Borit, B. Edmonds, *Int. J. Soc. Res. Methodol*. **25**, 511 (2022).
10. A. Abid, M. Farooqi, J. Zou, *Nat. Mach. Intell*. **3**, 461 (2021).
11. S. Fazelpour, D. Danks, *Philos. Comp.* **16**, e12760 (2021).
12. Y. Bai *et al*., *arXiv*:2212.08073 [cs.CL] (2022).
13. L. Weidinger *et al*., in *2022 ACM Conference on Fairness, Accountability, and Transparency* (ACM, 2022), pp. 214–229.
14. J. C. Peterson, D. D. Bourgin, M. Agrawal, D. Reichman, T. L. Griffiths, *Science* **372**, 1209 (2021).
15. I. Grossmann *et al*., *Psychol. Inq*. **31**, 103 (2020).

ACKNOWLEDGMENTS

The authors thank T. Charlesworth, R. Saxe, and S. Fazelpour for their feedback on earlier versions of the draft. This work was funded by Social Sciences and Humanities Research Council of Canada Connection grant 611-2020-0190 (to I.G.), Social Sciences and Humanities Research Council of Canada Insight grant 435-2014-0685 (to I.G.), and John Templeton Foundation grant no. 62260 (to I.G. and P.E.T.).

10.1126/science.adi1778

ARTIFICIAL INTELLIGENCE

Art and the science of generative AI

Understanding shifts in creative work will help guide AI's impact on the media ecosystem

By **Ziv Epstein[1], Aaron Hertzmann[2,3], *and* the Investigators of Human Creativity***

The capabilities of a new class of tools, colloquially known as generative artificial intelligence (AI), is a topic of much debate. One prominent application thus far is the production of high-quality artistic media for visual arts, concept art, music, and literature, as well as video and animation. For example, diffusion models can synthesize high-quality images (*1*), and large language models (LLMs) can produce sensible-sounding and impressive prose and verse in a wide range of contexts (*2*). The generative capabilities of these tools are likely to fundamentally alter the creative processes by which creators formulate ideas and put them into production. As creativity is reimagined, so too may be many sectors of society. Understanding the impact of generative AI—and making policy decisions around it—requires new interdisciplinary scientific inquiry into culture, economics, law, algorithms, and the interaction of technology and creativity.

Generative AI tools, at first glance, seem to fully automate artistic production—an impression that mirrors past instances when traditionalists viewed new technologies as threatening "art itself." In fact, these moments of technological change did not indicate the "end of art," but had much more complex effects, recasting the roles and practices of creators and shifting the aesthetics of contemporary media (*3*). For example, some 19th-century artists saw the advent of photography as a threat to painting. Instead of replacing painting, however, photography eventually liberated it from realism, giving rise to Impressionism and the Modern Art movement. By contrast, portrait photography did largely replace portrait painting. Similarly, the digitization of music production (e.g., digital sampling and sound synthesis) was decried as "the end of music." Instead, it altered the ways people produce and listen to music, and helped spawn new genres, including hip hop and drum'n'bass. Like these historical analogs, generative AI is not the harbinger of art's demise, but rather is a new medium with its own distinct affordances. As a suite of tools used by human creators, generative AI is positioned to upend many sectors of the creative industry and beyond—threatening existing jobs and labor models in the short term, while ultimately enabling new models of creative labor and reconfiguring the media ecosystem.

Unlike past disruptions, however, generative AI relies on training data made by people. The models "learn" to generate art by extracting statistical patterns from existing artistic media. This reliance on training data raises new issues—such as where the data is sourced, how it influences the outputs, and how to determine authorship. By leveraging existing work to automate aspects of the creative process, generative AI challenges conventional definitions of authorship, ownership, creative inspiration, sampling, and remixing and thus complicates existing conceptions of media production. It is therefore important to consider generative AI's impacts on aesthetics and culture, legal questions of ownership and credit, the future of the creative work, and impacts on the contemporary media ecosystem. Across these themes, there are key research questions to inform policy and beneficial uses of this technology (*4*).

> **"...generative AI is not the harbinger of art's demise, but rather is a new medium with its own distinct affordances."**

To properly study these themes, it is first necessary to understand how the language used to describe AI affects perceptions of the technology. The very term "artificial intelligence" might misleadingly imply that these systems exhibit human-like intent, agency, or even self-awareness. Natural language–based interfaces now accompany generative AI models, including chat interfaces that use the "I" pronoun, which may give users a sense of human-like interaction and agency. These perceptions can undermine credit to the creators whose labor underlies the system's outputs (*5*) and deflect responsibility from developers and decision-makers when these systems cause harm (*6*). Future work is needed to understand how perceptions of the generative process affect attitudes toward outputs and authors. This could facilitate the design of systems that disclose the generative process and avoid misleading interpretations.

Generative AI's specific affordances in turn give rise to new aesthetics that may have a long-term effect on art and culture. As these tools become more widespread, and their use becomes commonplace (as with photography a century ago), it remains an open question how the aesthetics of their outputs will affect artistic outputs. A low barrier to entry for generative AI could increase the overall diversity of artistic outputs by expanding the set of creators who engage with artistic practice. At the same time, aesthetic and cultural norms and biases embedded in the training data might be captured, reflected, and even amplified—thereby decreasing diversity (*7*). AI-generated content may also feed future models, creating a self-referential aesthetic flywheel that could perpetuate AI-driven cultural norms. Future research should explore ways to quantify and increase output diversity and study how generative AI tools may influence aesthetics and aesthetic diversity.

The opaque, engagement-maximizing recommender algorithms of social media platforms could further reinforce aesthetic norms through feedback loops (*8*) that produce sensational and shareable content. As algorithm and content creators try to maximize engagement, this may further homogenize content. However, some preliminary experiments (*9*) suggest that incorporating engagement metrics when curating AI-generated content can, in some cases, diversify content. It remains an open question what styles are amplified by recommender algorithms, and how that prioritization affects the types of content creators make and share. Future work must explore the complex, dynamic systems formed by the interplay between generative models, recommender algorithms, and social media platforms, and their resulting impact on aesthetics and conceptual diversity.

Generative AI's reliance on training data to automate aspects of creation raises legal and ethical challenges regarding authorship and thus should prompt technical research into the nature of these systems. Copyright law must balance the benefits to creators, users of generative AI tools, and society at large. Laws could treat the use of training

[1]Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Adobe Research, San Francisco, CA, USA. [3]University of Washington, Seattle, WA, USA. Corresponding author: zive@mit.edu *The Investigators of Human Creativity coauthors and affiliations are listed at the end of this paper.

data as noninfringing if protected works are not directly copied; fair use if training involves a substantial transformation of the underlying data; allowed only if creators give an explicit license; or subject to statutory compulsory licensing that allows data to be used for training provided that creators are compensated. Much of copyright law relies on judicial interpretations, so it is not yet clear if collecting third-party data for training or mimicking an artist's style would violate copyright. Legal and technical issues are entwined: Do models directly copy elements from the training data, or produce entirely new works? Even when models do not directly copy from existing works, it is not clear whether and how artists' individual styles should be protected. What mechanisms could protect and compensate artists whose work is used for training, or even permit them to opt out, while still allowing new cultural contributions to be made with generative AI models? Answering these questions and determining how copyright law should treat training data require substantial technical research to develop and understand the AI systems, social science research to understand perceptions of similarity, and legal research to apply existing precedents to new technology. Of course, these points represent only an American legal perspective.

A distinct legal question concerns who can claim ownership over model outputs. Answering this requires understanding the creative contributions of a system's users versus other stakeholders, such as the system's developers and creators of the training data. AI developers could claim ownership over outputs through terms of use. By contrast, if users of the system have engaged in a meaningfully creative way (e.g., the process is not fully automated, or does not emulate specific works), then they might be considered as the default copyright holders. But how substantial must users' creative influence be for them to claim ownership? These questions involve studying the creative process of using AI-based tools and may become more complex if users gain more direct control.

Regardless of legal outcomes, generative AI tools are likely to transform creative work and employment. Prevailing economic theory [i.e., skill-biased technological change (SBTC)] assumes that cognitive and creative workers face less labor disruption from automation because creativity is not readily encodable into concrete rules (i.e., Polanyi's paradox) (*10*). Yet, new tools have sparked employment concerns for creative occupations such as composers, graphic designers, and writers. This conflict arises because SBTC fails to differentiate between cognitive activities such as analytical work and creative ideation. A new framework is needed to characterize the specific steps of the creative process, which of those steps might be affected by generative AI tools, and the effects on workplace requirements and activities of varying cognitive occupations (*11*).

Although these tools may threaten some occupations, they could increase the productivity of others and perhaps create new ones. For example, historically, music automation technologies enabled more musicians to create, even as earnings skewed (*12*). Generative AI systems can create hundreds of outputs per minute, which may accelerate the creative process through rapid ideation. However, this acceleration might also undermine aspects of creativity by removing the initial period of prototyping associated with a tabula rasa. In either case, production time and costs will likely fall. The production of creative goods may become more efficient, leading to the same amount of output with fewer workers. In turn, demand for creative work may increase. However, the production of creative goods may become more efficient, leading to the same amount of output with fewer workers. Furthermore, many work-for-hire occupations using conventional tools, like illustration or stock photography, could be displaced. Several historical examples bear this out. Most notably, the Industrial Revolution enabled mass production of traditionally artisanal crafts (e.g., ceramics, textiles, and steelmaking) with the labor of non-artisans; hand-made goods became specialty items. Similarly, photography replaced portrait painting. Digitization of music removed constraints of learning to physically manipulate instruments and enabled more complex arrangements with more contributors. These tools may change who can work as an artist, in which case artists' employment may rise even as average wages fall.

As these tools affect creative labor, they also introduce potential downstream harms to the broader media ecosystem. As the cost and time to produce media at scale decreases, the media ecosystem may become vulnerable to AI-generated misinformation through the creation of synthetic media, particularly media that provides probative evidence for claims (*13*). These new possibilities for generating photorealistic synthetic media may undermine trust in authentically captured media through the so-called "liar's dividend" (fake content benefits liars by undermining trust in the truth) (*14*) and also increase threats of fraud and nonconsensual sexual imagery. This raises important research questions: What is the role of platform interventions such as tracking source provenance and detecting synthetic media downstream in governance and building trust (*15*)? And how does the proliferation of synthetic media affect trust in real media, such as unedited journalistic photographs? As content production increases, collective attention spans may decrease (*16*). The explosion of AI-generated content may in turn hamper society's ability to collectively discuss and act in important arenas such as climate and democracy.

Every artistic medium mirrors and comments on the issues of its time, and the debates surrounding contemporary AI-generated art reflect present issues surrounding automation, corporate control, and the attention economy. Ultimately, we express our humanity through art, so understanding and shaping the impact of AI on creative expression is at the center of broader questions about its impact on society. New research into generative AI should inform policy and beneficial uses of the technology while engaging with critical stakeholders, particularly artists and creative laborers themselves, many of whom actively engage with difficult questions at the vanguard of societal change. ■

REFERENCES AND NOTES

1. R. Rombach *et al*., *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022). pp. 10684–10695.
2. A. Vaswani *et al*., "Attention is all you need" in *Advances in Neural Information Processing Systems* 30 (2017).
3. A. Hertzmann, *Arts*. **7** (no. 2) (2018).
4. Z. Epstein *et al*., Art and the science of generative AI: A deeper dive. arXiv2306.04141 [cs.CY] (2023).
5. Z. Epstein *et al*., *iScience* **23**, 101515 (2020).
6. M. C. Elish, *Engag. Sci. Technol. Soc*. **5**, 40 (2019).
7. S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York Univ. Press, 2018).
8. A. J. B. Chaney *et al*., in *Proceedings of the 12th ACM Conference on Recommender Systems* (2018), pp. 224–232.
9. Z. Epstein, M. Groh, A. Dubey, A. Pentland, *Proc. ACM Hum. Comput. Interact*. **5** (CSCW2), 1 (2021).
10. C. B. Frey, M. A. Osborne, *Technol. Forecast. Soc. Change* **114**, 254 (2017).
11. M. R. Frank *et al*., *Proc. Natl. Acad. Sci. U.S.A*. **116**, 6531 (2019).
12. D. Hesmondhalgh *et al*., "Music creators' earnings in the digital era." Intellectual Property Office Research Paper (2021).
13. J. A. Goldstein *et al*., arXiv 2301.04246 [cs.CY] (2023).
14. B. Chesney, D. Citron, *Calif. Law Rev*. **107**, 1753 (2019).
15. H. Farid, *J. Online Trust Safety* 10.54501/jots.v1i4.56 (2022).
16. P. Lorenz-Spreen *et al*., *Nat. Commun*. **10**, 1759 (2019).

ACKNOWLEDGMENTS

Z.E. and M.G. helped assess DALL-E 2 and M.R.F. helped assess ChatGPT for OpenAI. A.H. and L.H. work for Adobe, which makes generative AI tools; the ideas expressed here do not represent those of the company.

Investigators of Human Creativity Memo Akten[1], Hany Farid[2], Jessica Fjeld[3], Morgan R. Frank[4], Matthew Groh[5], Laura Herman[6,7], Neil Leach[8], Robert Mahari[3,5], Alex "Sandy" Pentland[5], Olga Russakovsky[9], Hope Schroeder[5], Amy Smith[10]

[1]University of California, San Diego, San Diego, CA, USA. [2]University of California, Berkeley, Berkeley, CA, USA. [3]Harvard Law School, Cambridge, MA, USA. [4]University of Pittsburgh, Pittsburgh, PA, USA. [5]Massachusetts Institute of Technology, Cambridge, MA, USA. [6]University of Oxford, Oxford, UK. [7]Adobe, Inc., London, UK. [8]Florida International University, Miami, FL, USA. [9]Princeton University, Princeton, NJ, USA. [10]Queen Mary University of London, London, UK.

10.1126/science.adh4451

BOOKS *et al.*

Perception reflects an interplay between our brai
models of the environment and our internal states.

NEUROSCIENCE

The body's predictive processor

Brains begin with complete models of the world and work backward to anticipate sensory data, argues a philosopher

By **Zoe Drayson**

How do our brains process, filter, and encode all the sensory data with which we are continually bombarded so that we experience a world of objects, people, and opportunities? Perhaps they do not. Proponents of "predictive processing" suggest that the bulk of our brains' processing operates in the opposite direction, starting with fully formed models and working backward to predict the sensory inputs that would be expected from such a world. If the brain's predictions do not match incoming sensory input, then error signals are fed back to correct the models. The more accurate the brain's predictions, the less processing of sensory inputs required.

In *The Experience Machine*, philosopher Andy Clark masterfully synthesizes recent work on the predictive brain into an accessible and captivating book for the nonspecialist. He explains not only how predictive processing in the brain can account for accurate perception but also—and perhaps more fascinatingly—what happens when things go wrong. If sensory feedback is ambiguous or limited, for example, or if a brain assigns too much weight to an existing model of the world, then errors can occur. Clark illustrates these scenarios with familiar and compelling examples, describing why we sense phantom vibrations from our cell phones, why we hear speech in white noise, and why we experience a blue-and-black dress as white and gold.

The Experience Machine: How Our Minds Predict and Shape Reality
Andy Clark
Pantheon, 2023. 304 pp.

In Clark's hands, the predictive brain is not merely a framework for understanding how our perceptual systems might work. Instead, it is a unifying theory of human minds. He weaves a compelling narrative that explores how perception, action, pain, emotion, and consciousness can all be understood as reflections of the relationship between our brain's model of the external world and its model of our internal physiological states. And he explores how the predictive brain can explain various phenomena, including posttraumatic stress disorder, sporting expertise, the aesthetic response to artistic beauty, and even Patrick Swayze's character's dismissive attitude toward pain—"Pain don't hurt"—in the 1989 movie *Road House*.

In one of the book's most memorable examples, Clark explores what happens in the central nervous system when police officers misidentify a nonharmful object or movement as a threat. The predictive brain, he proposes, takes the police officer's own heightened internal bodily sensations (such as increased heart rate) as evidence of a threatening situation, which predisposes the visual system to detect a weapon where there is none.

Clark's focus on neural processing may surprise readers familiar with the author's earlier work on embodied and extended cognition. He has spent several decades arguing that many of our cognitive abilities are not brain-bound but rather are inextricable from our bodies and environments. However, Clark is adamant that we should not misunderstand this new book's framework as a return to neurocentrism. He proposes instead that the predictive brain's ability to combine inward-looking and outward-looking sensory information is precisely what enables it to exploit non-neural resources to extend its cognitive reach. This notion is the book's most intriguing and its most metaphorical. Clark proposes that human minds are "seething, swirling oceans of prediction, continuously orchestrated by brain, body, and world."

What are the practical implications of understanding the brain as a predictive machine? Throughout the book, Clark explores the possibilities that specialized treatments and medical interventions could be used to shape our brain's predictive processing. Biofeedback, psychedelic drugs, and immersive virtual reality are some of the options considered. He also raises the question of how we might "hack" our own brains using the power of conscious thought, in a way that sometimes sounds ominously close to the pseudoscientific interpretations, such as manifestation, that Clark cautions against. "We must at some level strongly predict that we will occupy the states that we can plausibly attain and that best realize our goals," he writes. "Realistic optimism is thus the order of the day."

Scientific work and philosophical work on predictive processing have been careful to distinguish neural prediction and precision weighting from what we do as conscious human beings when we estimate outcomes and consider probabilities. In this book, however, Clark is keen to unite our understanding of conscious and unconscious cognitive strategies. He suggests that our best understanding of affirmation and visualization techniques, for example, is in terms of the predictive brain's error minimization strategy. Those hoping for an explanation of how our conscious expectations interact with nonconscious neural prediction mechanisms, however, will be left dissatisfied. This part of the puzzle, he acknowledges, is still missing. Nonetheless, *The Experience Machine* provides plenty of fodder to inspire the next generation of researchers to take on this task. ■

The reviewer is at the Department of Philosophy, University of California, Davis, CA 95616, USA. Email: zdrayson@ucdavis.edu

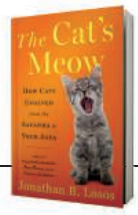
10.1126/science.adh9416

ANIMAL BIOLOGY

Nine lives

A biologist explores the cat's origins, evolution, and complicated history with humans

The Cat's Meow: How Cats Evolved From the Savanna to Your Sofa
Jonathan B. Losos
Viking, 2023. 400 pp.



By **Greger Larson**

I recently received campaign flyers from two local candidates for office. Both featured smiling, balding, middle-aged men wearing sweaters over button-down shirts. Both men cradled cats. In an effort to convince residents of their suitability for public office, these men did not choose to be photographed with children or dogs or spouses. They chose cats.

Of all the animals with whom we have formed domestic relationships, how is it that the cat has become so globally ubiquitous and popular? Evolutionary biologist Jonathan Losos's book *The Cat's Meow* has answers. Across 20 chapters, Losos explores every aspect of cats with charm and breathless enthusiasm, from their origins, history, and behavior to humans' penchant for molding new breeds using rare morphological traits.

Instead of expressing exasperation or frustration at how little is known about cats, Losos sees opportunity and in more than a dozen specific instances suggests how filling a particular knowledge gap would be an excellent research project. Do adult cats of other species "knead" as house cats do? What are the behavioral and anatomical differences between North and South African wildcats? To what degree has human preference been a factor in promoting the appearance and spread of noncamouflage coat colors? How much longer do indoor cats live relative to those who spend most of their time outside? His many questions have the makings of at least five PhD dissertations and two master's projects.

While there is much still to learn, research over the past two decades has established a few facts. Although house cats can and do produce fertile offspring with a range of similarly sized wildcat species distributed across Eurasia and Africa, our domestic versions are derived from just one primary ancestor: the North African wildcat. The first cat found in an archaeological context was discovered alongside a woman in a burial on the island of Cyprus dated to 9500 years ago. Whether this cat would be recognized as a domestic cat from a modern perspective is uncertain, but this site demonstrates that the affinity between people and cats dates back millennia.

Ancient Egyptians had a particular fondness for cats. Losos details how the creatures were absent from Egyptian iconography until ~4000 years ago, when they became regular features on temple walls. By 3000 years ago, hundreds of thousands of people made annual pilgrimages to pay homage to the Egyptian goddess Bastet, and they did so by purchasing and placing a mummified cat in a temple. To meet the demand, huge catteries bred millions of cats that were sacrificed as juveniles on an industrial scale and mummified.

A pair of Egyptian Maus sit on either side of a statue of the ancient Egyptian goddess Bastet.

Losos packs in scores of other cat stories and observations that never cease to fascinate. For the most part, he does so with careful yet accessible science that avoids overly simplistic conclusions. On two topics, however, he accepts conventional wisdom when a more incisive approach would have been welcome.

Despite frequently discussing the substantial variability of individual cats gleaned from both scientific studies and anecdotal stories, Losos never questions the assumption that behavior can be predicted by a cat's breed. A recent study of more than 15,000 dogs, however, showed that breed designation does not predict individual behavior, and cats are likely no different (*1*).

Losos admirably presents a carefully worded deconstruction of the biological species concept and its lack of relevance when applied to domestic animals. A similar discussion of the terms "fully domesticated" or "semidomesticated"—both of which are arbitrary and effectively meaningless—would also have been welcome. Never mind that, unlike other domesticates, cats often sleep on people's heads, Losos concludes that because domestic cats have not been as "greatly transformed" as other animals, they are rightly placed in the "semidomesticated" camp. But these are minor quibbles.

Cats are amazing. They fascinate and bewilder us. They enhance our lives, and they murder wildlife. Since forming a tight-knit relationship with us, they have altered their cries and purrs to more effectively manipulate us. Feral cats in Australia can sense the locations of wildfires and travel miles to hunt on the now open landscape. There are hundreds of emerging cat stories detailed here that will only heighten the reader's amazement. And Losos has done cats at least as proud as we imagine they pride themselves. ■

The reviewer is at the Palaeogenomics and Bio-Archaeology Research Network, School of Archaeology, University of Oxford, Oxford OX1 3TG, UK. Email: greger.larson@arch.ox.ac.uk

REFERENCES AND NOTES

1. K. Morrill *et al.*, *Science* **376**, eabk0639 (2022).

10.1126/science.adi1153

PHOTO: YVES LANCEAU/NPL/MINDEN PICTURES

An abandoned fishing ship sits on the drained floor of the desiccated Aral Sea.

Edited by **Jennifer Sills**

Address the Aral Sea crisis with cooperation

At the United Nations water conference held in New York in March, Central Asian countries issued a joint statement that highlighted the ecological disaster caused by the desiccation of the Aral Sea. Once the world's fourth largest lake (*1*, *2*), the shrinking Aral Sea has transformed the lake's basin into a 5.5-million-hectare desert, posing challenges to Uzbekistan, Kazakhstan, Turkmenistan, Tajikistan, and Kyrgyzstan. Addressing this crisis will require international support and regional cooperation.

The primary catalyst behind the Aral Sea's desiccation was the overextension of irrigation within the sea's catchment area between the mid-1950s and mid-1980s. The excessive demand disrupted the region's ecological balance and resulted in a substantial decrease of water flowing into the lake from rivers (*3*). The loss of the Aral Sea's fresh water has led to the degradation of the lake's aquatic ecosystems as well as disruptions to major terrestrial ecosystems (*4*–*6*). Local mesophytic plant communities, for example, have decreased in both distribution and diversity (*7*). Meanwhile, the release of toxic salty dust particles from the exposed seabed has caused a decline in human health and welfare, including increased frequency of typhus, gastritis, epilepsy, leukemia, bronchitis, and asthma in local residents (*8*).

Donors and international organizations, such as the United Nations Development Programme, can provide resources and expertise to assist the affected countries in implementing sustainable solutions (*9*). International support can help countries in the Aral Sea basin implement more efficient irrigation systems and adopt modern farming practices. These technological advancements would save water, which could then be reused for agricultural crops. By mobilizing financial resources, the international community can provide funding for sustainable infrastructure and development in the region.

Continued cooperation among the five Central Asian countries that share the Aral Sea basin is also critical (*10*). In 2018, the Presidents of Central Asian countries convened in Turkmenistan and established the Comprehensive Program of Measures to Mitigate the Consequences of the Aral Disaster and the Development of the Aral Sea Region, which addressed water conservation, improved ecological protection, and coordinated efforts among Central Asian nations (*11*). In the past 5 years, more than 500 initiatives have been executed.

However, despite the efforts to address the environmental aspects of desiccation, Central Asian nations continue to grapple with critical water scarcity and ongoing conflicts and disputes over water resources. These countries should increase collaboration to equitably distribute water resources and invest in programs that restore and preserve the ecological balance of the Aral Sea, such as the International Fund for Saving the Aral Sea (*12*). With international support and regional cooperation, the Aral Sea countries can integrate the use of water resources, protect the environment, and conserve biological assets.

Lei Wang[1]*, Zhenyong Zhao[1], Khabibullo Shomurodov[2], Changyan Tian[1]
[1]State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China. [2]Institute of Botany, Academy of Sciences of the Republic of Uzbekistan, Tashkent 100053, Uzbekistan.
*Corresponding author.
Email: egiwang@ms.xjb.ac.cn

REFERENCES AND NOTES

1. P. Micklin, in *Dying and Dead Seas Climatic versus Anthropic Causes* (Kluwer Academic Publishers, 2004), pp. 99–123.
2. P. Micklin, *Lake Reserv. Res. Manag.* **15**, 193 (2010).
3. N. V. Aladin, *Arid Ecosyst.* **16**, 76 (2009).
4. R. Stone, *Science* **284**, 30 (1999).
5. M. Loodin, *Model. Earth Syst. Environ.* **6**, 2495 (2020).
6. P. Micklin, *Annu. Rev. Earth Planet. Sci.* **35**, 47 (2007).
7. B. Adilov *et al.*, *J. Arid Land* **13**, 71 (2021).
8. S. Vasilev, "Aral Sea: Causes of drying up and its consequences" (2021). https://24smi.org/facts/287526-aralskoe-more-prichiny-vysykhaniia-i-ego-posledstv.html [in Russian].
9. M. J. Deng, A. H. Long, *J. Glaciol. Geocryol.* **33**, 1376 (2011) [in Chinese].
10. N. V. Aladin, J. T. Høeg, I. Plotnikov, *Science* **370**, 1283 (2020).
11. "Heads of Central Asian states met in Turkmenbashi," Gazeta.uz (2018); https://www.gazeta.uz/ru/2018/08/24/meeting/ [in Russian].
12. International Fund for Saving the Aral Sea (2017); https://kazaral.org/en/.

10.1126/science.adi2199

Protect fragile mountaintop ecosystems

Protecting fragile mountaintop ecosystems is a crucial component of the Sustainable Development Goals of the United Nations' 2030 Agenda for Sustainable Development (*1*). Mountaintops are among the ecosystems most vulnerable to climate change

(*2*–*4*). Global warming could lead to the extinction of about 10% of eukaryotic species on Earth by 2100 (*5*). Mountaintop ecosystems preserve vast amounts of biodiversity that lives in extreme environmental conditions, provide food and water for human well-being, and are important for global carbon storage (*6*). Thus, they have irreplaceable economic and ecological values in the biodiversity, stability, and sustainability of global and local ecosystems. The protection of mountaintop ecosystems needs a targeted conservation framework.

Mountaintops are exposed to various hostile environments, including high levels of UV radiation, dramatic fluctuations in moisture and temperature, low levels of oxygen partial pressure, low water-holding capacity, and limited soil fertility (*4*, *6*). They are also highly vulnerable to disturbances, such as land-use change, unsustainable resource use, alien species invasion, and hydrological events (*7*, *8*). Although montane species shift their distributions upslope to track their optimal climate and even increase the species richness in the mountaintops under global warming (*9*), this upslope migration could lead to increased interspecific competition and great extinction debt of species already at the summit.

To date, global conservation plans and actions do not pay enough attention to mountaintops despite the ongoing extinction in this ecosystem (*5*, *8*, *10*). This lack of focus has allowed mountaintop mining to cause 40% loss of aquatic biodiversity in West Virginia (*11*). In addition, bushfires have led to the extinction risk of mountaintop plants in Western Australia's Stirling Range (*12*).

Mountaintop ecosystems will receive more attention if international organizations (such as International Union for Conservation of Nature) add them to the global agenda of biodiversity conservation, monitoring, and assessment. Governments can include more mountaintops in national protected areas and safeguard their biodiversity through law enforcement. Finally, more data about the status of mountaintop species and ecosystems is urgently needed. Scientists should conduct research and ensure that accurate information is incorporated into conservation targets and plans.

Shi-Kang Shen*, Xiong-Li Zhou, Si-Qi Wang, Zhenyu Lyu, Rui Zhang, Liu Yang, Bo Long
Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology, Institute of Biodiversity, School of Ecology and Environmental Science, Yunnan University, Kunming, Yunnan 650504, China.
*Corresponding author. Email: ssk168@ynu.edu.cn

REFERENCES AND NOTES

1. "Sustainable Development Goals: 17 Goals to Transform our World," United Nations (2015); https://www.un.org/en/exhibits/page/sdgs-17-goals-transform-world.
2. C. M. Costion *et al.*, *Biol. Conserv.* **191**, 322 (2015).
3. L. Monteiro *et al.*, *Flora* **238**, 234 (2018).
4. M. C. Urban, *Science* **348**, 571 (2015).
5. B. G. Freeman, M. N. Scholer, V. Ruiz-Gutierrez, J. W. Fitzpatrick, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11982 (2018).
6. F. Hagedorn, K. Gavazov, J. M. Alexander, *Science* **365**, 1119 (2019).
7. M. A. Palmer *et al.*, *Science* **327**, 148 (2010).
8. L. Baumbach, D. L. Warren, R. Yousefpour, M. Hanewinkel, *Commun. Biol.* **4**, 869 (2021).
9. M. J. Steinbauer *et al.*, *Nature* **556**, 231 (2018).
10. D. Hoffmann, M. F. de Vasconcelos, G. W. Fernandes, *Perspect. Ecol. Conserv.* **18**, 257 (2020).
11. M. Simonin et al., *Ecol. Appl.* **31**, e02389 (2021).
12. Y. Rack, *BioScience* 10.1093/biosci/biad031 (2023).

10.1126/science.adi3604

Child nutrition requires affordable healthy food

The United Nations' Sustainable Development Goal 2 focuses on efforts to "end hunger, achieve food security and improved nutrition, and promote sustainable agriculture," (*1*) but progress has been insufficient to meet these goals by 2030. The cost of healthy, nutritious food is a major barrier to adequate nutrition, especially in low-income countries (*1*). To address malnutrition and food insecurity, national governments, with assistance from development partners and technical experts, must reform and strengthen agri-food systems to make low-cost, nutrient-dense foods affordable to all.

In 2020, approximately 42% (3.07 billion) of the world's population—including 88% and 69% of the population in low-income and lower-middle-income countries, respectively—could not afford a healthy diet (*1*–*3*). In 12 countries in Sub-Saharan Africa, more than 90% of the population was in this category (*1*). In the 55 countries and territories that faced food crises (*4*), 75.2 million children experienced impaired growth and development (i.e., stunting), and 15.8 million children under the age of 5—a third of which lived in northern Nigeria, Ethiopia, the Democratic Republic of the Congo, and Sudan—suffered from severe degradation of muscle and fat tissue (wasting syndrome) (*4*). Undernutrition causes nearly 45% of child deaths, and stunting affects nearly 155 million children globally (*5*). The unaffordability of nutritious foods is a contributing factor to all forms of malnutrition (*6*).

The World Bank's Food Commodity Price Index increased by 80% between 2020 and 2022 (*7*). Increased prices jeopardize child nutrition because low-income families often rely on cheap, processed foods that are high in calories but low in nutrients. Such foods are often cheaper than fresh fruits and vegetables (*1*). The deterioration of food security and nutritional status among the most vulnerable populations is expected to continue (*3*).

To ensure that all children have access to healthy, nutritious food, national governments should enact policies that improve food affordability, food availability, and nutrition awareness. Policy measures to improve affordability directly could include food vouchers for children and food subsidies for food producers. Government authorities should also promote free and fair trade, eliminate trade barriers, and encourage competition (*8*) to reduce price distortions in local markets (*9*, *10*) and accelerate the use of imported technologies that facilitate cost-effective food production (*1*). To increase food supply and availability, which will contribute to lower prices, governments can boost production by increasing investment in the food and agriculture sectors (*1*). Finally, governments should engage in public education. With the help of community leaders, government agencies should initiate community meetings, distribution of leaflets, and other behavioral change communication activities (*11*) that encourage parents to prioritize nutritious foods.

Md Moyazzem Hossain[1], Mohammed Nazmul Huq[1]*, Faruq Abdulla[2]
[1]Department of Statistics, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh.
[2]Cancer Care and Research Trust Bangladesh (CCRTB), Dhaka 1204, Bangladesh.
*Corresponding author. Email: nhuq@juniv.edu

REFERENCES AND NOTES

1. UN Food and Agriculture Organization, International Fund for Agricultural Development, UN Children's Fund, World Food Programme, World Health Organization, "The state of food security and nutrition in the world 2022: Repurposing food and agricultural policies to make healthy diets more affordable" (2022).
2. W. A. Masters, E. M. Martinez, F. Greb, A. Herforth, S. Hendriks, in *Science and Innovations for Food Systems Transformation*, J. von Braun, K. Afsana, L. O. Fresco, M. H. A. Hassan, Eds. (2023), pp. 603–623.
3. UN Food and Agriculture Organization, International Fund for Agricultural Development, UN Children's Fund, World Food Programme, World Health Organization, "The State of Food Security and Nutrition in the World 2020: Transforming food systems for affordable healthy diets" (2020).
4. *Lancet Child Adolesc. Health* **5**, 459 (2021).
5. World Health Organization, "Malnutrition" (2021).
6. S. S. Morris, A. Garg, R. E. Black, *Nutrit. Rev.* **79**, 1 (2021).
7. J. Baffes, K. Temaj, "Food prices continued their two-year-long upward trajectory," World Bank (2022).
8. A. Lee, D. Patay, L.-M. Herron, E. Parnell Harrison, M. Lewis, *Int. J. Equity Health* **20**, 153 (2021).
9. E. Dabla-Norris, R. Duval, "How lowering trade barriers can revive global productivity and growth," *International Monetary Fund Blog* (2016).
10. UN Conference on Trade and Development, "Trade policies for combating inequality: Equal opportunities to firms, workers and countries" (2019).
11. K. Everett-Murphy, A. De Villiers, E. Ketterer, K. Steyn, *Health Educ. Res.* **30**, 882 (2015).

10.1126/science.adi3972

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

Three-dimensional reconstruction of a fossil lycopod with a spiral morphology that differs from most modern plants

PLANT EVOLUTION

Fossil leaf arrangements

Aerial parts of plants arrange their organs around stems, and this arrangement defines their structure. In most existing plant species, organs emerge at 137.5° from the previous organ. This results in continuous spirals of organs, with the number of clockwise and anticlockwise spirals forming consecutive numbers in a Fibonacci sequence. Reconstructing early lycophyte phyllotaxis from Rhynie Chert fossils, Turner *et al.* found alternative leaf arrangements suggesting that Fibonacci-style patterning was not ancestral to living land plants. The authors also found support for a hypothesis that lycophyte leaves originated from modified reproductive structures rather than evolving de novo. The work provides insight into how diverse early plants evolved into the plant forms seen today. —MRS

Science, adg4014, this issue p. 1189

OPTOELECTRONICS

Broadband graphene photodetectors

The band structure of graphene, together with its superlative transport properties, has long been predicted to be useful in the manufacture of broadband optoelectronic devices such as photodetectors. However, limitations on the absorption and carrier extraction have hindered realizing this goal. Koepfli *et al*. experimentally verified 500-gigahertz bandwidth in a graphene-based photodetector. They incorporated a metamaterial to enhance absorption and engineered the contact design to maximize carrier extraction. An ultrafast photodetector operating with such a wide bandwidth across a large spectral range (1400 to 4200 nanometers) will be useful for applications in sensing and telecommunications. —ISO

Science, adg8017, this issue p. 1169

CHEMICAL PHYSICS

Mastering slow electrons

When condensed matter is exposed to high-energy radiation, low-energy electrons are among the most common and harmful products and can cause damage to biological tissues. The energy-dependent activity of such electrons remains largely unexplored because of difficulties with performing controlled low-energy studies in situ. Using double-imaging photoelectron-photoion coincidence spectroscopy and quantum chemical modeling applied to ultraviolet photoexcitation of metal-ammonia clusters, Hartweg *et al*. demonstrated an unconventional electron transfer–mediated decay process that leads to efficient generation of a high yield of slow electrons. The proposed technique for in situ low-energy electron formation opens the door to practical applications for a broad range of scientific disciplines. —YS

Science, adh0184, this issue p. 1161

CATALYSIS

From support to starring role

In heterogeneous catalysis, some reactions can be optimized by changing the size of the metal nanoparticles. Muravev *et al*. showed that the size of the support of cerium oxide nanoparticles changes the reactivity of highly dispersed palladium catalysts for

low-temperature carbon monoxide oxidation. Single atoms of palladium on the smallest support (4 nanometers in size) were most active when the reaction mixture was carbon monoxide rich, but medium-sized supports (8 nanometers in size) worked better in an oxygen-rich feed. These differences were attributed to the redox properties of the palladium–cerium oxide interface depending on the size of the support. —PDS

Science, adf9082, this issue p. 1174

BASALT FLOWS

Later lava

The Ontong Java Plateau, a submarine, large, igneous province covering nearly 2 million square miles, was formed during the Cretaceous Period in what may have been the largest volcanic eruption in geological history. The canonical age of its emplacement is 128 to 120 million years ago. However, Davidson *et al*. measured high-precision $^{40}Ar/^{39}Ar$ ages that date it to a later time: 117 to 108 million years ago. These dates indicate that the formation of the Ontong Java Plateau did not cause Ocean Anoxic Event 1a, as was previously suggested. —HJS

Science, ade8666, this issue p. 1185

BIOCATALYSIS

A one-pot solution for oligos

Therapeutic oligonucleotides are short stretches of nucleic acid that modulate gene expression and usually have modifications on the sugar or at the phosphate to protect them from degradation. Widespread use of this drug class has been limited in part because of difficulties in synthesis, delivery, and stability. Moody *et al*. developed an isothermal biocatalytic process using a polymerase and an endonuclease. The former extends a template strand with modified nucleotides, and the latter releases the product strand and regenerates the template. The authors demonstrated the versatility of this approach by synthesizing several clinically relevant oligonucleotides. According to a process mass intensity analysis, the biocatalytic approach is potentially less resource and waste intensive than traditional synthesis. —MAF

Science, add5892, this issue p. 1150

COLITIS

Colitis-soothing gut microbiota

Immune checkpoint inhibitor treatment can result in immune-mediated colitis, but remedies for this adverse event that do not compromise cancer treatment are lacking. Halsey *et al*. show in a large case series that fecal microbiota transfer (FMT) from healthy donors promotes resolution of refractory immune checkpoint inhibitor–mediated colitis in cancer patients. Of 12 patients with severe diarrhea or colitis who received FMT, more than half eventually achieved remission of these side effects after FMT alone. This response was linked to increased gut *Collinsella* and *Bifidobacterium* species and decreased colonic $CD8^+$ T cells. —CC

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abq4006

MAGNETISM

Closing in on an unusual phase transition

Transitions between phases with unrelated symmetries are expected to be discontinuous. Nearly two decades ago, a different, continuous kind of transition between such phases was proposed under the name deconfined quantum critical point (DQCP). However, observing the DQCP experimentally has proven to be extremely challenging. Cui *et al*. used nuclear magnetic resonance to study the magnetic field–driven transitions of the layered material $SrCu_2(BO_3)_2$ at various pressures. At high pressures, the experimental results combined with numerical calculations pointed to the existence of a nearby DQCP. —JS

Science, adc9487, this issue p. 1179

IN OTHER JOURNALS

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

STEM WORKFORCE

Trickle-down DEI

Undergraduate STEM populations are becoming more diverse, yet tenured faculty continue to be predominately white men. Marchiondo *et al*. investigated whether informal factors rooted in social influence can heighten faculty support of diversity, equity, and inclusion (DEI) efforts, particularly among men, as a way to increase faculty support for DEI. The team developed a model in which academic leaders' DEI-related attitudes interacted with faculty gender to predict faculty endorsement of institutional DEI initiatives over time. The model was tested using two climate surveys of academic leaders and faculty in the United States. The results showed that men, but not women, reported greater perceived bias in academia when their leaders held more positive DEI attitudes, suggesting that academic leaders are DEI role models and educators, particularly for men. —MMc *J. Divers. High. Educ.* (2023) 10.1037/dhe0000333

SIGNAL TRANSDUCTION

Kinetic proofreading roles

Receptors discriminate between ligands through differences in binding affinities. In some cases, this discrimination is higher than can be expected from the differences in free energy of equilibrium binding. In such cases, kinetic proofreading mechanisms that add steps with intermediate bound states that include energy consumption may increase fidelity. Kirby and Zilman used mathematical modeling to examine the potential role of kinetic proofreading in biological scenarios such as those observed with T cell receptors. By considering the important effect of stochastic intrinsic molecular noise, the authors found that proofreading amplified noise and usually did not help ligand discrimination. Thus, kinetic proofreading mechanisms

ILLUSTRATION: JON KRAUS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

AGING

A cellular view of the aging fly

Aging is a fundamental process in multicellular life, and many phenotypes associated with aging are conserved across mammals and beyond. However, characterizing cellular processes across tissues over the life of an organism has been historically difficult to achieve. Lu *et al*. performed single-nucleus RNA sequencing at multiple time points across the life span of the fruit fly, *Drosophila melanogaster*. After examining more than 850,000 nuclei across 163 cell types, the authors identified changes in gene expression and cellular composition during the aging process and used them to develop an aging clock. This atlas will serve as a resource for studying the process of aging in this important model organism. —CNS

Science, adg0934, this issue p. 1145

NEUROSCIENCE

Neuronal activity alters astrocytes

Astrocytes can respond to input from neuromodulators. However, the roles for neuromodulators in astrocytes in functioning brain circuits are poorly defined. Sardar *et al*. discovered that loss of the astrocytic neuromodulator transporter Slc22a3 resulted in reduced levels of intracellular serotonin and impaired calcium ion–mediated serotonin responses (see the Perspective by Vasile and Rouach). These deficits in serotonin manifest in astrocytes with a Slc22a3 deletion that have reduced histone serotonylation. The expression of genes for the synthesis of γ-aminobutyric acid (GABA) is regulated by histone serotonylation. These changes in GABA gene expression are accompanied by reduced GABA release from astrocytes that lack Slc22a3. Inhibition of histone serotonylation in olfactory bulb astrocytes leads to reduced GABA release and impaired olfactory sensory processing. —PRS

Science, ade0027, this issue p. 1146; see also adi5765, p. 1105

INORGANIC CHEMISTRY

A solid bond for beryllium

Most people who have heard of beryllium know of it as a structural component of emeralds. Its molecular chemistry is more obscure but has provoked decades of controversy over whether the element can be coaxed into bonding with itself. Boronski *et al*. now report a long, elusive example of a molecule stable in solid state at room temperature that features a beryllium–beryllium bond (see the Perspective by Dutton). Each beryllium center is also bonded to a cyclopentadienyl ligand. The compound was characterized by x-ray crystallography and reacted reductively with metal iodides to form bonds with aluminum and zinc. —JSY

Science, adh4419, this issue p.1147; see also adi5762, p. 1106

CONSERVATION

Reef sharks in decline

In recent years, much attention has been given to catastrophic declines in sharks. Most of this attention has focused on large pelagic species that are highly threatened by direct and indirect harvest. Simpfendorfer *et al*. looked globally at the smaller, coral reef–associated species of sharks and rays and found steep declines in shark species (see the Perspective by Shiffman). Five of the most common reef shark species have experienced a decline of up to 73%. As shark species decline on coral reefs, ray species increase, indicating a community-wide shift. Species are best protected when active protections are in place, suggesting routes for better conservation. —SNV

Science, ade4884, this issue p. 1155; see also adi5759, p. 1104

CHEMISTRY

Polariton chemistry is possible

Hybrid light-matter states called polaritons, which are formed by strong interactions between resonant molecular transitions and photonic modes in microcavities, could be used to control chemical reactions with electromagnetic fields, a long-standing goal in chemistry. Unfortunately, such "polariton chemistry" still lacks a series of convincing demonstrations. Ahn *et al*. performed a joint experimental and theoretical study of alcoholysis of phenyl isocyanate with cyclohexanol under various strong light-matter coupling conditions. Through a rigorous analysis of their theoretical and experimental results, the authors provide compelling arguments for how cavity-altered reactivity may arise. These results are needed in this emerging field because they provide an important corroboration of earlier observations that became controversial after several reports of failed attempts. —YS

Science, ade7147, this issue p. 1165

CANCER IMMUNOLOGY

Radiation and immunity

The treatment of solid tumors can include applications of conformal radiotherapy, which allows for the delivery of radiation from several directions to better target tumor tissue. By contrast, preclinical animal models of solid tumors use a less targeted, nonconformal radiotherapy approach, which has been linked to immunosuppressive immune responses. Tadepalli *et al*. show that application of conformal radiotherapy to murine tumor models promotes the recruitment and activation of monocytes, which facilitates the infiltration and function of effector $CD8^+$ T cells that target tumor cells. These recruited monocytes produce interferon I, which potentiates antitumor immune responses. Conversely, nonconformal radiotherapy causes damage to healthy tissue, thus limiting recruitment of these monocytes to tumors, and results in immunosuppressive responses. —CNF

Sci. Immunol. (2023) 10.1126/sciimmunol.add7446

BIOCHEMISTRY

Directing PDK1 substrate specificity

Phosphorylation by the kinase PDK1 is required for the activation of several kinases, including Akt. Sacerdoti *et al.* identified different conformations of PDK1 with distinct substrate specificities (see the Focus by Martínez-Arenas and Bayascas). Small molecules binding to separate sites stabilized distinct monomeric conformations with differential abilities to phosphorylate Akt and other kinases. Because PDK1 activates several kinases that promote cell growth and survival, these results may provide ways to selectively manipulate PDK1 activity for anticancer therapies. —WW

Sci. Signal. (2023) 10.1126/scisignal.add3184, 10.1126/scisignal.adh5114

low-temperature carbon monoxide oxidation. Single atoms of palladium on the smallest support (4 nanometers in size) were most active when the reaction mixture was carbon monoxide rich, but medium-sized supports (8 nanometers in size) worked better in an oxygen-rich feed. These differences were attributed to the redox properties of the palladium–cerium oxide interface depending on the size of the support. —PDS

Science, adf9082, this issue p. 1174

BASALT FLOWS

Later lava

The Ontong Java Plateau, a submarine, large, igneous province covering nearly 2 million square miles, was formed during the Cretaceous Period in what may have been the largest volcanic eruption in geological history. The canonical age of its emplacement is 128 to 120 million years ago. However, Davidson *et al*. measured high-precision $^{40}Ar/^{39}Ar$ ages that date it to a later time: 117 to 108 million years ago. These dates indicate that the formation of the Ontong Java Plateau did not cause Ocean Anoxic Event 1a, as was previously suggested. —HJS

Science, ade8666, this issue p. 1185

BIOCATALYSIS

A one-pot solution for oligos

Therapeutic oligonucleotides are short stretches of nucleic acid that modulate gene expression and usually have modifications on the sugar or at the phosphate to protect them from degradation. Widespread use of this drug class has been limited in part because of difficulties in synthesis, delivery, and stability. Moody *et al*. developed an isothermal biocatalytic process using a polymerase and an endonuclease. The former extends a template strand with modified nucleotides, and the latter releases the product strand and regenerates the template. The authors demonstrated the versatility of this approach by synthesizing several clinically relevant oligonucleotides. According to a process mass intensity analysis, the biocatalytic approach is potentially less resource and waste intensive than traditional synthesis. —MAF

Science, add5892, this issue p. 1150

COLITIS

Colitis-soothing gut microbiota

Immune checkpoint inhibitor treatment can result in immune-mediated colitis, but remedies for this adverse event that do not compromise cancer treatment are lacking. Halsey *et al*. show in a large case series that fecal microbiota transfer (FMT) from healthy donors promotes resolution of refractory immune checkpoint inhibitor–mediated colitis in cancer patients. Of 12 patients with severe diarrhea or colitis who received FMT, more than half eventually achieved remission of these side effects after FMT alone. This response was linked to increased gut *Collinsella* and *Bifidobacterium* species and decreased colonic $CD8^+$ T cells. —CC

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abq4006

MAGNETISM

Closing in on an unusual phase transition

Transitions between phases with unrelated symmetries are expected to be discontinuous. Nearly two decades ago, a different, continuous kind of transition between such phases was proposed under the name deconfined quantum critical point (DQCP). However, observing the DQCP experimentally has proven to be extremely challenging. Cui *et al*. used nuclear magnetic resonance to study the magnetic field–driven transitions of the layered material $SrCu_2(BO_3)_2$ at various pressures. At high pressures, the experimental results combined with numerical calculations pointed to the existence of a nearby DQCP. —JS

Science, adc9487, this issue p. 1179

IN OTHER JOURNALS

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

STEM WORKFORCE

Trickle-down DEI

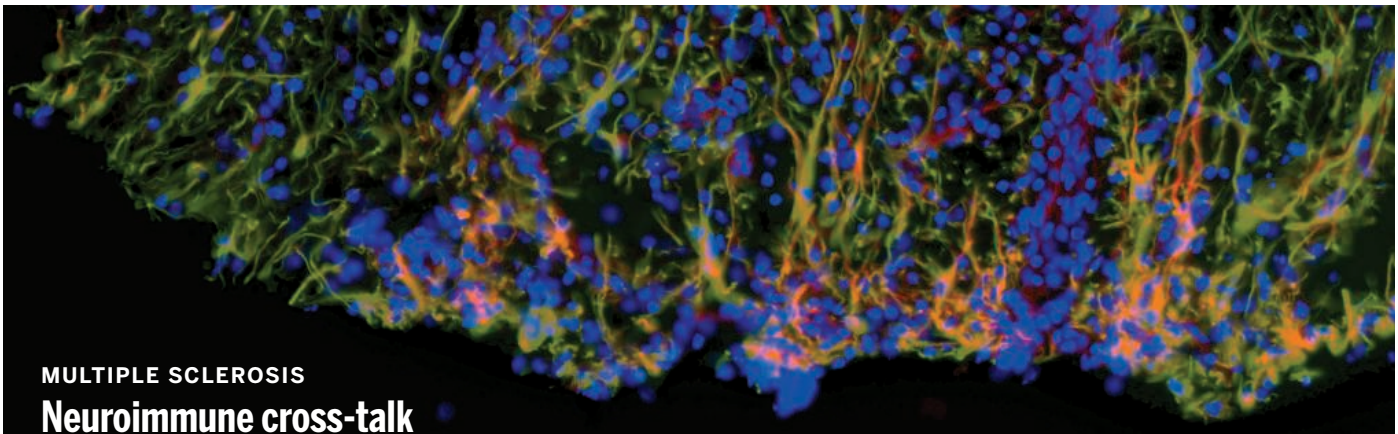
Undergraduate STEM populations are becoming more diverse, yet tenured faculty continue to be predominately white men. Marchiondo *et al.* investigated whether informal factors rooted in social influence can heighten faculty support of diversity, equity, and inclusion (DEI) efforts, particularly among men, as a way to increase faculty support for DEI. The team developed a model in which academic leaders' DEI-related attitudes interacted with faculty gender to predict faculty endorsement of institutional DEI initiatives over time. The model was tested using two climate surveys of academic leaders and faculty in the United States. The results showed that men, but not women, reported greater perceived bias in academia when their leaders held more positive DEI attitudes, suggesting that academic leaders are DEI role models and educators, particularly for men. —MMc *J. Divers. High. Educ.* (2023) 10.1037/dhe0000333

SIGNAL TRANSDUCTION

Kinetic proofreading roles

Receptors discriminate between ligands through differences in binding affinities. In some cases, this discrimination is higher than can be expected from the differences in free energy of equilibrium binding. In such cases, kinetic proofreading mechanisms that add steps with intermediate bound states that include energy consumption may increase fidelity. Kirby and Zilman used mathematical modeling to examine the potential role of kinetic proofreading in biological scenarios such as those observed with T cell receptors. By considering the important effect of stochastic intrinsic molecular noise, the authors found that proofreading amplified noise and usually did not help ligand discrimination. Thus, kinetic proofreading mechanisms

ILLUSTRATION: JON KRAUS



MULTIPLE SCLEROSIS

Neuroimmune cross-talk

Multiple sclerosis (MS) is an immune-mediated disease that affects the central nervous system (CNS). To better understand the underlying inflammatory process that contributes to MS, Kiss *et al.* investigated the cross-talk between immune cells and glial cells. The authors identified two cell types, CNS-resident astrocytes and $CD44^{hi}CD4^{+}$ T lymphocytes, that could produce the cytokine interleukin-3 (IL-3). The production of IL-3 resulted in an immune migratory response by myeloid cells that expressed IL-3 receptor-α (IL-3Rα). IL-3:IL-3Rα cross-talk was found to increase CNS immune cell infiltration and to exacerbate MS in a mouse model. In patient CNS tissue, IL-3:IL-3Rα levels correlated with immune cell recruitment and MS severity. —PNK

Immunity (2023) 10.1016/j.immuni.2023.04.013

Immunofluorescent image of the spinal cord of a mouse with multiple sclerosis showing colocalization (orange) of IL–3 (magenta) with an astrocyte marker (green); blue shows DAPI staining for nuclear DNA

seen at biological receptors may serve other purposes, such as the detection of ligand concentration, multiple ligands in mixtures, or other environmental signals. —LBR

Proc. Natl. Acad. Sci. U. S. A. (2023) 10.1073/pnas.2212795120

NEUROSCIENCE

Brain white matter and reaction time

Modeling human behavior has mostly focused on the analysis of mean performance. However, there is much trial-to-trial variability of behavior around an individual's mean. McCormick *et al.* combined reaction time measurements with brain scanning. Using a technique called "dynamic structural equation modeling," they fitted a model that separates out four distinct components of human reaction time behavior: mean performance, variability in response times, linear trends across the task, and trial-to-trial inertia in performance. In a large group of participants 18 to 102 years of age, they used this model to examine the associations between these performance components across aging and individual differences in white matter microstructure. Poorer white matter microstructure predicted increased variability and decreased mean level performance. —PRS

J. Neurosci. (2023) 10.1523/JNEUROSCI.1042-22.2023

BLACK MORTALITY

Discriminatory deaths

While the United States continues to address health inequities between Black and white Americans, there is a need to evaluate how much progress was made during the past two decades. To understand the health consequences of race, important metrics to track include measures of excess deaths among Black compared with white individuals and potential years of life lost. Caraballo *et al.* tracked this parameter from 1999 to 2020 using death certificate data from the Centers for Disease Control and Prevention (CDC). Among Black men, the mortality rate substantially declined between 1999 and 2011, plateaued until 2019, and by 2020 returned to 2005 levels. Among Black women from 1999 to 2015, the mortality rate significantly declined, but by 2020, it also returned to 2005 levels. This was largely driven by heart disease in both sexes and cancer among men. The findings carry policy implications for the role of structural racism in exacerbating disparities in 2020 during the early stages of the pandemic. —EEU

JAMA (2023) 10.1001/jama.2023.7022

HYBRID MATERIALS

Rational design of hybrid materials

Solid-state hybrid organic-inorganic materials form intricate networks of noncovalent interactions, which complicates our understanding of these systems' formation process and electronic properties. Using periodic density functional theory calculations and experimental calorimetry data, Rajapaksha *et al.* present a method to explore the relationships between structural and energetic descriptors in uranyl tetrahalide systems. They demonstrated the importance of analyzing the entire three-dimensional noncovalent interaction network, not just simple molecular systems, because it significantly affects the thermodynamic stability and overall properties of hybrid materials. In the future, the proposed technique could be combined with machine learning approaches, guiding the rational design of new hybrid materials. —YS

Angew. Chem. Int. Ed. (2023) 10.1002/anie.202305073

TOPOCHEMISTRY

Swinging by to react

Product formation in photocyclization reactions of alkene can be controlled by crystallizing the reactants and thus fixing their relative orientation, provided that the double bonds are close enough to react (about 4 angstroms). Nakagawa *et al.* show that an alkene, *trans*-4-styrylpyridine, confined in a metal-organic framework, can undergo [2+2] photocyclization despite the double bonds being almost 6 angstroms apart. Deuterium solid-state nuclear magnetic resonance studies suggest that a dynamic "swing" motion of the guest molecules brings them close enough to undergo reaction. —PDS

J. Am. Chem. Soc. (2023) 10.1021/jacs.3c01225

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Induction of astrocytic Slc22a3 regulates sensory processing through histone serotonylation

Debosmita Sardar, Yi-Ting Cheng†, Junsung Woo†, Dong-Joo Choi, Zhung-Fu Lee, Wookbong Kwon, Hsiao-Chi Chen, Brittney Lozzi, Alexis Cervantes, Kavitha Rajendran, Teng-Wei Huang, Antrix Jain, Benjamin R. Arenkiel, Ian Maze, Benjamin Deneen*

INTRODUCTION: Astrocytes are intimately associated with neurons and participate in a host of essential roles that facilitate synaptic transmission and circuit function. In neurons, heightened activity induces the expression of "immediate early genes," which are predominately transcription factors that modify gene expression programs and activity-dependent epigenomic states, ultimately regulating circuit activity, plasticity, and associated behavioral outputs. However, whether heightened neuronal activity induces an analogous immediate early gene–like response in mature astrocytes and how this sculpts astrocytic transcriptional and epigenomic responses to regulate circuit function remain unclear.

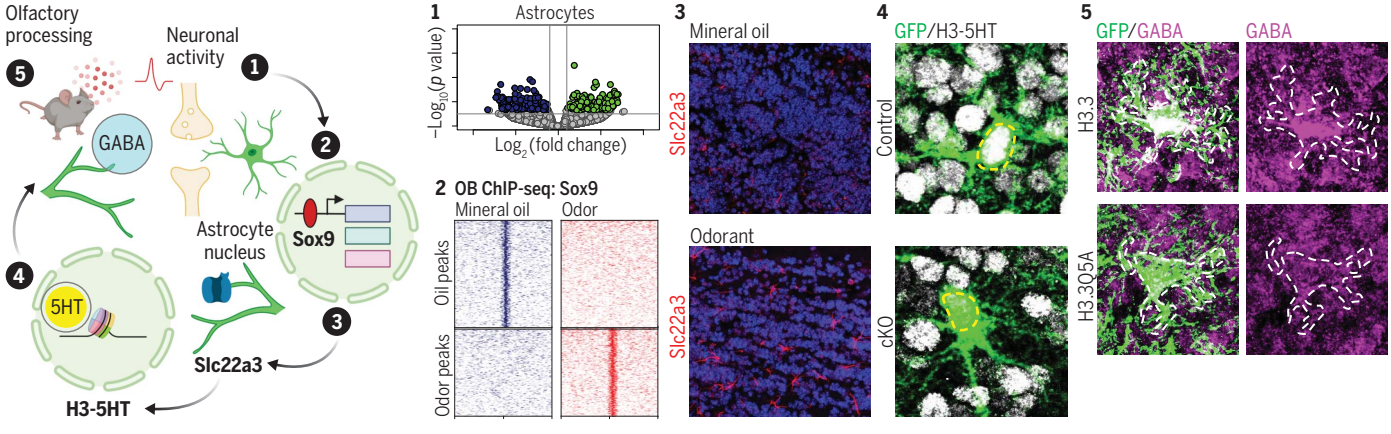
RATIONALE: Prior studies have shown that non-neuronal cells undergo transcriptional changes in response to neuronal activity. A series of recent studies demonstrated that astrocytic transcription factors regulate region specific–circuits in the adult brain, indicating that diverse transcriptional states in astrocytes can influence neuronal activity. Therefore, we used astrocyte-specific transcription factors to interrogate drivers of neuronal activity–dependent changes in astrocytes. This revealed astrocyte transcription factor targets that are critical for maintaining astrocyte-neuron communication, and, using a model of olfactory processing in combination with an astrocyte conditional knockout mouse model, we further investigated how an astrocytic activity–dependent target regulates epigenomic states in astrocytes to regulate animal behavior.

RESULTS: We show that neuronal activity induces widespread transcriptional changes in astrocytes, and, in the olfactory bulb, these transcriptional changes are driven by alterations in DNA binding of the transcription factor Sox9. Using Sox9 as a molecular entry point, we screened through neuronal activity–dependent targets in astrocytes, which led to the identification of *Slc22a3* as an activity-inducible astrocyte gene that encodes neuromodulator transporter Slc22a3, which regulates sensory processing in the olfactory bulb. Region-specific and cell type–specific deletion of olfactory bulb astrocytic Slc22a3 led to defects in olfactory bulb circuits and odor processing, implying that the loss of astrocytic Slc22a3 affects astrocyte-neuron communication in the olfactory bulb. Loss of astrocytic Slc22a3 led to reduced astrocyte morphological complexity and diminished calcium activity in the presence of neurotransmitters and neuromodulators. Mechanistically, Slc22a3 regulates the transport of serotonin into olfactory bulb astrocytes and coordinates histone serotonylation (H3-5HT) to regulate the expression of astrocytic γ-aminobutyric acid (GABA)–associated genes and olfactory behaviors. Finally, attenuation of astrocytic H3-5HT in the olfactory bulb was achieved using mutant histone variant 3.3 (H3.3Q5A), which was compared against control (H3.3). This led to reduced expression of astrocytic GABA biosynthetic genes and astrocytic tonic GABA release, culminating in olfactory deficits.

CONCLUSION: Our study reveals how neuronal activity orchestrates transcriptional responses in astrocytes and identifies the astrocyte transcription factor Sox9 as a mediator of these responses. We identify *Slc22a3* as an odor-responsive, astrocyte gene in the olfactory bulb that encodes neuromodulator transporter Slc22a3 and is required to maintain astrocyte-neuron communication and olfactory sensory processing. Mechanistically, Slc22a3 uses epigenomic processes that involve the transport of the monoamine neuromodulator serotonin into olfactory bulb astrocytes, which subsequently regulates gene expression through histone serotonylation. Taken together, these results identify new mechanisms for how astrocytes integrate neuromodulator signaling to gate neurotransmitter release for sensory processing. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: deneen@bcm.edu
†These authors contributed equally to this work.
Cite this article as D. Sardar *et al.*, *Science* **380**, eade0027 (2023). DOI: 10.1126/science.ade0027

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.ade0027



Neuronal activity–dependent astrocyte transcriptional and epigenomic changes control olfactory processing. Neuronal activity induces widespread transcriptional changes in astrocytes (1), which is exemplified by Sox9 DNA-binding changes after odor-evoked neuronal activation (2). Slc22a3 (3) represents a Sox9-regulated and activity-dependent olfactory bulb target that controls astrocytic serotonin (5HT) levels to coordinate histone serotonylation (4), a previously uncharacterized epigenomic mediator of astrocytic GABA release that ultimately affects olfactory processing (5). GFP, green fluorescent protein; OB, olfactory bulb; ChIP-seq, chromatin immunoprecipitation coupled with next-generation sequencing.

RESEARCH ARTICLE

NEUROSCIENCE

Induction of astrocytic Slc22a3 regulates sensory processing through histone serotonylation

Debosmita Sardar[1,2], Yi-Ting Cheng[1,2,3]†, Junsung Woo[1,2]†, Dong-Joo Choi[1,2], Zhung-Fu Lee[1,2,4], Wookbong Kwon[1,2], Hsiao-Chi Chen[1,2,5], Brittney Lozzi[1,2,6], Alexis Cervantes[1,2], Kavitha Rajendran[1,2], Teng-Wei Huang[1], Antrix Jain[7], Benjamin R. Arenkiel[3,8,9], Ian Maze[10,11,12], Benjamin Deneen[1,2,3,4,13]*

Neuronal activity drives alterations in gene expression within neurons, yet how it directs transcriptional and epigenomic changes in neighboring astrocytes in functioning circuits is unknown. We found that neuronal activity induces widespread transcriptional up-regulation and down-regulation in astrocytes, highlighted by the identification of *Slc22a3* as an activity-inducible astrocyte gene that encodes neuromodulator transporter Slc22a3 and regulates sensory processing in the mouse olfactory bulb. Loss of astrocytic Slc22a3 reduced serotonin levels in astrocytes, leading to alterations in histone serotonylation. Inhibition of histone serotonylation in astrocytes reduced the expression of γ-aminobutyric acid (GABA) biosynthetic genes and GABA release, culminating in olfactory deficits. Our study reveals that neuronal activity orchestrates transcriptional and epigenomic responses in astrocytes while illustrating new mechanisms for how astrocytes process neuromodulatory input to gate neurotransmitter release for sensory processing.

Astrocytes are intimately associated with neuronal activity and neurotransmission and participate in a host of essential roles that facilitate synaptic function (*1*–*5*). Cell-intrinsic activation of G protein subunit G_q or G_i signaling and associated calcium activity in astrocytes can influence a multitude of behavioral outputs (*6*–*10*). However, in native behavioral states, astrocyte activation results from cell-extrinsic sources, usually in response to a neuronal stimulus, and therefore typically follows neuronal activation (*11*). In neurons, heightened activity induces expression of "immediate early genes," which are predominately transcription factors (TFs) that modify gene expression programs and activity-dependent epigenomic states, ultimately regulating circuit activity, plasticity, and associated behavioral outputs (*12*). Astrocytic TFs can also regulate region-specific circuits in the adult brain (*13*, *14*), indicating that transcriptional states in astrocytes can influence neuronal activity. Evidence for the reciprocal relationship can be found in the developing cortex (*15*) and in vitro (*16*, *17*), where alterations in neuronal activity influence expression of synaptogenic genes in differentiating astrocytes. Single-cell transcriptomic studies have shown that non-neuronal cells undergo transcriptional changes in response to neuronal activity that is driven by visual stimuli (*18*). However, whether heightened neuronal activity induces an analogous "immediate early gene–like" response in mature astrocytes and how this sculpts astrocytic transcriptional and epigenomic responses to regulate circuit function remain unclear.

A central component of astrocyte-neuron communication is synaptic neurotransmitter signaling because astrocytes express receptors and transporters for both glutamate and γ-aminobutyric acid (GABA), the activation of which can elicit calcium-dependent responses that modify circuit activity (*5*, *19*). Maintenance of extracellular glutamate levels by means of astrocytic glutamate transporters are required for drug-seeking behaviors, circadian rhythms, and avoidance behaviors, among others (*20*–*23*). Similarly, disruption of GABA signaling in astrocytes leads to repetitive behaviors, impaired motor function, and deficits in learning and memory (*7*, *24*–*26*). Astrocytes also express a host of neuromodulator receptors and respond to noradrenaline, acetylcholine, and serotonin with increases in intracellular calcium (*16*). Manipulations of cholinergic inputs to hippocampal astrocytes influence sleep states, whereas noradrenaline primes cortical astrocyte responses to local neuronal activity during locomotion and dopamine promotes astrocyte-mediated depression of excitatory synapses, which influences drug-induced locomotor activity (*27*–*29*). Despite direct roles in gating responses to neuromodulators, the signaling mechanisms used by astrocytes to respond and process neuromodulatory cues remain unclear. Furthermore, how astrocytes coordinate the integration of neuromodulator with neurotransmitter signaling to regulate circuit function is also unclear.

Epigenomic regulation of gene expression participates in encoding information in the central nervous system (CNS), where alterations in synaptic activity induce epigenetic modifications that influence behavioral outcomes (*12*). For example, alterations in histone acetylation are directly linked to memory storage and synaptic plasticity in hippocampal neurons after learning (*30*). Among neuromodulators, serotonin and dopamine can be incorporated into histones, with histone serotonylation enabling permissive gene expression states in serotonergic neurons (*31*) and histone dopaminylation regulating cocaine-seeking behaviors in rodents (*32*). Nevertheless, several aspects of this biology are still unclear, including the extent and importance of these histone modifications throughout the brain and the fundamental mechanisms of how neuromodulators are transported to the nucleus and added to histones. Furthermore, how these and other epigenomic mechanisms are used by astrocytes to regulate gene expression, circuit function, and associated behavioral responses is unknown.

Neuronal activity directs transcriptional responses in astrocytes

To understand how neuronal activation influences transcriptional responses in astrocytes, we used chemogenetics to activate neurons in Aldh1l1–GFP (green fluorescent protein) reporter mice (Fig. 1, A and B). We used intraventricular injection of viral expression vector (pAAV; AAV, adeno-associated virus) containing hM3Dq DREADD (Gq-DREADD), a DREADD (designer receptors exclusively activated by designer drugs) based on human muscarinic receptor hM3, under the pan-neuronal synapsin promoter, which allowed widespread neuronal Gq-DREADD expression throughout the brain. When mice reached early adulthood (~8 weeks), we activated neuronal G_q signaling with intraperitoneal injection of clozapine-*N*-oxide (CNO) (0.3 mg/kg) or saline (represented as Gq-Saline or Gq-CNO herein) and harvested brains 30 min after treatment (Fig. 1C). We confirmed enhanced neuronal activity in Gq-CNO–treated samples using electrophysiology in brain-slice preparations,

[1]Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, TX, USA. [2]Center for Cancer Neuroscience, Baylor College of Medicine, Houston, TX, USA. [3]Program in Developmental Biology, Baylor College of Medicine, Houston, TX, USA. [4]Program in Development, Disease Models, and Therapeutics, Baylor College of Medicine, Houston, TX, USA. [5]The Integrative Molecular and Biomedical Sciences Graduate Program, Baylor College of Medicine, Houston, TX, USA. [6]Genetics and Genomics Graduate Program, Baylor College of Medicine, Houston, TX, USA. [7]Mass Spectrometry Proteomics Core, Baylor College of Medicine, Houston, TX, USA. [8]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [9]Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA. [10]Nash Family Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [11]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [12]Howard Hughes Medical Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [13]Department of Neurosurgery, Baylor College of Medicine, Houston, TX, USA.
*Corresponding author. Email: deneen@bcm.edu
†These authors contributed equally to this work.

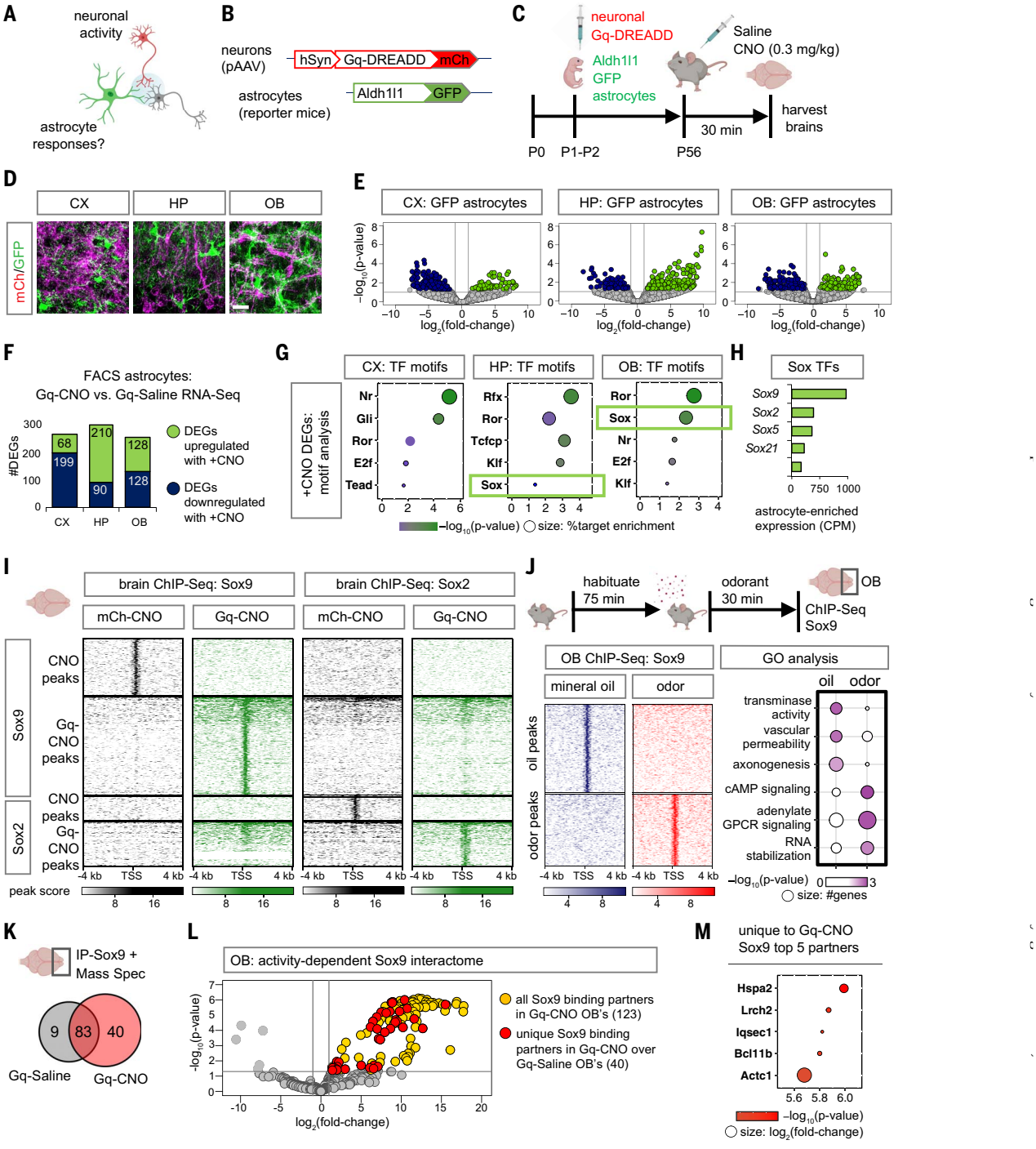


Fig. 1. Neuronal activity directs Sox9-regulated transcriptional responses in astrocytes. (**A** to **C**) Schematics illustrating the chemogenetic neuronal activation experimental design. (**D**) Sections showing distinct labeling of astrocytes (GFP) and Gq-DREADD neurons (mCh) in the CX, HP, and OB. Scale bar, 25 μm. (**E**) Volcano plots depicting RNA-seq from GFP astrocytes comparing the Gq-CNO and Gq-Saline groups. (**F**) Number of DEGs that are up-regulated (green) or down-regulated (blue) in GFP astrocytes in the Gq-CNO versus Gq-Saline group [n = 3 per cohort, $p < 0.05$, $\log_2$(fold change) >1 or <1]. (**G**) Significant TF motifs ($p < 0.05$) enriched in these DEGs and exhibiting astrocyte-specific expression. (**H**) Average transcript expression of Sox family TFs in GFP astrocytes. CPM, counts per million. (**I**) Comparison of heatmaps of ChIP-Sox9 and ChIP-Sox2 at 4 kb from the peak center in the Gq-CNO group versus mCh-CNO control group (n = 3 per cohort). (**J**) Schematic for odor-evoked neuronal activation in the OB. Heatmaps of ChIP-Sox9 at 4 kb from the peak center for mineral oil–exposed versus odor-exposed mice are shown on the left. Enriched GO terms associated with these peaks (n = 6 per cohort) are shown on the right. (**K**) Sox9 binding partners in OBs from the Gq-CNO versus Gq-Saline groups. (**L**) Volcano plot depicting immunoprecipitation coupled with mass spectrometry data of the Sox9 interactome in Gq-CNO OBs. Fold change was calculated over control lysates incubated with beads only and without antibody. Sox9 binding partners specific to the Gq-CNO versus Gq-Saline groups are highlighted in red [n = 9 to 12 per cohort, $p < 0.05$, $\log_2$(fold change) >1]. (**M**) Top 5 Sox9 interactors specific to the Gq-CNO group [$p < 0.0001$, $\log_2$(fold change) >8.5]. Color coding represents p values (x axis), and sizes represent the $\log_2$(fold change), with larger circles denoting greater binding affinity. Illustrations were created with Biorender.com.

coupled with expression of neuronal immediate early-gene marker c-Fos (fig. S1, A to C). To identify neuronal activity–dependent gene expression changes in astrocytes, we used the presence of distinct reporters in Gq-DREADD–expressing neurons (mCherry, or mCh) and astrocytes (GFP) to isolate these two cell types from the cortex (CX), hippocampus (HP), and olfactory bulb (OB) (Fig. 1D) by fluorescence-activated cell sorting (FACS) and performed RNA sequencing (RNA-seq) (fig. S1D). Analysis of Gq-CNO versus Gq-Saline cohorts from each region revealed widespread changes in gene expression in astrocytes, and we identified 267, 300, and 256 astrocyte-specific differentially expressed genes (DEGs) in the CX, HP, and OB, respectively (Fig. 1, E and F). These DEGs were specific to astrocytes, demonstrating that immediate early responses in neurons are distinct from those in astrocytes (fig. S2, A to C). Gene ontology (GO) analyses revealed distinct terms associated with astrocytes and neurons: Top astrocytic GO terms included protein phosphorylation, calcium ion transport, and serotonin transport in the CX, HP, and OB, respectively (fig. S2, D to F, and table S1). To control for possible transcriptional artifacts due to CNO treatment, we treated Aldh1l1-GFP mice with 0.3 mg/kg CNO (represented as CNO-control herein), harvested their OBs after 30 min, and isolated GFP-expressing astrocytes by FACS, which was followed by RNA-seq (fig. S3A). By comparing RNA-seq profiles of untreated OB astrocytes with those of CNO-control OB astrocytes, we identified widespread changes in gene expression (fig. S3B). However, these CNO-induced changes do not correspond to any of the GO groups that we identified in our activity-based profiling experiments (fig. S3C). A comparison of the activity-induced DEGs with the CNO-associated DEGs revealed miniscule overlap. Among the 256 DEGs induced in OB astrocytes by neuronal activity, only seven also demonstrated altered expression in the CNO-treated control Aldh1l1-GFP mice (fig. S3, D and E).

To dissect the transcriptional mechanisms that underlie these changes in astrocyte gene expression, we used a TF motif analysis on the astrocyte DEGs. By filtering for TFs that also exhibit enriched expression in astrocytes (fig. S4A), we identified a cohort of TFs with motifs associated with activity-dependent DEGs in astrocytes, including the Sox-family of TFs in both the HP and OB (Fig. 1G). The Sox family has multiple members that exhibit elevated expression in astrocytes, with Sox9 and Sox2 being the most notable (Fig. 1H). Sox9 is critical for olfactory circuit processing (*13*), whereas Sox2 plays a role in responses to brain injury (*33*). We next evaluated Sox9 and Sox2 protein levels after neuronal activation and observed no significant changes in protein expression, as measured by both immunostaining (fig. S4B) and immunoblotting (fig. S4, C and D). Therefore, we asked if these Sox TFs undergo changes in their transcriptional activity after neuronal activation. We performed chromatin immunoprecipitation coupled with next-generation sequencing (ChIP-seq) of Sox9 and Sox2 in CNO-treated Gq-DREADD (Gq-CNO)–expressing mice as well as CNO-treated control pAAV-mCh–expressing mice (represented as mCh-CNO herein) and harvested brains 30 min after CNO treatment. Although both Sox9 and Sox2 are equally immunoprecipitated in Gq-CNO and mCh-CNO control lysates (fig. S5A), we found that both of these Sox TFs exhibit >1.5-fold increase in DNA binding capacity in Gq-CNO brains compared with mCh-CNO control brains (Fig. 1I and fig. S5, B and C). Differential ChIP-seq peak analysis between Gq-CNO treatment and controls revealed 295 Sox9-specific peaks, 139 Sox2-specific peaks, and 117 Sox9 and Sox2 shared peaks (fig. S5D). GO analysis revealed that activity-dependent peaks were associated with dendrite development (Sox9-specific) and proteoglycans (Sox2-specific), whereas Sox9 and Sox2 shared peaks were enriched in metabolic and mRNA processing (fig. S5E).

We next examined whether such changes in TF DNA binding would also occur after native stimulation. Because Sox9 exhibited greater DNA binding capacity, and astrocytic Sox9 is critical for olfactory sensory circuit function in the OB (*13*), we examined whether neuronal activation driven by olfactory sensory input influences Sox9 transcriptional activity. In the OB, Sox9 expression is restricted to astrocytes (>96% Aldh1l1-GFP$^+$ cells) and has miniscule expression in the oligodendroglial lineage (2.79% Olig2$^+$ cells) (fig. S6, A and B). To examine Sox9 transcriptional activity after native olfactory-driven neuronal stimuli, we performed Sox9 ChIP-seq on OBs from mice sacrificed after exposure to the odorant isoamyl acetate or mineral oil control for 30 min, which revealed largely distinct Sox9 peaks (Fig. 1J and fig. S6C). Associated GOs showed that whereas Sox9 peaks in mineral oil–exposed OBs were enriched in transaminase activity GO terms, peaks in odor-exposed OBs were enriched in cyclic adenosine monophosphate (cAMP) signaling GO terms (Fig. 1J). To assess whether these changes in Sox9 DNA binding capacity after neuronal activation are due to changes in Sox9 binding partners, we performed Sox9-IP coupled with mass spectrometry and observed a >4-fold increase in Sox9 interactions after neuronal activation (Fig. 1K and table S2), identifying 40 proteins that specifically interact with Sox9 in the Gq-CNO–treated OBs compared with Gq-Saline control OBs (Fig. 1L). We observed overlap between the top activity-specific Sox9 interacting proteins (Fig. 1M) and TFs identified from motif analysis of activity-specific ChIP-Sox9 peaks in the OB (fig. S6D).

Slc22a3 is an activity-dependent gene target of Sox9 in OB astrocytes

To identify astrocyte genes that are induced after neuronal activity, we compared the activity-dependent astrocyte-specific DEGs in the OB (Fig. 1E) with existing RNA-seq data from OB astrocytes from Sox9 conditional knockout (Sox9-cKO) mice (*13*), in which Sox9 is deleted specifically from OB astrocytes. We identified 62 genes that are activity-dependent and Sox9 regulated, which represents 24.2% of activity-dependent DEGs in OB astrocytes (fig. S7A). Among these, the 34 genes that demonstrated decreased expression in the Sox9-cKO mice (Fig. 2A and fig. S7B) were further filtered based on the Sox9 motif binding score at their promoters (Fig. 2B). In a ChIP–PCR (polymerase chain reaction) screen, we immunoprecipitated DNA bound to Sox9 and performed PCR (fig. S7C) using primers spanning Sox9 motifs at promoter regions of candidate genes. Among the candidates that demonstrated Sox9 promoter occupancy in the OB specifically after G_q-based neuronal stimulation was *Slc22a3* (Fig. 2, C and D, and fig. S7D). To validate Sox9-regulation of Slc22a3, we evaluated its protein expression in Sox9-cKO mice (fig. S8A) (*13*) and found a 76% reduction of its expression in astrocytes from the OB (Fig. 2, E and F) but not the CX or HP (fig. S8, B to E).

We examined whether astrocytic Slc22a3 is regulated by neuronal activity by evaluating its protein and transcript expression in our Gq-DREADD–based paradigm. Its expression was up-regulated 2.22-fold in Gq-CNO–treated OB astrocytes compared with Gq-Saline and CNO-control OB astrocytes (Fig. 2, G and H, and fig. S9, A to E). We investigated whether neuronal activation from olfactory sensory input promotes Slc22a3 expression in the OB. We exposed habituated Aldh1l1-GFP mice to mineral oil or the odorant isoamyl acetate. Exposure was maintained for 40 min, and after removal of oil or odorant, mice were kept in the chamber for an additional 20 min to allow protein turnover (Fig. 2I). We evaluated Fos expression, a marker of neuronal activity, which revealed increased protein expression in lysates obtained from the OB (Fig. 2J and fig. S10A), and we observed a significant 88.8% increase in neurons, but not in astrocytes, of OBs of odor-exposed mice compared with OBs of mineral oil–exposed control mice (Fig. 2, K to O, and fig. S10B). Next, we assessed Slc22a3 and observed a significant 58.9% increase in its astrocytic expression in odor-exposed mice compared with mineral oil–exposed and nonhabituated control mice (Fig. 2, P to T, and fig. S10, C to E). These analyses were performed in the granular cell layer of the OB, given both Fos and

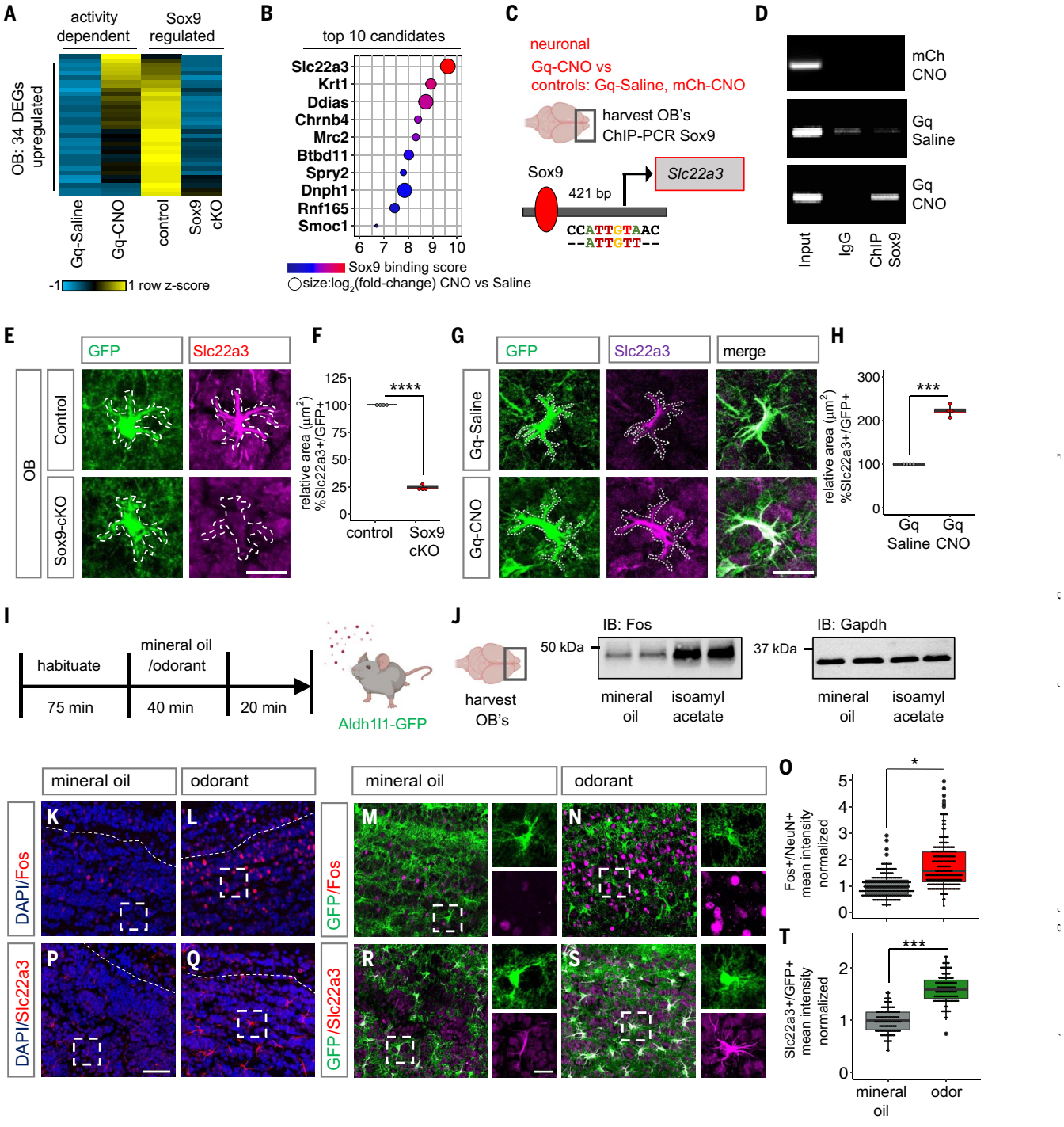


Fig. 2. Neuronal activation induces Slc22a3 expression in OB astrocytes. (**A** and **B**) Heatmap depicting the 34 neuronal activity–dependent and Sox9-regulated DEGs (n = 3, $p < 0.05$) (A) and the top 10 candidates filtered by Sox9 motif binding score 1000 base pairs (bp) from the transcription start site of candidates (B). (**C** and **D**) ChIP-PCR screen identifying activity-dependent Sox9 regulation at the *Slc22a3* promoter (n = 4 to 6 OBs). (**E** to **H**) Slc22a3 expression in control versus Sox9-cKO OB astrocytes and in Gq-Saline versus Gq-CNO [(E) and (G)] and box plots depicting the area covered by Slc22a3 in GFP astrocytes [(F) and (H)] (average of 116 to 122 cells per cohort; ****$p = 1.93 \times 10^{-8}$ and ***p = 0.00013, unpaired Student's two-tailed t test on n = 4 mice per cohort). Scale bar, 20 μm. (**I**) Schematic illustrating the odor-evoked neuronal activation experimental design. (**J**) Immunoblots (IBs) of Fos and loading control from OB lysates (n = 4 per cohort). (**K** to **O**) Immunostaining of Fos in Aldh1l1-GFP mice [(K) to (N)] and quantification of mean fluorescence intensity in NeuN$^+$ neurons (O) (244 to 250 cells per cohort; *p = 0.0285, Wilcoxon rank sum test on n = 4 mice per cohort). (**P** to **T**) Immunostaining of Slc22a3 in Aldh1l1-GFP mice [(P) to (S)] and quantification of mean fluorescence intensity in GFP$^+$ astrocytes (T) (130 or 131 cells per cohort; ***p = 0.00039, unpaired Student's two-tailed t test on n = 4 mice per cohort). Scale bar, 50 μm; inset scale bar, 10 μm. The dashed lines represent the boundary of the granule cell layer. See table S3 for data summary. For the box plots in (F), (H), (O), and (T), the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. Illustrations were created with Biorender.com.

Slc22a3 enrichment in this subregion of the OB. Moreover, we observed a small increase in the number of astrocytes that express Slc22a3 (fig. S10F), and increased levels of Slc22a3 expression were observed only in astrocytes and not in neurons, as evaluated by colabeling with NeuN (fig. S10G).

Astrocytic Slc22a3 is required for olfactory sensory processing

Slc22a3 is a membrane transport protein that is involved in both the release and uptake of the catecholamine family of chemicals, which includes the neuromodulators serotonin, dopamine, and noradrenaline (*34*–*38*). Slc22a3 is expressed in astrocytes (*36*, *37*) (fig. S11, A and B); however, its contributions to astrocyte function and associated brain circuits are unknown. We determined whether Slc22a3 contributes to OB astrocyte function and olfactory circuits. To achieve selective knockout of Slc22a3, we generated a transgenic mouse line containing a floxed Slc22a3 allele (Slc22a3-FF) in combination with OB injection of pAAV containing Cre that is driven by the astrocyte-specific Gfap promoter (fig. S11, C and D). Injections occurred at postnatal day 60 (P60), and mice were harvested at P90, which enabled region- and cell type–specific deletion of Slc22a3 in OB astrocytes (fig. S11, E and F) but not in neurons (fig. S11, G to I). To quantify Slc22a3 knockout efficiency in astrocytes, we generated *Slc22a3-FF; Aldh1l1-GFP* mice, a double-transgenic mouse line obtained by crossing Slc22a3-FF and Aldh1l1-GFP mice. For control experiments, we injected Gfap-mCh, and for knockout experiments, we injected Gfap-Cre-RFP viral vectors; as before, mice were injected at P60 and harvested at P90 (Fig. 3A). Quantification of Slc22a3 knockout efficiency across all Aldh1l1-GFP OB astrocytes revealed a significant 58.34% reduction in Slc22a3 expression in astrocytes (Fig. 3, B and C, and fig. S12, A to D) but not in neurons (Fig. 3, D and E). Finally, we did not observe any changes in the overall numbers of astrocytes and neurons after Slc22a3 deletion in the OB (fig. S12, E and F).

To determine how the loss of Slc22a3 in OB astrocytes affects olfactory behaviors, we used Slc22a3-cKO and control mice (Fig. 3A). We performed behavioral assays on these mice to evaluate potential alterations in odor behaviors (fig. S13A). In a three-chamber place-preference assay, mice were exposed to a novel odorant (*R*)-limonene in a series of increasing concentrations (fig. S13B). Odor detection was calculated based on the time spent investigating the odorant-containing chamber compared with the mineral oil–containing chamber and is represented as zone preference index. Slc22a3-cKO mice required a higher concentration of odorant to exhibit a preference for (*R*)-limonene compared with control mice, thus demonstrating a significant reduction in olfactory detection (Fig. 3, F and G). Next, we evaluated odor discrimination using a similar paradigm in which we compared the preference for structurally similar odorants (fig. S13B). Whereas control mice spent more time in the chamber containing the new odorant, thus demonstrating odor preference, Slc22a3-cKO mice showed no significant difference in preference, suggesting that Slc22a3-cKO mice are deficient in their ability to discriminate between chemically similar, but distinct, odorants (Fig. 3H). Given the observed defects in olfactory behaviors, we examined whether the electrophysiological properties of OB neurons are affected by astrocytic knockout of Slc22a3. We performed whole-cell recordings from granule cells in the OB and found that the excitability of granular cell layer neurons was unaffected (Fig. 3, I and J). Examination of their synaptic properties revealed a significant 52.78% decrease in spontaneous excitatory postsynaptic current (sEPSC) frequency (Fig. 3, K and L), coupled with a 17.78% increase in spontaneous inhibitory postsynaptic current (sIPSC) amplitude (Fig. 3, M and N) in Slc22a3-cKO mice.

Slc22a3 regulates astrocytic calcium responses to neuromodulators

The foregoing results indicate that the loss of astrocytic Slc22a3 disrupts olfactory function, which led us to examine how its loss affects core features of astrocytes in the OB. We performed RNA-seq on FACS-isolated astrocytes from control and Slc22a3-cKO OBs to identify the molecular features that affect core astrocyte properties after loss of Slc22a3 (Fig. 4A). Slc22a3 deletion (fig. S13C) affected astrocyte transcriptomics: 565 up-regulated and 1049 down-regulated genes were differentially expressed in Slc22a3-cKO red fluorescent protein (RFP) astrocytes in comparison to control mCh-injected control astrocytes (Fig. 4B and table S4). GO analysis of these DEGs revealed synaptic transmission as the most-affected category, in addition to cell-cell adhesion, interferon response, neurotransmitter transport, potassium ion transport, and calcium ion binding. Constituent genes of all these categories were mostly down-regulated. Only interferon response genes were up-regulated (Fig. 4, C and D). Given that synaptic transmission is affected by astrocyte-neuron communication, which is reliant upon their proximity, and facilitated by complex and elaborate astrocyte morphologies, we evaluated how the loss of Slc22a3 affects astrocyte morphology. Furthermore, Slc22a3-cKO astrocytes demonstrate down-regulation of several morphology-related genes (Fig. 4E). We performed high-resolution confocal imaging and three-dimensional reconstruction of OB astrocytes using GFP reporter in control (*Aldh1l1-GFP; Slc22a3-FF* + pAAV-Gfap-mCh) and Slc22a3-cKO (*Aldh1l1-GFP; Slc22a3-FF* + pAAV-Gfap-Cre-RFP) OBs (Fig. 4F). Sholl analysis of astrocytes revealed decreased morphological complexity in Slc22a3-cKO astrocytes in comparison to control astrocytes (Fig. 4G) based on a host of parameters, including the number of process intersections as a function of distance from the soma (Fig. 4H) (82.96% reduction), total process length (48.25% reduction), number of branches (21.23% reduction), and number of terminal points (54.24% reduction) (Fig. 4I).

Astrocyte calcium activity is considered a proxy for the physiological activities of astrocytes and has been linked to astrocyte regulation of neuronal function. Furthermore, our RNA-seq studies identified several genes linked to calcium activity that demonstrate significant down-regulation in Slc22a3-cKO astrocytes (Fig. 4E). To assess astrocytic calcium, we used pAAV-containing Gfap promoter–driven GCaMP6 that encodes a fluorescent calcium optical sensor, which we injected into OBs of control (*Slc22a3-FF*) and Slc22a3-cKO (*Slc22a3-FF* + pAAV-Gfap-Cre) mice (Fig. 4J). We generated OB slices from these mice and performed two-photon imaging to assess calcium dynamics under basal conditions from astrocytes located in the granular cell layer of the OB. We detected no change in soma spontaneous calcium amplitude and frequency between control and Slc22a3-cKO OB astrocytes (fig. S14, A to C).

Astrocyte calcium activity is modified in response to neurotransmitters and neuromodulators, which provides an indirect measure of their interactions with neurons (*16*). We observed a significant down-regulation of the ionotropic glutamate receptor *Grin2a* and the metabotropic serotonin receptor *Htr2c* in Slc22a3-cKO OB astrocytes (Fig. 4E). We therefore assessed calcium activity in astrocytes from OB slices after application of glutamate and serotonin. We bath-applied serotonin or glutamate onto OB slices while monitoring fluctuations in calcium elevations in astrocytes [with tetrodotoxin (TTX) in the solution]. Furthermore, because Slc22a3 has been implicated in the maintenance of extracellular serotonin levels in the brain (*38*), we were especially interested in serotonin-induced calcium dynamics. We performed two-photon imaging in control and Slc22a3-cKO mice while TTX (0.5 μM) was applied to block neuronal activity for 200 s before application of glutamate (300 μM) or serotonin (50 μM) for an additional 300 s. Quantification of fluorescence from astrocytic soma revealed a significant 66.56% reduction in serotonin-induced amplitude and 72.27% reduction in the frequency of calcium events in Slc22a3-cKO OB astrocytes compared with control astrocytes (Fig. 4, K to N). By contrast, we observed no significant difference between the amplitude of glutamate-induced calcium activity in Slc22a3-cKO astrocytes compared

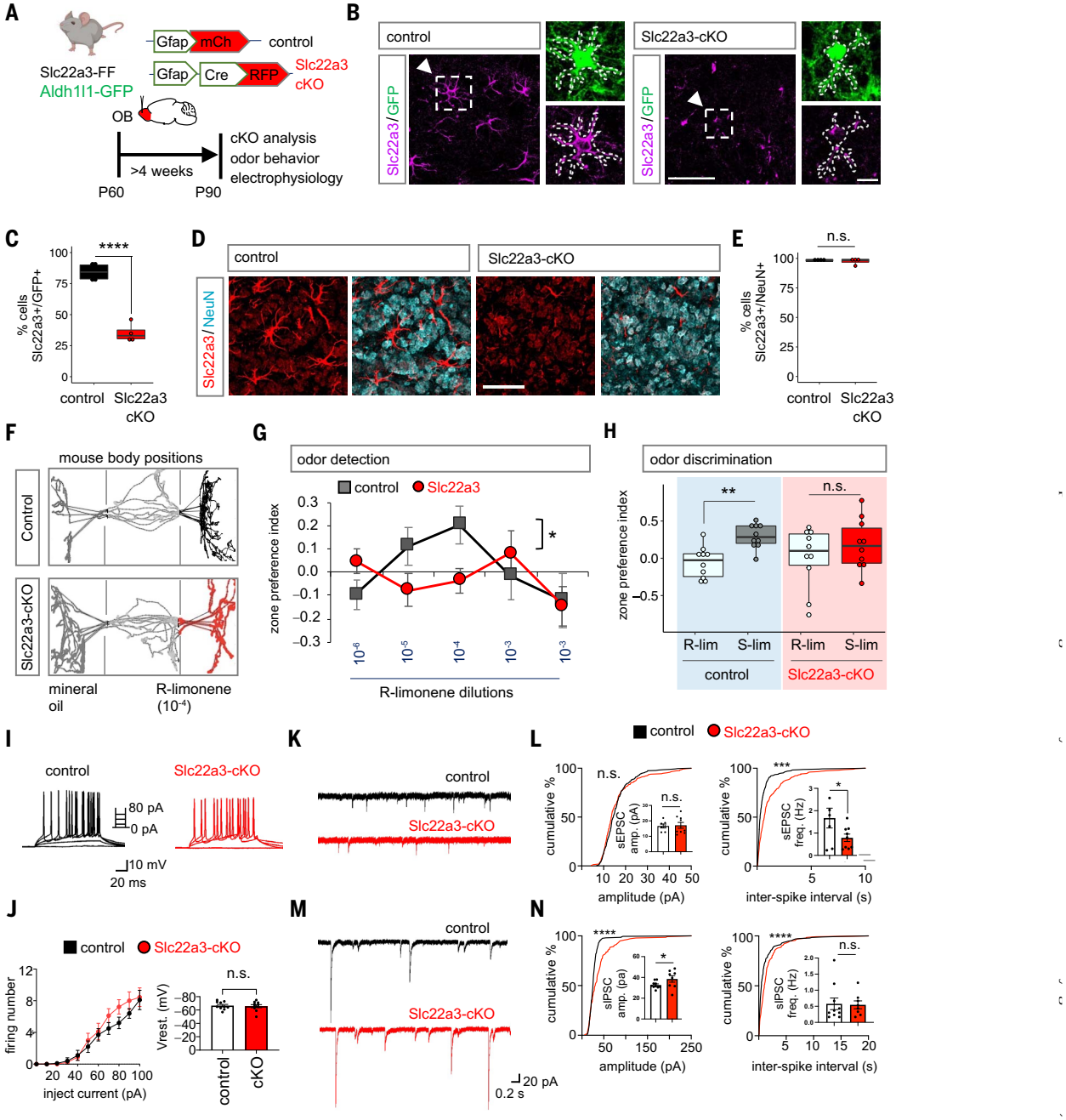


Fig. 3. Astrocytic Slc22a3 regulates olfactory circuit function. (**A**) Schematic illustrating viral vectors used for Slc22a3 conditional deletion from OB astrocytes. (**B** to **E**) Immunostaining and quantification of Slc22a3 in GFP$^+$ astrocytes [(B) and (C)] (average of 83 to 93 cells per cohort; ****$p = 9.59 \times 10^{-5}$, unpaired Student's two-tailed t test on $n = 4$ mice per cohort) and in NeuN$^+$ neurons [(D) and (E)] [average of 350 to 363 cells per cohort; $p = 0.5828$ (n.s., not significant), unpaired Student's two-tailed t test on $n = 4$ mice per cohort]. Scale bar, 50 μm; inset scale bar, 10 μm. In (B), white dashed lines represent zoomed-in astrocytes and their boundaries. (**F**) Schematic illustrating live mouse tracking in the three-chamber assay for odor detection. The top panel represents a control mouse exploring (*R*)-limonene at a dilution of 10^{-4}, whereas the bottom panel represents a Slc22a3-cKO mouse showing no preference for the same (*R*)-limonene concentration. (**G**) Quantification of odor detection in control and Slc22a3-cKO mice [$n = 10$ per cohort; *$p = 0.0211$, two-way repeated measures analysis of variance (ANOVA) with Sidak multiple comparison]. (**H**) Quantification of odor discrimination between (*R*)-limonene (R-lim) and (*S*)-limonene (S-lim) from the same cohorts of mice ($n = 10$ per cohort; **$p = 0.0052$, two-way repeated measures ANOVA with Sidak multiple comparisons). (**I** and **J**) Whole-cell patch clamp electrophysiology of granule cell firing number from stepped current injections ($n = 3$, 10 cells per cohort; $p = 0.0578$, two-way ANOVA with Sidak's multiple comparison correction). Data are presented as mean ± SEM. Vrest., resting membrane potential. (**K** and **L**) Traces and summary data of amplitude and frequency from sEPSC recordings (7 to 9 cells per cohort; sEPSC amplitude $p = 0.7546$ and sEPSC frequency *$p = 0.0473$, unpaired Student's two-tailed t test on $n = 3$ mice per cohort; ****$p < 0.0001$), Kolmogorov-Smirnov test. (**M** and **N**) Traces and quantification of amplitude and frequency from sIPSC recordings (8 to 10 cells per cohort; sIPSC amplitude *$p = 0.0164$ and sIPSC frequency $p = 0.8095$, unpaired Student's two-tailed t test on $n = 3$ mice per cohort; ****$p <$ 0.0001 Kolmogorov-Smirnov test). All recordings in (L) and (N) are in granule cells, and data are presented as mean ± SEM. See table S3 for data summary. For the box plots in (C), (E), and (H), the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. Illustrations were created with Biorender.com.

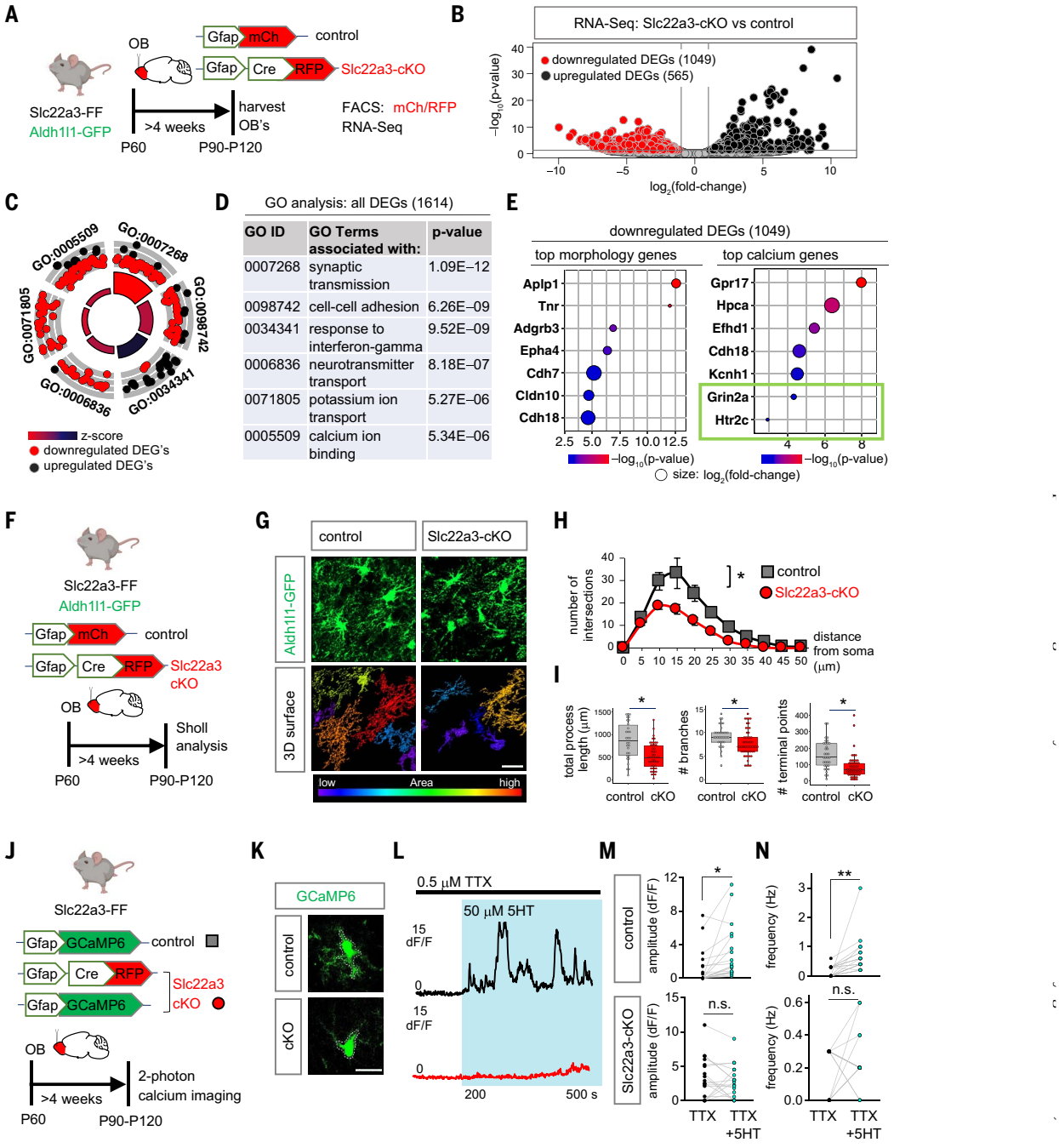


Fig. 4. Astrocytic Slc22a3 regulates astrocyte morphology and calcium activity. (**A**) Schematic illustrating viral vectors and timelines for the RNA-seq experiment. (**B**) Volcano plots depicting RNA-Seq from FACS-isolated astrocytes comparing Slc22a3-cKO samples with control samples. (**C** and **D**) GO circle plot (C) and table (D), showing the top GO terms found in the DEGs shown in (B). (**E**) Top morphology and calcium-associated genes ($p < 0.01$) in down-regulated DEGs in the Slc22a3-cKO OB astrocytes. (**F**) Schematic illustrating viral vectors and timelines for the evaluation of astrocyte morphology. (**G**) High-magnification confocal images of Aldh1l1-GFP from control and Slc22a3-cKO mice and 3D surface rendering of the same showing reduced astrocyte morphological complexity in Slc22a3-cKO OB astrocytes. Scale bar, 20 μm. (**H**) Sholl analysis of astrocyte complexity ($n = 4$, average of 34 to 54 cells per cohort; $^*p = 0.0117$, two-way repeated measures ANOVA with Sidak correction). Data are presented as mean ± SEM. (**I**) Quantification of total process length ($^*p = 0.0285$, Wilcoxon rank sum test), branch number ($^*p = 0.0198$, unpaired Student's two-tailed t test), and terminal points ($^*p = 0.0177$, unpaired Student's two-tailed t test) in control and Slc22a3-cKO OB astrocytes (34 to 54 cells per cohort, statistics on $n = 4$ mice per cohort). In the box plots, the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. (**J**) Schematic illustrating mice, viral vectors, and timelines for the expression of the optical calcium sensor in OB astrocytes. (**K** and **L**) Traces from two-photon, slice imaging of GCaMP6 activity from OB astrocyte soma in control and Slc22a3-cKO slices in the presence of TTX (0.5 μM) and serotonin (5HT, 50 μM). Scale bar, 10 μm. dF/F, change in fluorescence intensity relative to resting fluorescence intensity. (**M** and **N**) Quantification of amplitude and frequency from serotonin-induced calcium activity from astrocyte soma (19 or 20 cells per cohort; control amplitude $^*p = 0.0346$, Slc22a3-cKO amplitude $p = 0.6992$, control frequency $^{**}p = 0.0029$, and Slc22a3-cKO frequency $p = 0.5588$, paired Student's two-tailed t test on $n = 4$ mice per cohort; n.s., not significant). See table S3 for data summary. Illustrations were created with Biorender.com.

with control astrocytes, whereas there was a small reduction in frequency in Slc22a3-cKO astrocytes (fig. S14, D to F). Analysis of astrocyte microdomain calcium activity in response to serotonin or glutamate revealed reduced amplitude and frequency of calcium events for both in Slc22a3-cKO astrocytes (fig. S15, A to D).

Slc22a3 regulates histone serotonylation in astrocytes

We next examined intracellular serotonin levels within astrocytes in the OB (fig. S16A). Immunostaining for serotonin revealed a decrease in its intensity within Slc22a3-cKO astrocytes (fig. S16, B to D), suggesting a defect in serotonin transport. In neurons, serotonin can be added to histones, and the serotonylation of histones directly potentiates epigenetic mechanisms of gene regulation (*31*). Although both astrocyte morphology and calcium have been linked to a host of behavioral outcomes (*39*), the role of astrocytic histone modifications in the regulation of circuits and behavior remains unknown. Furthermore, Slc22a3 is a serotonin transporter in the CNS, and the role of serotonin in OB astrocytes remains relatively unexplored. We first asked whether the observed reduction in intracellular serotonin in Slc22a3-cKO astrocytes affects histone serotonylation. Using the Aldh1l1-GFP reporter mouse, we performed immunostaining with antibodies specific for the serotonylated histone mark on histone H3—H3K4me3Q5ser (H3-5HT)—and found that 86% of GFP^+ OB astrocytes contain H3-5HT (Fig. 5A and fig. S17A). Next, we immunostained for H3-5HT in Slc22a3-cKO (*Aldh1l1-GFP; Slc22a3-FF* + pAAV-Gfap-Cre) and control (*Aldh1l1-GFP; Slc22a3-FF* + pAAV-Gfap-mCh) OB astrocytes (Fig. 5B). We found an 18% reduction in the levels of H3-5HT in Slc22a3-cKO OB astrocytes (Fig. 5, C and D) but not in neurons of Slc22a3-cKO OBs (fig. S17B). We next performed ChIP-seq of H3-5HT on control and Slc22a3-cKO OBs to determine whether patterns of H3-5HT epigenetic modification are affected after astrocytic Slc22a3 loss and observed a 2.3-fold reduction in H3-5HT peaks in Slc22a3-cKO OBs (Fig. 5E).

Prior studies demonstrating that histone serotonylation activates gene expression (*31*) led us to cross-compare our ChIP-seq results with the RNA-seq data obtained from Slc22a3-cKO astrocytes. Of the 1049 down-regulated DEGs in the Slc22a3-cKO astrocytes, 538 genes (51% of the DEGs) also lose histone serotonylation modifications in the Slc22a3-cKO OBs (Fig. 5, F and G). In comparison, of the 565 up-regulated DEGs in Slc22a3-cKO astrocytes, only 26 genes (4.6% of the DEGs) acquire H3-5HT modifications in the Slc22a3-cKO OBs (fig. S17C). We next interrogated GO and pathways that exhibit differential H3-5HT modification in the OBs of Slc22a3-cKO mice. GO analysis between control and Slc22a3-KO H3-5HT peaks revealed that a prevalence of GO terms linked to GABAergic signaling are lost after astrocytic Slc22a3 deletion (Fig. 5H). Among these genes are various receptor subtypes for GABA and GABA biosynthetic enzymes *Maob* and *Aldh1a1* (fig. S17D). Further analysis of the GO terms from the 538 genes that lose H3-5HT and are down-regulated in Slc22a3-cKO astrocytes revealed categories of synaptic transmission, morphology, cell-cell adhesion, and calcium-ion binding, GO terms that were also observed in the Slc22a3-cKO RNA-seq results (Fig. 4D). GABA signaling pathway GO categories were also maintained in this smaller subset of filtered astrocytic genes (Fig. 5I).

Because the GABA pathway represents one of the top GO terms demonstrating a loss of histone serotonylation modifications in the Slc22a3-cKO OBs (Fig. 5H), we focused on the GABA synthesis pathway. Assessing astrocyte-specific expression of GABA and protein expression of *Maob* and *Aldh1a1* using localization with GFP in control and Slc22a3-cKO OBs, we found that both monoamine oxidase B (MAOB) and astrocytic GABA were significantly down-regulated (31.5 and 30.2%, respectively) in OB astrocytes from Slc22a3-cKO mice, whereas Aldh1a1 was unaffected (Fig. 6, A to D, and fig. S18, A and B). A similar analysis evaluating GABA and MAOB levels in neurons with NeuN colabeling revealed no significant differences (Fig. 6, E to H). However, we observed an overall reduction in mean intensity of GABA in whole sections but not in mean intensity of MAOB (fig. S18, C and D). Given that astrocytes are capable of synthesizing and releasing GABA (*25*, *40*) and our results suggest a general decrease in astrocytic GABA in Slc22a3-cKO astrocytes, we used slice electrophysiology recordings to measure tonic GABA release from OB astrocytes of control and Slc22a3-cKO mice (Fig. 6I). We found a 86.71% decrease in tonic GABA current from granular cell layer neurons of the OBs of Slc22a3-cKO mice. To exclude the possibility that the tonic GABA current is derived from neurons, we performed control recordings in the presence of TTX (0.5 μM) to dampen neuronal activity. This revealed a 75.24% decrease in tonic current in Slc22a3-cKO OBs, confirming that the observed reduction in tonic GABA current is driven by astrocytes and not by neurons (Fig. 6, J to L). Finally, to determine whether this reduction in tonic GABA currents was due to a decrease in high-affinity neuronal extrasynaptic GABA-A receptor function or expression, we recorded tonic GABA currents in the presence of 5 μM GABA, a saturating concentration for extrasynaptic GABA-A receptors. We identified no differences between control and Slc22a3-cKO groups (Fig. 6, M and N), further indicating that the observed reduction in tonic GABA currents (Fig. 6, J to L) results from a decrease in GABA release from astrocytes.

Inhibition of H3-5HT in OB astrocytes disrupts sensory processing

To examine whether H3-5HT modifications in OB astrocytes directly contribute to olfactory sensory processing, we used a dominant-negative mutant H3.3Q5A, where the histone glutamine residue that is modified by serotonin is mutated to an alanine to attenuate H3-5HT (*31*). The use of histone variant 3.3 (H3.3) ensures histone incorporation to replace canonical H3 during histone protein turnover (*41*). We generated pAAV containing control H3.3 and mutant H3.3Q5A under control of the Gfap promoter and introduced these into the OBs of wild-type mice (Fig. 7A). Astrocyte-specific expression of H3.3 and H3.3Q5A was confirmed by colocalization of the GFP reporter on H3.3 constructs with astrocyte marker Sox9 but not neuronal marker NeuN (fig. S19, A and B). We observed no differences in cell numbers of GFP^+ astrocytes and $NeuN^+$ neurons in OBs expressing these H3.3 constructs (fig. S19, C and D). Quantification of H3-5HT revealed a significant 46.3% reduction of H3-5HT in H3.3Q5A-expressing OB astrocytes compared with H3.3-expressing control astrocytes (Fig. 7, B and C) but no reduction in surrounding neurons (fig. S19E). We next assessed astrocyte morphology and neuronal electrophysiology. High-resolution confocal imaging and Sholl analysis revealed decreased morphological complexity in H3.3Q5A-expressing astrocytes compared with H3.3-expressing control astrocytes (Fig. 7D), as evaluated by the number of process intersections as a function of distance from the soma (59.5% reduction) (Fig. 7E), total process length (58.73% reduction), number of branches (26.44% reduction), and number of terminal points (63.03% reduction) (Fig. 7F). Astrocytes exhibit extensive structural plasticity in perisynaptic astrocyte processes (PAPs) (*42*, *43*). We therefore measured the number of astrocyte terminal process points and determined the distance between PAPs labeled by ezrin (Ezr) and postsynaptic marker PSD95. There was a significant reduction in the number of PSD95-expressing puncta localized to ezrin-expressing terminal astrocytic processes in H3.3Q5A-expressing astrocytes compared with control astrocytes (fig. S20, A to F), suggesting reduced interactions between astrocyte processes and neuronal synapses.

An examination of the basal properties of EPSCs (Fig. 7G) and IPSCs (Fig. 7H) revealed no changes in sEPSC or sIPSC frequency and amplitudes (Fig. 7, I and J). Because Slc22a3-cKO OBs showed a significant reduction in H3-5HT epigenomic regulation of GABA pathways, we assessed the expression of MAOB and GABA. We found a significant reduction in the

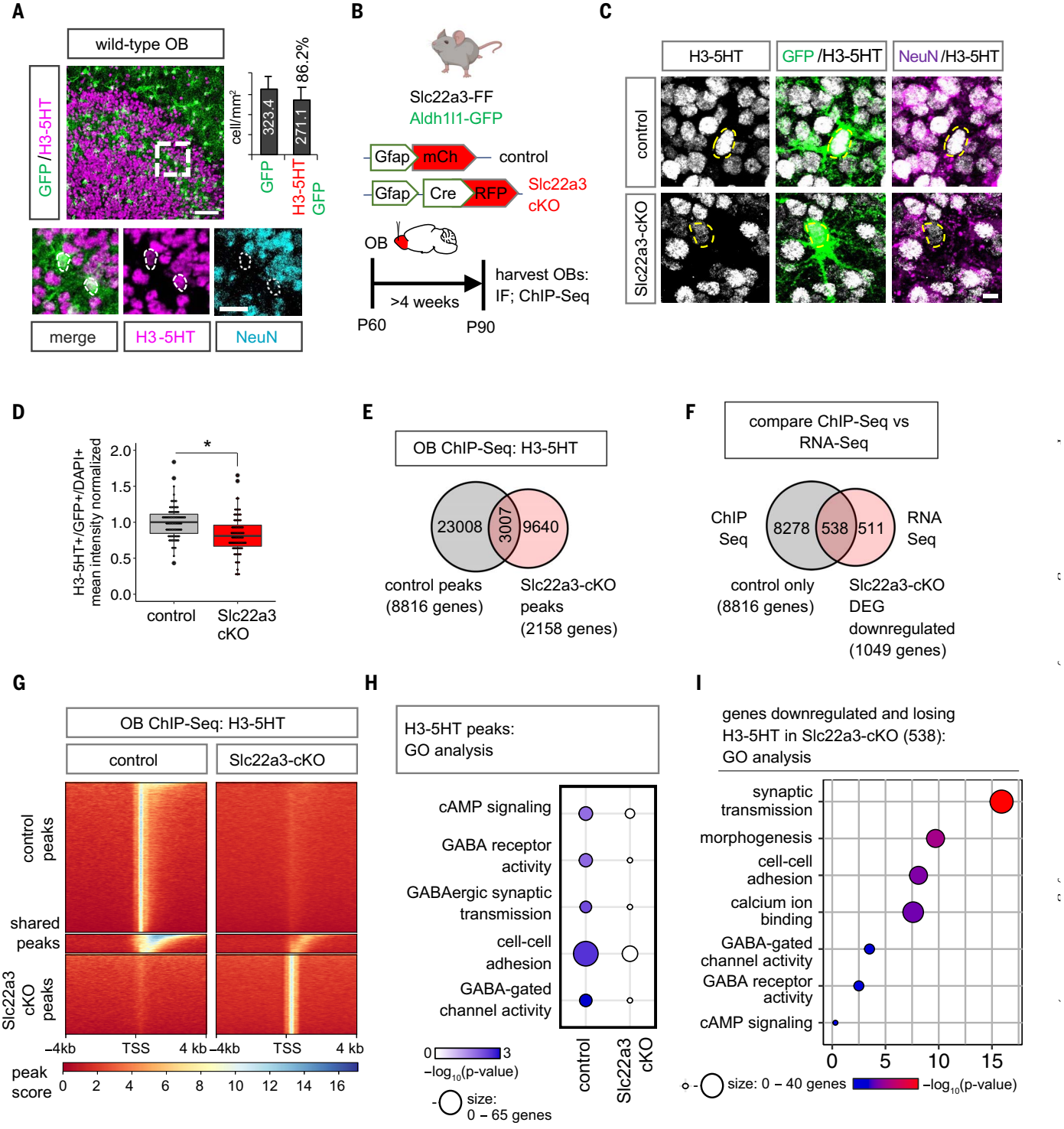


Fig. 5. Slc22a3 regulates histone serotonylation in OB astrocytes. (**A**) Immunostaining of H3-serotonin in OBs of Aldh1l1-GFP mice and quantification of GFP$^+$H3-5HT$^+$ colabeling (n = 3, 25 to 45 cells). White dashed lines in the zoom-in views represent the boundary of the astrocyte nucleus. Error bars represent mean ± SEM. Scale bar, 50 μm; zoom-in scale bar, 20 μm. (**B**) Schematic illustrating viral vectors and timelines for H3-5HT quantification and ChIP-seq. IF, immunofluorescence. (**C** and **D**) H3-5HT immunostaining (C) and quantification (D) in control and Slc22a3-cKO OB astrocytes (74 to 79 cells per cohort; *p = 0.0377, unpaired Student's two-tailed t test on n = 4 mice per cohort). Yellow dashed lines represent the boundary of the astrocyte nucleus. Scale bar, 5 μm. In the box plot, the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. DAPI, 4′,6-diamidino-2-phenylindole. (**E**) Venn diagram depicting the number of H3-5HT ChIP-seq peaks that are specific to or shared between control and Slc22a3-cKO OBs (n = 4 OBs per cohort). (**F**) Venn diagram depicting number of genes that both lose H3-5HT peaks and are down-regulated in Slc22a3-cKO astrocytes. (**G**) Heatmaps comparing ChIP H3-5HT at 4 kb from the peak center in control versus Slc22a3-cKO OBs. TSS, transcription start site. (**H**) GO analysis of genes at H3-5HT peaks revealing a loss of H3-5HT regulation at GABA-associated pathways in Slc22a3-cKO OBs. (**I**) GO analysis of the 538 overlapping genes shown in (F). See table S3 for data summary. Illustrations were created with Biorender.com.

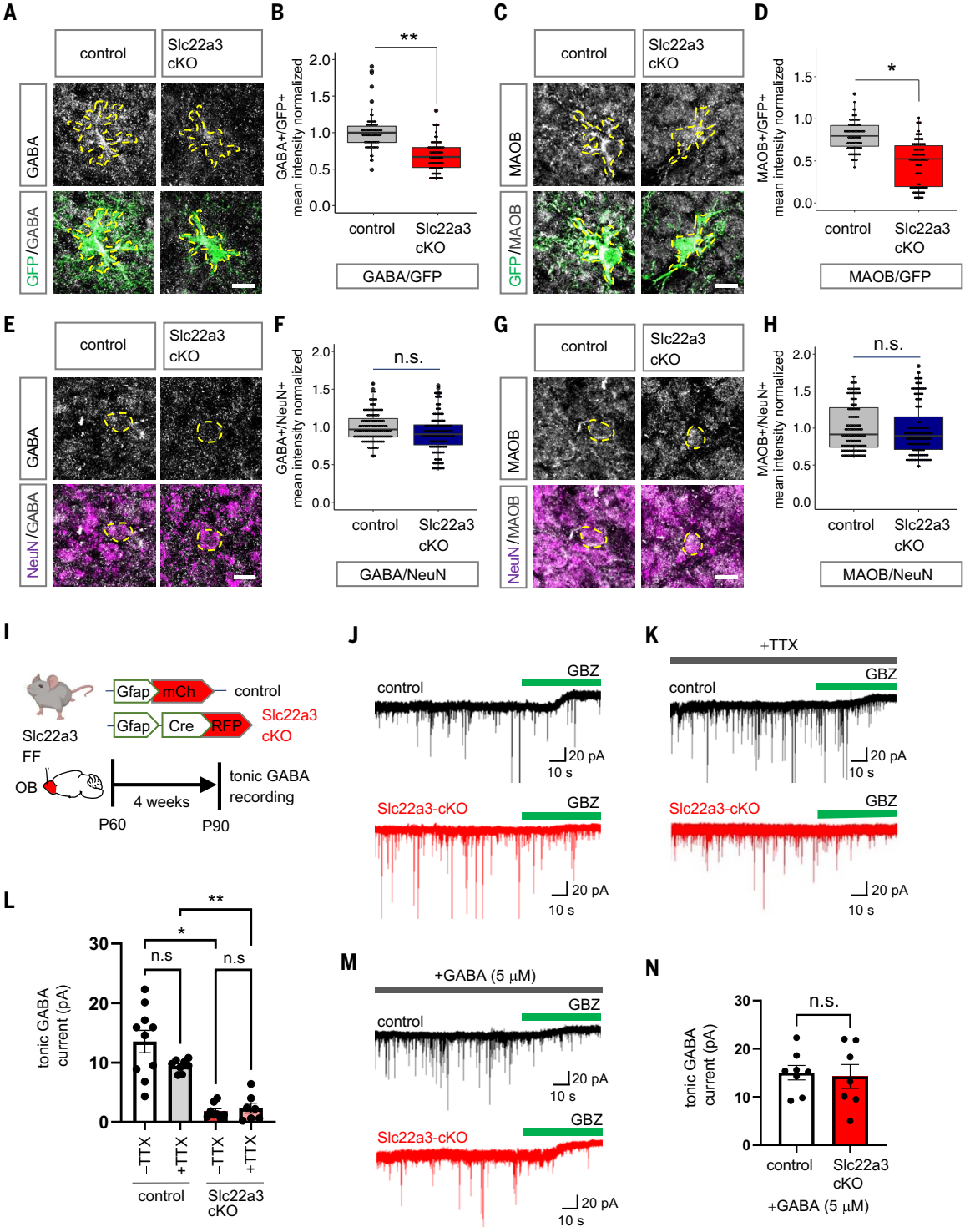


Fig. 6. Slc22a3 regulates tonic GABA release from OB astrocytes. (**A** to **H**) Immunostaining and quantification of GABA and MAOB in control versus Slc22a3-cKO GFP$^+$ astrocytes [(A) to (D)] (73 to 88 cells per cohort; GABA $^{**}p$ = 0.0032, MAOB $^{*}p$ = 0.0121) and NeuN$^+$ neurons [(E) to (H)] (144 cells per cohort; GABA p = 0.2119, MAOB p = 0.9355) in the OB (unpaired Student's two-tailed t test on n = 4 mice per cohort). In the box plots, the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. In the immunostaining images, yellow dashed lines represent the astrocyte boundary. Scale bars, 10 μm. (**I**) Schematic illustrating viral vectors and timelines for the tonic GABA current measurement experiments. (**J** and **K**) Traces of tonic GABA currents in granule cells in OBs from control and Slc22a3-cKO mice treated with gabazine (GBZ; 20 μM) and with (K) or without (J) treatment with TTX (0.5 μM). (**L**) Quantification of tonic GABA current (7 to 10 cells per cohort; without TTX $^{*}p$ = 0.0331 and with TTX $^{**}p$ = 0.0056, unpaired Student's two-tailed t test on n = 3 mice per cohort). (**M** and **N**) Traces (M) and quantification of measurement (N) of tonic GABA current in the presence of GABA (7 or 8 cells per cohort; p = 0.8483, unpaired Student's two-tailed t test on n = 3 mice per cohort). Data in (L) and (N) are presented as mean ± SEM. See table S3 for data summary. Illustrations were created with Biorender.com.

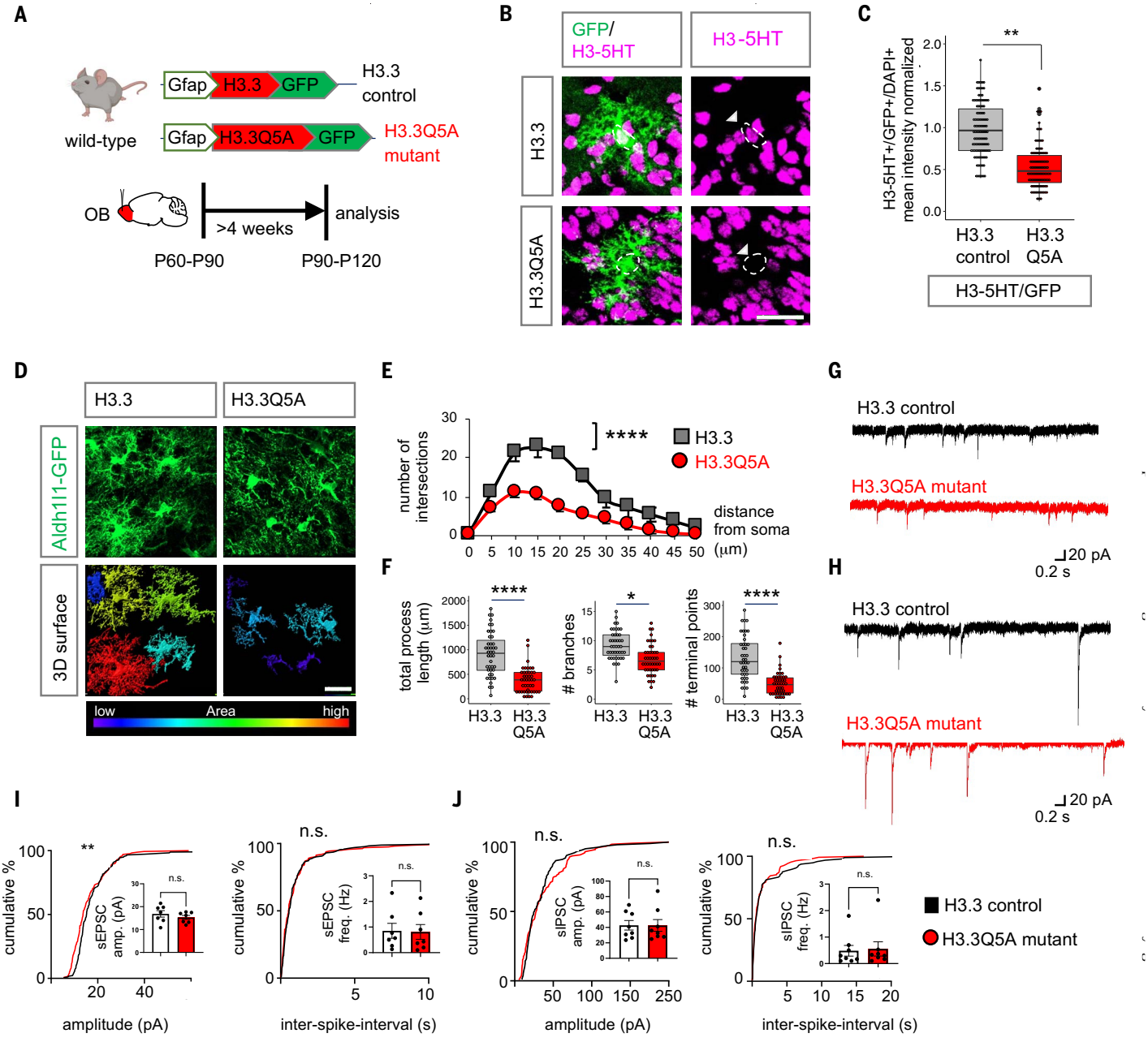


Fig. 7. Inhibition of H3-5HT in OB astrocytes disrupts astrocyte morphology. (**A**) Schematic illustrating the viral vectors used for H3.3 and H3.3Q5A expression in OB astrocytes. (**B**) H3-5HT and GFP colabeling in H3.3- and H3.3Q5A-expressing OBs (white dashed lines represent the boundary of the astrocyte nucleus; scale bar, 25 μm) and (**C**) a box plot depicting quantification of astrocytic H3-5HT (99 to 110 cells per cohort; $^{**}p = 0.0092$, unpaired Student's two-tailed t test on $n = 4$ mice per cohort). (**D**) High-magnification confocal images of H3.3-GFP and 3D surface rendering of the same showing reduced astrocyte morphological complexity in H3.3Q5A OB astrocytes. Scale bar, 20 μm. (**E**) Sholl analysis of astrocyte complexity ($n = 4$, average of 44 cells per cohort; $^{****}p = 1.9 \times 10^{-5}$, two-way repeated measures ANOVA with Sidak correction). Data are presented as mean ± SEM. (**F**) Quantification of total process length, branch number, and terminal points [44 cells per cohort; $^{****}p = 2.75 \times 10^{-5}$ (left), $^{*}p = 0.0122$ (middle), and $^{****}p = 8.73 \times 10^{-6}$ (right), unpaired Student's two-tailed t test on $n = 4$ mice per cohort]. (**G** and **H**) Traces and (**I** and **J**) summary data of amplitude and frequency from sEPSC recordings [(G) and (I)] ($n = 7$ cells per cohort; sEPSC amplitude $p = 0.2386$ and sEPSC frequency $p = 0.7917$, unpaired two-tailed Student's t test on $n = 3$ mice per cohort; $^{**}p = 0.0059$ Kolmogorov-Smirnov test) and from sIPSC recordings [(H) and (J)] ($n = 8$ cells per cohort; sIPSC amplitude $p = 0.8277$ and sIPSC frequency $p = 0.7128$, unpaired two-tailed Student's t test on $n = 3$ mice per cohort). All recordings are in granule cells from H3.3 and H3.3Q5A OBs, and data are presented as mean ± SEM. See table S3 for data summary. For the box plots in (C) and (F), the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. Illustrations were created with Biorender.com.

expression of both (MAOB: 17.3%; GABA: 23.5%) in OB astrocytes expressing H3.3Q5A compared with H3.3-expressing control OB astrocytes (Fig. 8, A to C), but we found no reduction in neurons (fig. S21, A and B). We also observed an overall reduction in GABA mean intensity in whole fields (fig. S21, C and D). We performed control experiments to confirm that the expression of Gfap-H3.3 alone does not alter levels of Slc22a3, GABA, and MAOB in OBs of astrocytic H3.3–expressing mice compared with Aldh1l1-GFP mice (fig. S21, E

to G). We used electrophysiology to measure GABA release from OB astrocytes expressing H3.3Q5A (mutant) or H3.3 (control) and observed a 92.97% reduction in tonic GABA current in the H3.3Q5A-expressing condition (Fig. 8, D to F). Control recordings were performed in the presence of TTX (0.5 μM), which revealed a 86.09% decrease in tonic current in H3.3Q5A-expressing OBs consistent with alterations in astrocyte-derived tonic GABA currents. Control recordings in the presence of saturating GABA (5 μM) further corroborated reduced tonic GABA currents due to a reduction of tonic GABA release from OB astrocytes expressing H3.3Q5A (fig. S22, A and B).

Given the links between GABA signaling and neuronal circuit function in the OB, we examined how inhibition of H3-5HT affects olfactory sensory processing. We performed odor detection and odor discrimination behavioral assays as previously described (Fig. 3). First, we compared odor behaviors of mice expressing H3.3 in OB astrocytes with those of wild-type control mice and found that astrocytic H3.3 expression alone has no significant impact on odor detection limit and odor discrimination ability (fig. S22, C and D). We then compared odor behaviors of H3.3-expressing (control) and H3.3Q5A-expressing (mutant) mice and found reduced odor detection and impaired odor discrimination in mice that expressed H3.3Q5A in OB astrocytes (Fig. 8, G and H). Mice that expressed astrocytic H3.3Q5A recapitulated the phenotypes of impaired tonic GABA and olfactory behaviors that were observed in the Slc22a3-cKO mice (Figs. 3, 4, and 6). Collectively, these findings suggest that the neuronal activity–dependent target Slc22a3 in OB astrocytes facilitates serotonin-mediated epigenetic regulation of GABA synthesis, which contributes to sensory processing of olfaction (Fig. 8I).

Discussion

Neuronal circuits and associated behaviors are intimately linked to astrocytes, yet how neuronal activation sculpts astrocyte transcriptional responses to support circuit activities is unclear. We identified neuronal activity–dependent transcriptional changes in astrocytes, highlighted by alterations in the DNA binding capacity of the TF Sox9 and induction of prospective immediate early gene *Slc22a3*, which encodes the neuromodulator transporter Slc22a3. Further examination of Slc22a3 in astrocytes revealed that it is required to maintain olfactory circuit function. Our study pinpointed Sox-family TFs as key components of activity-dependent responses in astrocytes in the OB. However, given that astrocyte function is regulated by region-specific transcriptional dependencies (*13*, *14*, *44*), there are likely region-specific astrocyte TFs that mediate these responses and responses associated with injury and disease (*45*). It is also likely that different forms of neuronal input (excitatory, inhibitory, and so on) or behavioral states influence these transcriptional responses in astrocytes. These findings illustrate that astrocytes exhibit a form of plasticity in response to neuronal activity and that this plasticity is reliant upon a combination of inputs from neuronal circuits and region-specific TFs. These observations also raise the question of how this form of transcriptional plasticity in astrocytes is regulated. The increase in Sox9-DNA binding in the presence of heightened neuronal activity is likely the result of activity-dependent protein interactions. Our mass spectrometry studies identified a cohort of 40 proteins that interact with Sox9 in the OB specifically after neuronal activation. These findings indicate that neuronal input influences the protein constituency of transcriptional complexes and suggests that these inputs shape protein interactions that drive gene expression. Understanding how neuronal input remodels transcriptional complexes and the specific roles of activity-dependent protein interactions in astrocyte-neuron communication are important areas of future investigation. From these findings, a model emerges in which neuronal activity orchestrates transcriptional responses in astrocytes to meet the demands of a functioning circuit.

We identified *Slc22a3* as a prospective immediate early gene that is specifically induced in OB astrocytes after exposure to odor, and analysis of OB astrocytes from Slc22a3-cKO mice revealed a host of phenotypes, including reduced morphological complexity, reduced calcium activity, decreased tonic GABA release, altered histone serotonylation, and impaired olfactory detection. These observations suggest that in addition to an acute function after odor exposure or neuronal activity, Slc22a3 also plays a chronic, longer-term role in astrocyte function and communication with neurons. The chronic role is likely due to epigenomic changes that result in increased MAOB expression, increased GABA production, and altered morphology. This chronic role is likely to reflect a homeostatic function for Slc22a3 and/or serotonin transport in maintaining the physiological activities of OB astrocytes. It is also possible that these two roles are interdependent, whereby the changes in GABA release from astrocytes influences neuronal activity in a way that affects astrocyte morphology. Alternatively, we must consider the possibility that shorter astrocytic processes may lead to a longer diffusion time for GABA to reach synaptic GABA receptors, ultimately leading to a reduced tonic GABA current. Another possibility is that there is overall less tonic GABA release from astrocytes, as suggested by our data of reduced astrocytic GABA biosynthetic enzyme MAOB and astrocytic GABA levels. Slc22a3 functions as a monamine transporter that regulates the transport of serotonin (*38*). Because the OB is densely innervated by serotonergic fibers, which activate interneuron granule cells (*46*), it raises the question of whether serotonin is a factor that modulates inhibitory outputs and the subsequent impact on odor discrimination. Our finding that odor-evoked neuronal activation leads to enrichment of cAMP-signaling GO terms in astrocytes, suggests that neuronal activation leads to astrocytic activation of G_s or G_i G protein–coupled receptors (GPCRs), including astrocytic serotonergic receptors. In turn, this leads to up-regulated expression of serotonin transporter Slc22a3 in astrocytes via Sox9. Our data demonstrating that serotonin-induced calcium signaling is reduced in Slc22a3-deficient astrocytes is further supported by our observation of reduced astrocytic serotonergic receptors in Slc22a3-cKO astrocytes. These collective findings indicate that astrocytes use serotonin through Slc22a3 to control tonic GABA inhibition, highlighting a new mechanism of serotonergic modulation of inhibitory outputs through astrocytes, which ultimately affects olfactory sensory processing.

Our observations that Slc22a3-cKO astrocytes display both reduced morphological complexity and reduced calcium activity raises questions regarding the relationship between astrocyte morphology and astrocyte calcium. Because the elaborate processes of astrocytes enable close proximity with synapses, it is likely that astrocyte morphology directly influences calcium. It is also likely that increases in astrocyte calcium can affect gene expression or local protein translation to increase structural complexity to enhance synaptic contact. However, whether astrocyte morphology and calcium directly influence each other is context dependent. For example, a recent study showed that hippocampal astrocytes from socially isolated animals have altered astrocyte calcium activity; however, no changes in astrocyte morphological complexity was observed (*47*). Therefore, additional studies are required to better understand whether there is any direct relationship between morphology and calcium.

Epigenetic mechanisms of gene regulation play a key role in all facets of cell physiology, but how these processes influence roles for astrocytes in circuits remains undefined. We found that serotonin is added to histones in astrocytes and that this modification is used to regulate gene expression and olfactory sensory processing.

Mechanistically, Slc22a3 regulates serotonin levels in astrocytes, which influences the extent of histone serotonylation, reflecting a long-term change in transcriptional activity that

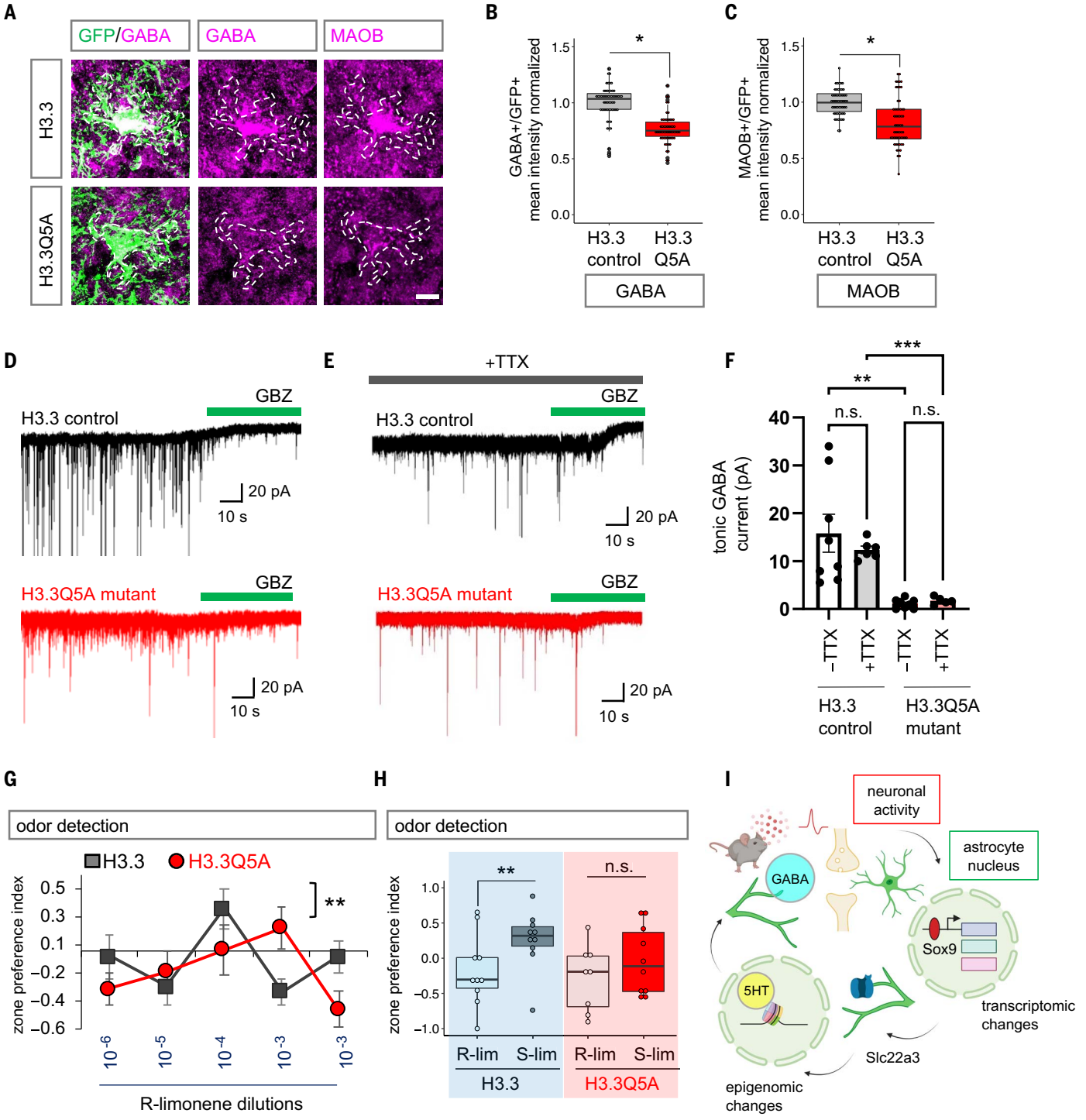


Fig. 8. Inhibition of H3-5HT disrupts astrocytic tonic GABA release and sensory processing. (**A** to **C**) Immunostaining (A) and box plots [(B) and (C)] depicting quantification of astrocytic GABA and MAOB in H3.3 and H3.3Q5A OBs (56 to 58 cells per cohort; GABA *p = 0.0117, unpaired two-tailed Student's t test on n = 4 mice per cohort; MAOB *p = 0.0285, Wilcoxon rank sum test on n = 4 mice per cohort). White dashed lines represent the astrocyte boundary. Scale bar, 10 μm. (**D** and **E**) Traces of tonic GABA currents in granule cells in OBs from H3.3 and H3.3Q5A mice treated with gabazine (GBZ; 20 μM) with (E) or without (D) TTX (0.5 μM). (**F**) Quantification of tonic GABA currents (5 to 8 cells per cohort; without TTX **p = 0.0017 and with TTX ***p = 0.0007, unpaired two-tailed Student's t test on n = 3 mice per cohort). Data are presented as mean ± SEM. (**G**) Quantification of odor detection in H3.3 and H3.3Q5A mice (n = 10 per cohort; **p = 0.0072, two-way repeated measures ANOVA with Sidak multiple comparison). Data are presented as mean ± SEM. (**H**) Quantification of odor discrimination between (*R*)-limonene and (*S*)-limonene from the same cohorts of mice (n = 10 per cohort; **p = 0.0058, two-way repeated measures ANOVA with Sidak multiple comparisons). See table S3 for data summary. (**I**) Model figure integrating activity-dependent transcriptional changes in astrocytes, with Slc22a3 function in olfactory circuits and histone serotonylation regulation of GABA in astrocytes. For the box plots in (B), (C), and (H), the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. Illustrations were created with Biorender.com.

ultimately regulates the expression of GABA-synthesis components. This represents a new mechanism of epigenomic regulation in astrocytes while also highlighting that a host of epigenomic phenomena remain undefined in astrocytes. A recent study demonstrated that Slc22a3 may have roles in noradrenergic signaling at astrocyte nuclear membranes, highlighting another way for Slc22a3 to modulate nuclear processes (*48*). Synaptic changes in neurons affect chromatin accessibility (*30*, *49*, *50*), and it is likely that classic histone acetylation and methylation states are also modified in astrocytes by neuronal activity. Indeed, Sox9 has been implicated in the regulation of epigenomic states in brain tumors (*51*). In olfactory circuits, a few studies have investigated the serotonergic system in the modulation of olfactory processing (*52*, *53*). Our findings reveal a new role for serotonin in astrocytes, wherein it gets deposited in the genome and indirectly regulates astrocytic release of GABA, which demonstrates how astrocytes use serotonin to gate GABA in the OB. Studies in mice and other species have established links between neuromodulators and neurotransmitter signaling (*28*, *29*, *54*, *55*), which is mediated by astrocytes. In this work, we identified an epigenomic intermediatory. It will be critical to decipher the extent to which these neuromodulatory signaling mechanisms in astrocytes directly regulate neurotransmitter signaling or go through an epigenomic intermediatory as well as the nature of this form of transcriptional regulation under different behavioral states.

Materials and methods

Animals

All animals were treated in compliance with the US Department of Health and Human Services, NIH guidelines, and Baylor College of Medicine institutional animal care and use committee (IACUC) guidelines. Mice were housed in a 12-hour light-dark cycle environment with food and water available at all times. Both male and female mice were used for all experiments, and littermates were randomly allocated to experimental groups. All mice used in this study were maintained on the C57BL/6J background. For all ex vivo and in vivo experiments, adult mice aged 2 to 5 months were used. For Sox9 conditional knockout, Sox9 flox/flox (FF) conditional mutant mice were crossed with Cag-CreER mice. Conditional deletion of astrocytic Sox9 has been characterized previously (*13*), and we followed the same protocol for Sox9 deletion (fig. S8A). For Slc22a3 conditional knockout, Slc22a3-FF conditional mutant mice were generated at the Baylor College of Medicine Genetically Engineered Mouse core facility, with floxed sites flanking exon 2. For FACS purification and visualization of astrocytes, the Aldh1l1-GFP reporter mouse line was used (*56*). Sox9-FF and Slc22a3-FF were crossed with Aldh1l1-GFP for histological analysis in conditional knockout mice. All above experimental procedures were approved by the Baylor College of Medicine IACUC.

Plasmids used for pAAV vectors

For chemogenetic experiments with Gq-DREADD, pAAV-hSyn-hM3D(Gq)-mCherry (Addgene plasmid no. 50474) was used. For Sox9 and Slc22a3 conditional knockout with Gfap-Cre-RFP, AAV-GFAP-iCre-P2A-TurboRFP (a kind gift from K. Ung and B. Arenkiel) was used (*13*). For controls with Gfap-mCh, pZac2.1-GfaAB1CD-mCherry was used. For two-photon calcium imaging experiments, we generated pAAV-Gfap-GCaMP6 from flexed GCaMP6m and pZac2.1-GfaBC1D-mCherry-hPMCA2w/b (Addgene no. 111568). For experiments with H3.3 control and H3.3Q5A mutant, we generated pAAV-Gfap-H3.3/H3.3Q5A-GFP from H3.3(wild type or Q5A)-Flag-HA(C-Ter) (*31*) in a pAAV-IRES-GFP backbone, wherein CMV promoter was replaced with Gfap promoter from AAV-GFAP-iCre-P2A-TurboRFP using Mlu I and Bam HI restriction sites. Purified plasmids were validated using Genewiz sequencing services. All viral expression vectors were packaged into AAV at the Viral Vector core facility at the Neurological Research Institute at Baylor College of Medicine.

Delivery of pAAV vectors into mouse brain by intraventricular or stereotaxic injection

For chemogenetic experiments, pAAV for Gq-DREADD (serotype 2/9, ~7.7 × 10^{12} genome copies (GC)/ml) was introduced by intraventricular injection into P1 to P2 pups. Trypan Blue dye (2.5 μl) was mixed with AAV (10 μl) before injection. Pups were anesthetized with hypothermia, and AAV mixed with dye (2 μl) was injected into each hemisphere. Pups were allowed to recover in cages placed partially on a low-voltage heating pad for at least 2 hours. Subsequent Gq-DREADD experiments was performed when pups reached adult age (>8 weeks). For region-specific AAV delivery into the OB, stereotaxis microinjection was performed in 2- to 3-month-old mice. The following viral titers were used: Gfap-mCh (~3 × 10^{12} GC/ml), Gfap-Cre (~2.6 × 10^{12} GC/ml), Gfap-GCaMP6-GFP (~6.8 × 10^{12} GC/ml), Gfap-H3.3-GFP (~5.2 × 10^{12} GC/ml), and Gfap-H3.3Q5A-GFP (~3.71 × 10^{12} GC/ml). Fast Green dye (1 μl) was mixed with AAV (10 μl) before injection. Mice were injected intraperitoneally with Buprenex (0.3 mg/ml), anesthetized with isoflurane, and placed into stereotaxic frame with the head secured by blunt ear bars and nose placed into anesthesia and ventilation system. A skin incision was made, followed by craniotomies of 2 to 3 mm in diameter above the OB powered by a high-speed drill. AAV was loaded into a microdispenser (Nanoject II, Drummond Scientific, no. 13-681-460) and injected using beveled glass pipettes (Drummond Scientific, no. 3-000-203-G/X) at a rate of 69 nl/s, at 10 s intervals 10 times in each OB using the following coordinates: 4.7 mm anterior to bregma, 0.6 mm lateral to midline, and depth of 1.5 mm beneath the surface of the skull. Glass pipettes were left in place for at least 3 min before and after AAV delivery before slow withdrawal. Surgical wounds were closed with sutures, and mice were allowed to recover in cages placed partially on a low-voltage heating pad for at least 2 hours. Buprenex (0.3 mg/ml) was administered, and mice were monitored for 3 days after surgery. All subsequent experimental analyses on these mice were performed 4 to 5 weeks after viral delivery. All animal procedures were performed in accordance with approved Baylor College of Medicine IACUC protocols.

In vivo neuronal activation

For chemogenetic neuronal activation (schematic shown in Fig. 1C), 2- to 3-month-old Gq-DREADD–expressing mice were administered CNO (Tocris no. 4936) at 0.3 mg/kg body weight intraperitoneally. After 30 min, mice were sacrificed, and brains were collected. Control animals were injected with saline. Additional controls were injected with CNO at the same dosage in mice expressing empty viral vector. Only Gq-DREADD–expressing mice that were administered CNO showed seizure-like behaviors within 10 to 15 min. For odor-evoked neuronal activation, experimental paradigm was based on a previous report (*57*) and modified here as shown in the schematics in Figs. 1J and 2I. Odorant isoamyl acetate was used because it is a standard nonbiological and nonpreferred odor. Control animals were exposed to mineral oil. Additional controls consisted of mice that were directly taken from home cages without habituation. For FACS purification and histological analyses, brains were collected as described below. For proteomic experiments, OBs were collected and snap frozen. For ChIP experiments, OBs were collected and processed immediately as described below.

Immunofluorescence on frozen brain tissues

Mice were anesthetized under isoflurane and perfused transcardially with phosphate-buffered solution (PBS) pH 7.4 followed by 4% paraformaldehyde (PFA) and brains were removed, fixed in 4% PFA overnight, placed in 20% sucrose overnight, and embedded in optimal cutting temperature (OCT) compound the next day. Cryosections of 20 μm were washed in PBS twice, and antigen retrieval was performed by incubating sections in 10 mM sodium citrate (pH 6.0; 0.05% Tween20) at 75°C for 10 min, blocking with 10% goat or donkey serum in PBS with 0.3% Triton X-100, and incubating with primary antibody dilutions

in blocking solution overnight at 4°C. The next day, sections were washed in PBS with 0.1% Triton X-100 and incubated with species-specific secondary antibody dilutions for 1 hour at room temperature. After Hoechst nuclear counterstaining (1:10,000; Thermo Fisher no. H3570), coverslips were mounted with VECTASHIELD antifade mounting medium.

The following primary antibodies were used: mouse anti-mCherry (1:500, Abcam no. ab125096), chicken anti-GFP (1:1000, Abcam no. ab13970), rabbit anti-Cre (1:500, Cell Signaling no. 15036), rabbit anti-Slc22a3 (1:250, Alomone Labs no. ACT-013), rabbit anti-Slc22a3 (1:100, Alpha Diagnostics no. OCT-31A), rabbit anti-Fos (1:200, Cell Signaling no. 2250), rabbit anti-H3K4me3Q5ser (1:500, EMD Millipore no. ABE2580), mouse anti-NeuN (1:500, EMD Millipore no. MAB377), guinea pig anti-GABA (1:200, EMD Millipore no. AB175), rabbit anti-MAOB (1:100, Proteintech no. 12602-1-AP), rabbit anti-Aldh1a1 (1:200, Abcam no. ab52492), goat anti-serotonin (1:200, Abcam no. ab66047), mouse anti-Gfap (1:500, EMD Millipore no. MAB360), mouse anti-Ezrin (1:100, BioLegend no. 866401), and rabbit anti-PSD95 (1:500, Invitrogen no. 51-6900).

The following secondary antibodies were used (1:500, 0.1% Triton X-100 in PBS): goat anti-mouse Alexa Fluor 568 (Thermo Fisher no. A11004), goat anti-mouse Alexa Fluor 488 (Thermo Fisher no. A11001), goat anti-mouse Alexa Fluor 647 (Thermo Fisher no. A21235), goat anti-chicken Alexa Fluor 488 (Thermo Fisher no. A11039), goat anti-rabbit Alexa Fluor 568 (Thermo Fisher no. A11036), goat anti-rabbit Alexa Fluor 488 (Thermo Fisher no. A11034), goat anti-rabbit Alexa Fluor 647 (Thermo Fisher no. A21244), goat anti–guinea pig Alexa Fluor 568 (Thermo Fisher no. A11075), donkey anti-goat Alexa Fluor 568 (Thermo Fisher no. A11057), donkey anti-chicken Alexa Fluor 488 (Thermo Fisher no. 78948), and donkey anti-mouse Alexa Fluor 647 (Thermo Fisher no. A31571).

Confocal imaging and analyses

Fluorescent images were acquired using a Zeiss LSM 980 confocal microscope with a 20× or 40× oil objective. All images were taken from the granular cell layer of the OB in the region near the internal plexiform layer (fig. S11B). We focused on this region because both Fos and Slc22a3 showed a robust expression pattern in this region. For quantification, both control and experimental groups were immunostained on the same day, images were acquired in one session on the same day, and identical laser power settings were maintained for all cohorts under comparison. All quantification was done on raw unprocessed images. Quantification was done using ImageJ, wherein the region of interest (ROI) was detected manually based on the Aldh1l1-GFP–labeled astrocyte boundary, and area or mean intensity values were extracted. Astrocytes for quantification were selected based on clear GFP reporter labeling, and all analyses were done blind to conditions. Quantification of cell numbers was perfromed using the Cell Counter or Analyze Particles function in ImageJ. For visualization, box plots were made in R-studio using ggplot2.

FACS purification

Different mouse brain regions (CX, HP, and OB) were collected and dissociated using a previously described protocol (*58*). Dissociated cells were sorted on a BD FACSAria III instrument (100-μm nozzle). Around 95,000 cells were collected per 1.5-ml tube, which contained Buffer RLT (650 μl, Qiagen no. 79216) with 1% β-mercaptoethanol. Finally, each sample was vortexed and rapidly frozen on dry ice.

RNA extraction, library preparation, and RNA-seq

Total RNA was extracted using the RNeasy Micro Kit (Qiagen no. 74004), and quality control was performed using the High Sensitivity RNA Analysis Kit (Agilent no. 472-0500) on a 12-capillary fragment analyzer. cDNA synthesis and library construction with 8–base pair (bp) single indices were done from 10 ng total RNA using Trio RNASeq System (NuGEN no. 0507-96). The resulting libraries were validated using the Standard Sensitivity NGS Fragment Analysis Kit (Agilent no. DNF-473-0500) and quantified using Quant-it dsDNA Assay kit (Thermo Fisher no. Q33120). Samples were diluted to equimolar concentrations (2 nM), and pooled and denatured according to the manufacturer's instructions. A final library dilution of 1.3 pM was subjected to paired-end (read 1: 75, read 2: 75) sequencing of ~40 million reads per sample using the High Output v2 kit (Illumina no. FC-404-2002) on a NextSeq550 instrument.

RNA-seq bioinformatics analysis

Sequencing files from each flow cell lane were downloaded in fastq files and merged. Quality control was performed using fastQC (v0.10.1) and MultiQC (v0.9). Reads were mapped to the mouse genome mm10 assembly using STAR (v2.5.0a). In R (v4.1.2), mapped reads were used to build count matrices using Bioconductor packages GenomicAlignments (v1.26.0) and GenomicFeatures (v1.42.2) (*59*). University of California Santa Cruz transcripts were downloaded from Illumina iGenomes in the GTF file format. DESeq2 (v1.20.0) (*60*) was used for normalization and differential gene expression analysis. Motif analysis was performed using Hypergeometric Optimization of Motif Enrichment (HOMER, v4.10) to identify transcription factor motifs enriched within 1000 bp before and 500 bp after the transcription start site. Motifs with $p < 0.05$ were selected and further filtered based on two criteria: (i) >1.5-fold expression in GFP astrocytes over mCh neurons and (ii) transcript expression >150 counts per million in GFP astrocytes. GOs were determined using Enrichr, and significant GO terms ($p < 0.01$) were selected. For visualization, plots were made using ggplot2 (v3.3.5). Gene expression heatmaps were generated using ComplexHeatmap (v2.6.2).

Reverse transcription quantitative PCR

RNA extraction and cDNA preparation were performed as described above for sample preparation of RNA-seq. Reverse transcription quantitative PCR (RT-qPCR) was performed using PerfeCTa SYBR Green Fast Mix (Quantabio no. 95072-012) on a Roche Light Cycler 480 instrument. Reactions were set up using 2 ng cDNA, 250 nM primers, and 1× SYBR mix. qPCR was carried out at 95°C for 30 s, 40 cycles of 95°C for 5 s, and 60°C for 30 s, with a subsequent melting-curve analysis. The expression of transcripts of target genes was normalized to glyceraldehyde phosphate dehydrogenase (*Gapdh*). The RT-qPCR primers that were used are as follows: forward 5′- GGAGACCCACTC-TACCATCGT-3′, reverse 5′-GCTGCATAGCCCA-AGGTAAAA-3′ (for *Slc22a3*); and forward 5′-TGGCCTTCCGTGTTCCTAC-3′, reverse 5′- GAG-TTGCTGTTGAAGTCGCA-3′ (for *Gapdh*).

ChIP

For ChIP-Sox9 and ChIP-Sox2 in the brain, we pooled cortices, hippocampi, and OBs from three mice for each experimental cohort. For ChIP-Sox9 from odor exposed mice, we pooled OBs from six to eight mice for each experimental cohort. For ChIP-H3S5HT, we pooled OBs from four mice for each experimental cohort and performed two independent sequencing runs. Immediately after harvesting, tissues were dissociated in cold PBS using a pellet homogenizer on ice. Chromatin was cross-linked using a freshly prepared 1.1% formaldehyde solution with rocking at room temperature for 10 min, followed by addition of 0.1 M glycine. Cell pellets were collected by centrifugation at 3500 rpm for 5 min at 4°C, washed with PBS, and frozen at 80°C until further processing. Pellets were resuspended with PBS–phenylmethylsulfonyl fluoride (PMSF) containing 0.5% Igepal to release nuclei followed by washing with cold ChIP-Buffer (0.25% Triton-X100, 10 mM EDTA, 0.5 mM EGTA, 10 mM HEPES pH 6.5), and nuclei were lysed with ChIP lysis buffer (0.5% SDS, 5 mM EDTA, 25 mM Tris-HCl pH 8) for 15 to 20 min at room temperature. Lysates were sonicated to 250 to 350 bp using Diagenode Bioruptor. Chromatin was quantified using the Quant-iT double-stranded DNA (dsDNA) Assay kit (Thermo Fisher, no. Q33120), diluted fivefold (2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 20 mM

Tris-HCl; pH 8.0; with protease inhibitors), and incubated with antibody overnight at 4°C with rotation. For ChIP-Sox9, 70 to 100 μg of chromatin was incubated with rabbit anti-Sox9 (7 to 10 μg; Abcam no. ab5535). For ChIP-Sox2, 70 to 100 μg of chromatin was incubated with rabbit anti-Sox2 (7-10 μg; EMD Millipore no. AB5603). For ChIP-H35HT, 10 to 15 μg chromatin was incubated with rabbit anti-H3K4me3Q5ser (3 μg; EMD Millipore no. ABE2580). The next day, lysates were incubated with Protein A/G magnetic beads (Thermo no. 88802) for 5 to 6 hours at 4°C, followed by washing with Tris-SDS-EDTA-I buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl; pH 8.0), Tris-SDS-EDTA-II buffer (TSEI buffer with 500 mM NaCl), LiCl buffer (250 mM LiCl, 1% Nonidet P-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0), and Tris-EDTA buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). To release DNA fragments, samples were incubated in freshly prepared elution buffer (1% SDS, 0.1 M $NaHCO_3$) for 20 min at 65°C twice. Elutions were treated with proteinase K (0.4 mg/ml: Thermo Fisher no. AM2546) and NaCl (0.125 M) overnight at 65°C for reverse cross-linking. Subsequently, ChIP-DNA was purified using a PCR purification kit (Qiagen no. 28104) and quantified using the Quant-iT dsDNA Assay kit. For ChIP-PCR, additional control samples were prepared, wherein sonicated lysates were incubated with rabbit anti–immunoglobulin G (IgG) (R&D Systems no. AB-105-C). Subsequently, purified ChIP-DNA was analyzed in PCR reactions using AccuPrime Pfx DNA polymerase (Thermo Fisher no. 12344-032) that were carried out at 95°C for 5 min; 40 cycles of 95°C for 30 s, 55°C for 1 min, and 68°C for 30 s; and followed by 68°C for 2 min. *Slc22a3* primers used for ChIP-PCR were designed at the Sox9 binding site at the *Slc22a3* promoter (forward 5′-CTGTCCCTCT-GTCCCATTGT-3′, reverse 5′- TTCCAGGATCA-CCCAGACTC-3′). For ChIP-seq experiments, 10 to 12 ng of ChIP-DNA was used for library preparation as described below.

ChIP-seq library preparation, sequencing, and bioinformatic analysis

ChIP libraries were prepared using the TruSeq ChIP Library Preparation Kit (Illumina no. IP-202-1012), according to the manufacturer's instructions. Libraries ranging from 250 to 350 bp were extracted from gel incisions using the QIAquick Gel Extraction Kit (Qiagen no. 28706), PCR amplified, and purified using AMPure XP beads (Beckman Coulter Life Science no. A63882). The quality of the resulting libraries was analyzed on the Standard Sensitivity NGS Fragment Analysis Kit (Agilent no. DNF-473-0500) on a 12-capillary Fragment Analyzer. Libraries were quantified using the Quant-iT dsDNA assay kit (Thermo Fisher no. Q33120), and equal concentrations (2 nM) of libraries were pooled and subjected to single-end (read 1: 150) sequencing of ~60 million to 80 million reads per sample using the High Output v2 kit (Illumina no. FC-404-2002) on a NextSeq550 following the manufacturer's instructions.

Sequencing files from each flow cell lane were downloaded, and the resulting fastq files were merged. Quality control was performed using fastQC (v0.11.17) and MultiQC (v0.9). Reads were mapped to the mouse genome mm10 assembly using bowtie2 (v 2.2.6) (*61*). Using the HOMER (v4.10) software suite (*62*), bedgraph files and tag directories were made. The findPeaks command in factor (for Sox9, Sox2) or histone (for H3-5HT) mode was used to filter ChIP peaks enriched over input control. Annotation of enriched peaks was performed using annotatePeaks with mm10 assembly. Integrated Genome Browser–compatible files were made using samtools (v1.9), sort and index, deepTools (v3.2.0), and bamCompare (*63*, *64*). Overlapping peaks, unique peaks, and differentially bound peaks were obtained using mergePeaks or getDifferentialPeaks, and peaks were visualized using computeMatrix and plotHeatmap. Motif analysis for TFs was performed using findMotifsGenome.pl at 1000 bp from the peak center. GOs were determined by submitting genes associated with ChIP peaks at Enrichr, and significant GO terms ($p < 0.01$) were selected for visualization using ggplot.

Immunoprecipitation followed by mass spectrometry

We collected OBs that were immediately snap frozen from six to eight mice for each experimental cohort. Tissues were thawed, pellets homogenized, and nuclear lysates extracted using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher no. 78833) according to the manufacturer's instructions. Lysates were ultracentrifuged at 200,000*g* for 20 min at 4°C, and the supernatant (3 to 5 mg total protein) was used for immunoprecipitation with anti-Sox9 (5 μg, Abcam no. ab5535) for 1 hour at 4°C, followed by incubation with protein A Sepharose slurry (GE Healthcare Life Sciences) for another 1 hour at 4°C. The beads were collected by centrifugation at 1000*g* for 1 min, washed with NETN buffer (50 mM Tris pH 7.3, 1 mM EDTA, 0.5% NP-40) multiple times, and heated for 10 min at 90°C with 20 ml of 2X SDS loading dye to elute bound proteins. Negative control samples were prepared using the same methods without addition of anti-Sox9 antibody. The immunoprecipitated samples were resolved on NuPAGE 10% Bis-Tris Gel (Life Technologies), and the gel pieces were processed for in-gel digestion using trypsin enzyme (GenDEPOT no. T9600). The tryptic peptides were analyzed on a nano-LC 1200 system (Thermo Fisher Scientific) coupled to a Orbitrap Fusion Lumos (Thermo Fisher Scientific) mass spectrometer. The tandem mass spectrometry spectra was searched using Mascot algorithm (Mascot 2.4, Matrix Science) against the mouse National Center for Biotechnology Information (NCBI) refseq protein database in the Proteome Discoverer (PD1.4, Thermo Fisher) interface. The precursor mass tolerance was confined to 20 parts per million (ppm), and a fragment mass tolerance of 0.5 dalton and maximum of two missed cleavage were allowed. Dynamic modification of oxidation on methionine, protein N-terminal acetylation, destreak on cysteine, and phosphorylation on serine, threonine, and tyrosine were allowed. The assigned peptides were filtered at 5% false discovery rate (FDR) using the percolator *q* value. Protein quantification was performed using the iBAQ approach.

Western blot

Whole cell lysates were prepared in RIPA lysis buffer, run on a 10% SDS polyacrylamide gel, followed by wet transfer to a nitrocellulose membrane at 400 mA for 45 min. The membrane was blocked by 5% milk in Tris-buffered saline with Tween20 (TBST), followed by incubation overnight at 4°C. The following primary antibodies were used: rabbit anti-Sox9 (1:500 dilution, Abcam no. ab5535), rabbit anti-Sox2 (1:500, EMD Millipore no. AB5603), rabbit anti-Fos (1:500, Cell Signaling no. 2250S), and mouse anti-Gapdh (1:500 dilution, EMD Millipore no. MAB374). The next day, membranes were washed three times with TBST, incubated at room temperature for 1 hour in horseradish peroxidase–conjugated IgG at 1:2000 dilution in 5% milk, washed again three times with TBST, and developed using luminol reagent (Santa Cruz Biotechnology no. sc2048). For Western blot after IP, nuclear lysates were prepared as described above for sample preparation for mass spectrometry. The following primary antibodies were used: rabbit anti-Sox9 (1:500 dilution, Abcam no. ab5535) and rabbit anti-Sox2 (1:500, EMD Millipore no. AB5603). Subsequent pull-down was performed by adding protein A agarose beads (Thermo Fisher no. 15918-014) for an additional 5 hours at 4°C. The beads were collected, washed, and boiled in 2× SDS gel loading dye to elute bound proteins and run for Western blot.

Olfactory behavior

Behavioral assays for odor detection and odor discrimination were performed as described previously (*13*). Odor detection was performed by olfactory habituation and dishabituation in a three-chamber place-preference assay. Mice were acclimated and familiarized to the testing chamber for 4 min each day for 3 days

before testing. Mice were first habituated to the outer chambers, which contained mineral oil, and the middle chamber, which served as a neutral barrier, three times, after which mineral oil in one of the outer chambers was swapped with a novel odorant (*R*)-limonene in a series of increasing concentrations (10^{-6}, 10^{-5}, 10^{-4}, and 10^{-3} v/v dilution in oil). Odor detection was calculated based on the time spent investigating the odorant-containing chamber compared with the mineral oil–containing chamber and expressed as zone preference index. To assay for differences in odor discrimination, a similar experimental paradigm was used, wherein a different but structurally similar odorant (*S*)-limonene was introduced and mineral oil was replaced with (*R*)-limonene. Mice were first habituated to a high concentration of (*R*)-limonene (10^{-3} v/v dilution) in both outer chambers before introduction of (*S*)-limonene (10^{-3} v/v dilution) in one of the chambers. Videos were captured with a Logitech HD 1080p camera. Preference for an odor was determined based on the time spent in each chamber and calculated as zone preference index using MATLAB software with the Optimouse plug-in for analysis of mouse positions (*65*).

Imaging and Sholl analysis

Fluorescent images for morphological evaluation were acquired using a Zeiss LSM 980 confocal microscope with a 63× oil immersion objective with frame size at 1024 × 1024 and bit depth at 12. Serial images at the *z* axis were taken at an optical step of 0.5 mm, with an overall *z*-axis range encompassing the whole section. Images were imported to Imaris Bitplane software, and only astrocytes with their soma between the *z*-axis range were chosen for further analysis. We performed 3D surface rendering using the Imaris Surface module, and color coded the reconstructed surface images based on the surface area of each astrocyte. Morphological analysis was performed using the Imaris Filament module. Astrocyte branches and processes were outlined by Autopath, with a starting point set at 8 mm and a seed point set at 0.7 mm, and statistical outputs including "filament number Sholl intersections" were extracted and plotted. To measure the shortest distance between Ezrin and PSD 95, fluorescent images were acquired using a Zeiss LSM 880 laser scanning confocal microscope with a frame size of 1024 × 1024 and a bit depth of 8. Serial images at the *z*-axis were taken at an optical step of 1 μm. The images were imported and converted into ims format with IMARIS software (Bitplane 9.2.1). The converted images were imported to IMARIS software (Bitplane 9.7.0). For visualization of Ezrin and PSD 95, we performed each 3D surface rendering of Ezrin and PSD 95 based on the GFP^+ astrocytes using the IMARIS Surface module and color coded the reconstructed surface images based on the surface area. For analyzing the shortest distance between Ezrin and PSD95, the surfaces of Ezrin and PSD95 were analyzed by object × object statistics in the Surface module. The parameters of distances between Ezrin and PSD95 were automatically calculated, and we collected the number of attached PSD95 on Ezrin in each other groups (the distance = 0 μm).

Two-photon GCaMP6 calcium imaging and analysis

Animals were anesthetized with isoflurane, and isolated brains were submerged in ice-cold artificial cerebrospinal fluid (ACSF) solution (130 mM NaCl, 24mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 3.5 mM KCl, 1.5 mM $CaCl_2$, 1.5 mM $MgCl_2$, and 10 mM D(+)-glucose, pH 7.4). The 300-μm OB slices were cut using a vibratome (DSK Linear Slicer) oxygenated in ACSF. Slices were then recovered in oxygenated ACSF for 15 min and allowed to acclimate to room temperature for at least 30 min before imaging. We recorded calcium activity using a two-photon resonant microscope (LSM 7MP, Zeiss) equipped with a Coherent Chameleon Ultra (II) Ti-sapphire laser tuned to 900 nm and a 20×, 1.0 NA Zeiss objective. Calcium activity was typically sampled at ~1 Hz. Optical signals were recorded at 1024-pixel–by–1024-pixel resolution. We recorded data from astrocytes at depths of 30 μm below the surface. All multiphoton imaging experiments were performed within 2 to 4 hours of slicing. For serotonin-induced calcium imaging, optical signals were recorded after slices were bathed in 500 nM TTX for 5 min. After 200 s of recording under TTX treatment, brain slices were bathed in 50 μM serotonin (Thermo Fisher no. AAB2126303) and recorded for an additional 300 s. For glutamate (Sigma no. G1251)–induced calcium imaging, slices were bathed in 300 μM instead of serotonin. Image analysis of spontaneous or induced Ca^{2+} were quantified using GECIquant algorithm with ImageJ software, wherein detection of ROIs for soma was performed in a semiautomated manner as described in a previous study (*14*, *66*). After thresholding from temporally projected stack images with a maximum intensity projection, a polygon selection was manually drawn around the approximate astrocyte territory of interest, and the selection was added to the ImageJ ROI manager. The area criterion was 30 μm to infinity for soma within the GECIquant ROI detection function. Intensity values for each ROI were extracted in ImageJ and converted to change in fluorescence intensity relative to resting fluorescence intensity (d*F*/*F*) values. For each ROI, basal *F* was determined during 40 s periods with no fluctuations. Clampfit 10.7 software was used to detect and measure amplitude and frequency values for the somatic and microdomain transients. We counted the response after with these criteria: amplitude (>0.5 d*F*/*F*), pre–trigger time (3 ms), and minimum duration (5 ms). Data points were plotted on Prism software.

Slice recording for EPSC, IPSC, and tonic GABA

Animals were anesthetized with isoflurane, and isolated brains were submerged in ice-cold ACSF solution (130 mM NaCl, 24mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 3.5 mM KCl, 1.5 mM $CaCl_2$, 1.5 mM $MgCl_2$, and 10 mM D(+)-glucose, pH 7.4). Then, 300-μm slices were cut using a vibratome (DSK Linear Slicer) oxygenated in ACSF at room temperature for 1 hour and then acclimated at room temperature with continuous perfusion with ASCF solution (2 ml/min). Slices were placed in the recording chamber, and target cells were identified with an upright Olympus microscope with a 60× water immersion objective with infrared differential interference contrast optics. Whole-cell recording was performed with pCLAMP10 and MultiClamp 700B amplifier (Axon Instrument, Molecular Devices) at room temperature from OB granule cells. The holding potential was −60 mV. Pipette resistance was typically 5 to 8 megohm. The pipette was filled with an internal solution (in mM): 135 CsMeSO4, 8 NaCl, 10 HEPES, 0.25 EGTA, 1 Mg-ATP, 0.25 Na2-GTP, 30 QX-314, pH adjusted to 7.2 with CsOH (278-285 mOsmol) for EPSC measurement, 135 CsCl, 4 NaCl, 0.5 $CaCl_2$, 10 HEPES, 5 EGTA, 2 Mg-ATP, 0.5 Na2-GTP, 30 QX-314, pH adjusted to 7.2 with CsOH (278-285 mOsmol) for IPSCs and tonic current measurement. IPSC and tonic current were measured in the presence of ionotropic glutamate receptor antagonists, (2*R*)-amino-5-phosphonovaleric acid (APV) (50 μM, Tocris), and cyanquixaline (CNQX) (20 μM, Tocris). Electrical signals were digitized and sampled at 50-μs intervals with Digidata 1550B and Multiclamp 700B amplifier (Molecular Devices) using pCLAMP 10.7 software. Data were filtered at 2 kHz. The recorded current was analyzed with ClampFit 10.7 software. Data points were plotted on Prism software.

Statistical analysis

Sample sizes and statistical tests are provided in all figure legends. Sample sizes were determined based on published research (*13*, *14*). Mice from each cohort were randomly allocated to experimental groups. All analyses were done blind to experimental conditions. Summary data of all means, SEMs, *p* values, sample sizes, and statistical methods used are provided in table S3. Data were tested for normality using the Shapiro-Wilk tests and for homogeneity of variance using the Levene test. Parametric tests were used for normally distributed datasets and for data with a small sample size ($n = 3$). For comparison of two

groups, unpaired or paired Student's *t* test was used, and for comparison of three groups, one-way analysis of variance (ANOVA) followed by Tukey's tests was used. For multiple comparisons, two-way ANOVA with Sidak's correction was used. When data did not follow a normal distribution, nonparametric Wilcoxon rank sum tests were applied. Significant differences are denoted by asterisks in associated graphs. Data in bar graphs are presented as mean ± SEM. Box plots were generated using ggsignif (v0.6.0) with ggplot2 (v3.3.2). Levels of statistical significance are indicated as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and **** $p < 0.0001$.

REFERENCES AND NOTES

1. Y.-T. Cheng, J. Woo, B. Deneen, Sculpting astrocyte diversity through circuits and transcription. *Neuroscientist* **2022**, 10738584221082620 (2022). doi: 10.1177/10738584221082620; pmid: 35373633
2. P. Kofuji, A. Araque, Astrocytes and behavior. *Annu. Rev. Neurosci.* **44**, 49–67 (2021). doi: 10.1146/annurev-neuro-101920-112225; pmid: 33406370
3. B. S. Khakh, B. Deneen, The emerging nature of astrocyte diversity. *Annu. Rev. Neurosci.* **42**, 187–207 (2019). doi: 10.1146/annurev-neuro-070918-050443; pmid: 31283899
4. G. Dallérac, J. Zapata, N. Rouach, Versatile control of synaptic circuits by astrocytes: Where, when and how? *Nat. Rev. Neurosci.* **19**, 729–743 (2018). doi: 10.1038/s41583-018-0080-6; pmid: 30401802
5. N. Bazargani, D. Attwell, Astrocyte calcium signaling: The third wave. *Nat. Neurosci.* **19**, 182–189 (2016). doi: 10.1038/nn.4201; pmid: 26814587
6. A. Kol *et al*., Astrocytes contribute to remote memory formation by modulating hippocampal-cortical communication during learning. *Nat. Neurosci.* **23**, 1229–1239 (2020). doi: 10.1038/s41593-020-0679-6; pmid: 32747787
7. J. Nagai *et al*., Hyperactivity with disrupted attention by activation of an astrocyte synaptogenic cue. *Cell* **177**, 1280–1292.e20 (2019). doi: 10.1016/j.cell.2019.03.019; pmid: 31031006
8. A. Adamsky *et al*., Astrocytic activation generates de novo neuronal potentiation and memory enhancement. *Cell* **174**, 59–71.e14 (2018). doi: 10.1016/j.cell.2018.05.002; pmid: 29804835
9. M. Martin-Fernandez *et al*., Synapse-specific astrocyte gating of amygdala-related behavior. *Nat. Neurosci.* **20**, 1540–1548 (2017). doi: 10.1038/nn.4649; pmid: 28945222
10. G. Perea, A. Yang, E. S. Boyden, M. Sur, Optogenetic astrocyte activation modulates response selectivity of visual cortex neurons in vivo. *Nat. Commun.* **5**, 3262 (2014). doi: 10.1038/ncomms4262; pmid: 24500276
11. J. L. Stobart *et al*., Cortical circuit activity evokes rapid astrocyte calcium signals on a similar timescale to neurons. *Neuron* **98**, 726–735.e4 (2018). doi: 10.1016/j.neuron.2018.03.050; pmid: 29706581
12. E.-L. Yap, M. E. Greenberg, Activity-regulated transcription: Bridging the gap between neural activity and behavior. *Neuron* **100**, 330–348 (2018). doi: 10.1016/j.neuron.2018.10.013; pmid: 30359600
13. K. Ung *et al*., Olfactory bulb astrocytes mediate sensory circuit processing through Sox9 in the mouse brain. *Nat. Commun.* **12**, 5230 (2021). doi: 10.1038/s41467-021-25444-3; pmid: 34471129
14. A. Y.-S. Huang *et al*., Region-specific transcriptional control of astrocyte function oversees local circuit activities. *Neuron* **106**, 992–1008.e9 (2020). doi: 10.1016/j.neuron.2020.03.025; pmid: 32320644
15. I. Farhy-Tselnicker *et al*., Activity-dependent modulation of synapse-regulating genes in astrocytes. *eLife* **10**, e70514 (2021). doi: 10.7554/eLife.70514; pmid: 34494546
16. D. Sardar *et al*., Mapping astrocyte transcriptional signatures in response to neuroactive compounds. *Int. J. Mol. Sci.* **22**, 3975 (2021). doi: 10.3390/ijms22083975; pmid: 33921461
17. P. Hasel *et al*., Neurons and neuronal activity control gene expression in astrocytes to regulate their development and metabolism. *Nat. Commun.* **8**, 15132 (2017). doi: 10.1038/ncomms15132; pmid: 28462931
18. S. Hrvatin *et al*., Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018). doi: 10.1038/s41593-017-0029-5; pmid: 29230054
19. B. S. Khakh, K. D. McCarthy, Astrocyte calcium signaling: From observations to functions and the challenges therein. *Cold Spring Harb. Perspect. Biol.* **7**, a020404 (2015). doi: 10.1101/cshperspect.a020404; pmid: 25605709
20. K. D. Fischer, A. C. W. Houston, G. V. Rebec, Role of the major glutamate transporter GLT1 in nucleus accumbens core versus shell in cue-induced cocaine-seeking behavior. *J. Neurosci.* **33**, 9319–9327 (2013). doi: 10.1523/JNEUROSCI.3278-12.2013; pmid: 23719800
21. M. Brancaccio, A. P. Patton, J. E. Chesham, E. S. Maywood, M. H. Hastings, Astrocytes control circadian timekeeping in the suprachiasmatic nucleus via glutamatergic signaling. *Nat. Neurosci.* **21**, 120–129 (2018).
22. J. A. Gomez *et al*., Ventral tegmental area astrocytes orchestrate avoidance and approach behavior. *Nat. Commun.* **10**, 1455 (2019). doi: 10.1038/s41467-019-09131-y; pmid: 30926783
23. M. D. Scofield, P. W. Kalivas, Astrocytic dysfunction and addiction: Consequences of impaired glutamate homeostasis. *Neuroscientist* **20**, 610–622 (2014). doi: 10.1177/1073858413520347; pmid: 24496610
24. X. Yu *et al*., Reducing astrocyte calcium signaling in vivo alters striatal microcircuits and causes repetitive behavior. *Neuron* **99**, 1170–1187.e9 (2018). doi: 10.1016/j.neuron.2018.08.015; pmid: 30174118
25. J. Woo *et al*., Control of motor coordination by astrocytic tonic GABA release through modulation of excitation/inhibition balance in cerebellum. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5004–5009 (2018). doi: 10.1073/pnas.1721187115; pmid: 29691318
26. S. Mederos *et al*., GABAergic signaling to astrocytes in the prefrontal cortex sustains goal-directed behaviors. *Nat. Neurosci.* **24**, 82–92 (2021). doi: 10.1038/s41593-020-00752-x; pmid: 33288910
27. M. Paukert *et al*., Norepinephrine controls astroglial responsiveness to local circuit activity. *Neuron* **82**, 1263–1270 (2014). doi: 10.1016/j.neuron.2014.04.038; pmid: 24945771
28. T. Papouin, J. M. Dunphy, M. Tolman, K. T. Dineley, P. G. Haydon, Septal cholinergic neuromodulation tunes the astrocyte-dependent gating of hippocampal NMDA receptors to wakefulness. *Neuron* **94**, 840–854.e7 (2017). doi: 10.1016/j.neuron.2017.04.021; pmid: 28479102
29. M. Corkrum *et al*., Dopamine-evoked synaptic regulation in the nucleus accumbens requires astrocyte activity. *Neuron* **105**, 1036–1047.e5 (2020). doi: 10.1016/j.neuron.2019.12.026; pmid: 31954621
30. P. Mews *et al*., Acetyl-CoA synthetase regulates histone acetylation and hippocampal memory. *Nature* **546**, 381–386 (2017). doi: 10.1038/nature22405; pmid: 28562591
31. L. A. Farrelly *et al*., Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **567**, 535–539 (2019). doi: 10.1038/s41586-019-1024-7; pmid: 30867594
32. A. E. Lepack *et al*., Dopaminylation of histone H3 in ventral tegmental area regulates cocaine seeking. *Science* **368**, 197–201 (2020). doi: 10.1126/science.aaw8806; pmid: 32273471
33. C. Chen *et al*., Astrocyte-specific deletion of *Sox2* promotes functional recovery after traumatic brain injury. *Cereb. Cortex* **29**, 54–69 (2019). doi: 10.1093/cercor/bhx303; pmid: 29161339
34. F. Petrelli *et al*., Dysfunction of homeostatic control of dopamine by astrocytes in the developing prefrontal cortex leads to cognitive impairments. *Mol. Psychiatry* **25**, 732–749 (2020). doi: 10.1038/s41380-018-0226-y; pmid: 30127471
35. W. Song *et al*., Organic cation transporter 3 (Oct3) is a distinct catecholamines clearance route in adipocytes mediating the beiging of white adipose tissue. *PLOS Biol.* **17**, e2006571 (2019). doi: 10.1371/journal.pbio.2006571; pmid: 30653498
36. M. Cui *et al*., The organic cation transporter-3 is a pivotal modulator of neurodegeneration in the nigrostriatal dopaminergic pathway. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8043–8048 (2009). doi: 10.1073/pnas.0900358106; pmid: 19416912
37. V. Vialou *et al*., Altered aminergic neurotransmission in the brain of organic cation transporter 3-deficient mice. *J. Neurochem.* **106**, 1471–1482 (2008). doi: 10.1111/j.1471-4159.2008.05506.x; pmid: 18513366
38. N. L. Baganz *et al*., Organic cation transporter 3: Keeping the brake on extracellular serotonin in serotonin-transporter-deficient mice. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18976–18981 (2008). doi: 10.1073/pnas.0800466105; pmid: 19033200
39. J. F. Oliveira, V. M. Sardinha, S. Guerra-Gomes, A. Araque, N. Sousa, Do stars govern our actions? Astrocyte involvement in rodent behavior. *Trends Neurosci.* **38**, 535–549 (2015). doi: 10.1016/j.tins.2015.07.006; pmid: 26316036
40. S. Lee *et al*., Channel-mediated tonic GABA release from glia. *Science* **330**, 790–796 (2010). doi: 10.1126/science.1184334; pmid: 20929730
41. I. Maze *et al*., Critical role of histone turnover in neuronal transcription and plasticity. *Neuron* **87**, 77–94 (2015). doi: 10.1016/j.neuron.2015.06.014; pmid: 26139371
42. Y. Bernardinelli *et al*., Activity-dependent structural plasticity of perisynaptic astrocytic domains promotes excitatory synapse stability. *Curr. Biol.* **24**, 1679–1688 (2014). doi: 10.1016/j.cub.2014.06.025; pmid: 25042585
43. A. Perez-Alvarez, M. Navarrete, A. Covelo, E. D. Martin, A. Araque, Structural and functional plasticity of astrocyte processes and dendritic spine interactions. *J. Neurosci.* **34**, 12738–12744 (2014). doi: 10.1523/JNEUROSCI.2401-14.2014; pmid: 25232111
44. H. Chai *et al*., Neural circuit-specialized astrocytes: Transcriptomic, proteomic, morphological, and functional evidence. *Neuron* **95**, 531–549.e9 (2017). doi: 10.1016/j.neuron.2017.06.029; pmid: 28712653
45. J. E. Burda *et al*., Divergent transcriptional regulation of astrocyte reactivity across disorders. *Nature* **606**, 557–564 (2022). doi: 10.1038/s41586-022-04739-5; pmid: 35614216
46. D. Brunert, Y. Tsuno, M. Rothermel, M. T. Shipley, M. Wachowiak, Cell-type-specific modulation of sensory responses in olfactory bulb circuits by serotonergic projections from the raphe nuclei. *J. Neurosci.* **36**, 6820–6835 (2016). doi: 10.1523/JNEUROSCI.3667-15.2016; pmid: 27335411
47. Y.-T. Cheng *et al*., Social deprivation induces astrocytic TRPA1-GABA suppression of hippocampal circuits. *Neuron* **111**, 1301–1315.e5 (2023). doi: 10.1016/j.neuron.2023.01.015; pmid: 36787749
48. K. C. Benton *et al*., Norepinephrine activates β_1-adrenergic receptors at the inner nuclear membrane in astrocytes. *Glia* **70**, 1777–1794 (2022). doi: 10.1002/glia.24219; pmid: 35589612
49. Y. Su *et al*., Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat. Neurosci.* **20**, 476–483 (2017). doi: 10.1038/nn.4494; pmid: 28166220
50. L.-F. Chen *et al*., Enhancer histone acetylation modulates transcriptional bursting dynamics of neuronal activity-inducible genes. *Cell Rep.* **26**, 1174–1188.e5 (2019). doi: 10.1016/j.celrep.2019.01.032; pmid: 30699347
51. D. Sardar *et al*., Sox9 directs divergent epigenomic states in brain tumor subtypes. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202015119 (2022). doi: 10.1073/pnas.2202015119; pmid: 35858326
52. G. C. Petzold, A. Hagiwara, V. N. Murthy, Serotonergic modulation of odor input to the mammalian olfactory bulb. *Nat. Neurosci.* **12**, 784–791 (2009). doi: 10.1038/nn.2335; pmid: 19430472
53. V. Kapoor, A. C. Provost, P. Agarwal, V. N. Murthy, Activation of raphe nuclei triggers rapid and distinct effects on parallel olfactory bulb output channels. *Nat. Neurosci.* **19**, 271–282 (2016). doi: 10.1038/nn.4219; pmid: 26752161
54. Z. Ma, T. Stork, D. E. Bergles, M. R. Freeman, Neuromodulators signal through astrocytes to alter neural circuit activity and behaviour. *Nature* **539**, 428–432 (2016). doi: 10.1038/nature20145; pmid: 27828941
55. Y. Mu *et al*., Glia accumulate evidence that actions are futile and suppress unsuccessful behavior. *Cell* **178**, 27–43.e19 (2019). doi: 10.1016/j.cell.2019.05.050; pmid: 31230713
56. N. Heintz, Gene expression nervous system atlas (GENSAT). *Nat. Neurosci.* **7**, 483 (2004). doi: 10.1038/nn0504-483; pmid: 15114362
57. A. K. Bepari, K. Watanabe, M. Yamaguchi, N. Tamamaki, H. Takebayashi, Visualization of odor-induced neuronal activity by immediate early gene expression. *BMC Neurosci.* **13**, 140 (2012). doi: 10.1186/1471-2202-13-140; pmid: 23126335
58. C.-C. J. Lin *et al*., Identification of diverse astrocyte populations and their malignant analogs. *Nat. Neurosci.* **20**, 396–405 (2017). doi: 10.1038/nn.4493; pmid: 28166219
59. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
60. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
61. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
62. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for

macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). doi: 10.1016/j.molcel.2010.05.004; pmid: 20513432
63. H. Li *et al*., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
64. F. Ramírez *et al*., deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016). doi: 10.1093/nar/gkw257; pmid: 27079975
65. Y. Ben-Shaul, OptiMouse: A comprehensive open source program for reliable detection and analysis of mouse body and nose positions. *BMC Biol.* **15**, 41 (2017). doi: 10.1186/s12915-017-0377-3; pmid: 28506280
66. R. Srinivasan *et al*., Ca^{2+} signaling in astrocytes from *Ip3r2*$^{-/-}$ mice in brain slices and during startle responses in vivo. *Nat. Neurosci.* **18**, 708–717 (2015). doi: 10.1038/nn.4001; pmid: 25894291

ACKNOWLEDGMENTS

We thank A. Y.-S. Huang for providing training in stereotaxic injection and K. Ung for comments regarding odor behavioral assays.

Funding: This work was supported by grants from the National Institutes of Health (NIH) [National Institute of Neurological Disorders and Stroke (NINDS) R01-NS071153 to B.D., R01-AG071687 to B.D., National Institute of Mental Health (NIMH) R01-MH116900 to I.M., and National Institute on Deafness and Other Communication Diseases (NIDCD) 1K99-DC019668 to D.S.]. The Baylor College of Medicine Mass Spectrometry Proteomics Core is supported by a Dan L. Duncan Comprehensive Cancer Center NIH award (P30 CA125123), a CPRIT Core Facility Award (RP210227), and an NIH High End Instrument award (S10OD026804). The Baylor College of Medicine Genetically Engineered Rodent Model Core is funded in part by a NIH Cancer Center Grant (P30 CA125123). **Author contributions:** Conceptualization: D.S., B.D.; Methodology and data acquisition: D.S., Y.-T.C., J.W.; D.-J.C., Z.-F.L., W.K., H.-C.C., B.L., A.C., K.R., A.J.; Resources: T.-W.H., B.R.A., I.M., and B.D.; Funding acquisition: D.S., B.D.; Writing – original draft: D.S., B.D.; Review and editing: D.S., I.M., B.D. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The RNA-Seq and ChIP-Seq datasets generated in this study are available at the NCBI GEO website. RNA-Seq accession GEO: GSE229641; ChIP-Seq accession GEO: GSE229644.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade0027
Figs. S1 to S22
Tables S1 to S4
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 18 July 2022; resubmitted 30 January 2023
Accepted 28 April 2023
10.1126/science.ade0027

RESEARCH ARTICLE SUMMARY

AGING

Aging Fly Cell Atlas identifies exhaustive aging features at cellular resolution

Tzu-Chiao Lu†, Maria Brbić†, Ye-Jin Park, Tyler Jackson, Jiaye Chen, Sai Saroja Kolluru, Yanyan Qi, Nadja Sandra Katheder, Xiaoyu Tracy Cai, Seungjae Lee, Yen-Chung Chen, Niccole Auld, Chung-Yi Liang, Sophia H. Ding, Doug Welsch, Samuel D'Souza, Angela Oliveira Pisco, Robert C. Jones, Jure Leskovec, Eric C. Lai, Hugo J. Bellen, Liqun Luo, Heinrich Jasper*, Stephen R. Quake*, Hongjie Li*

INTRODUCTION: Aging is a natural process that is associated with the gradual decline of tissues in the body. This process increases the risk of developing various diseases, such as cardiovascular and neurodegenerative diseases and cancers. The study of aging has a long history, and several aging hypotheses have been proposed. Nonetheless, there are still many unanswered questions when it comes to understanding the effects of aging on the composition and maintenance of different cell types. It is also not clear whether all cell types age at the same rate or whether the transcriptome of one cell type can be used to predict age. Additionally, the genes and signaling pathways that contribute to aging in different cell types are not yet fully understood.

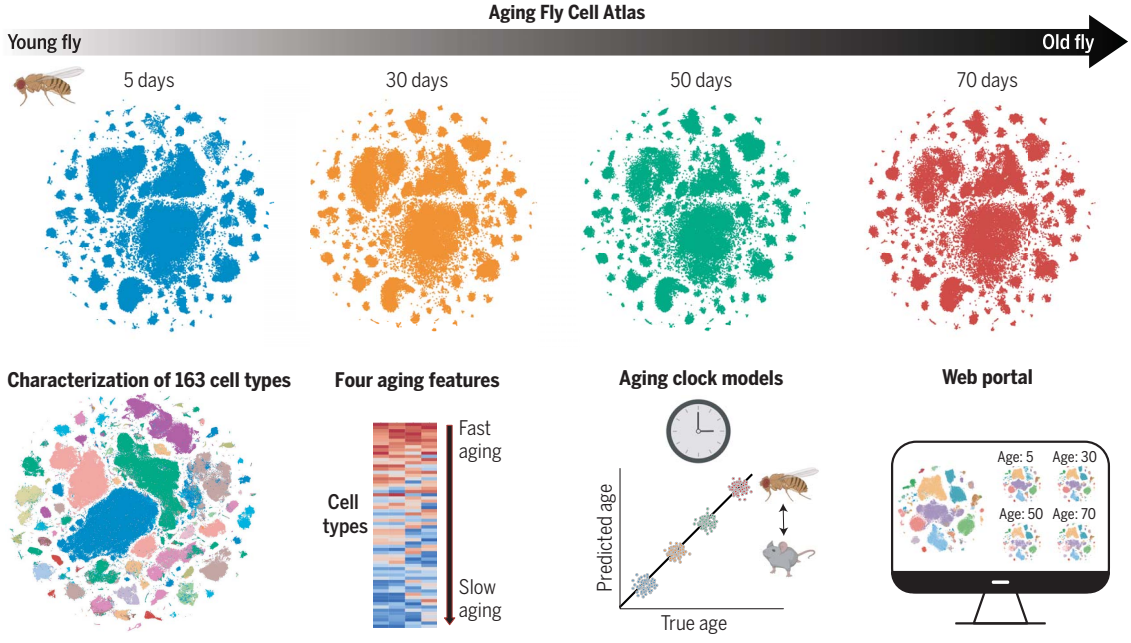
RATIONALE: *Drosophila melanogaster*, commonly known as the fruit fly, has played a vital role in advancing the fields of genetics, neurobiology, development, and aging. A large portion (~75%) of genes associated with human diseases have counterparts with functional similarity in the fly. The fly is also a useful model organism for studying the aging process, as it displays several age-related functional changes observed in humans, such as decreased motor activity, learning and memory, cardiac function, and fertility. Therefore, a comprehensive understanding of the molecular and genetic mechanisms underlying age-related decline in flies can provide valuable insights not only for aging studies in this species but also in other organisms, including humans.

RESULTS: Advancements in single-cell RNA sequencing technologies and the creation of the Fly Cell Atlas (FCA) have enabled the investigation of aging phenotypes at the single-cell level in *D. melanogaster*. Here, we present the Aging Fly Cell Atlas (AFCA), a single-nucleus transcriptomic map that characterizes changes in most tissues of male and female flies across their life span. Our analysis provides insights into age-related gene expression changes, alterations in cell composition, and common p[illegible] ways that correlate with aging. Notably, [illegible] observed an increase in fat body nuclei and elevated apoptotic markers in old indirect flight muscles, which potentially contribute to the age-related decrease in muscle nuclei. We also developed aging clock models that predict an animal's age from single-nucleus transcriptomic data. Additionally, we found variances in aging for expressed gene number and cell type identity, with different cell types being differentially affected by different aging features.

CONCLUSION: The AFCA is a valuable resource and will be of interest to the aging research community. It provides an important and timely resource for studying aging and age-related diseases. It has the potential to serve as a reference of whole-organism aging that can be used as a baseline for exploring different age-related diseases and understanding how different longevity perturbations increase life span at a cellular resolution. The ease of creating whole-fly aging atlases makes *D. melanogaster* a key model organism for examining the intersection of genetics, cell biology, and physiology going forward. We have developed a user-friendly data portal and provided access at the CELLxGENE. All resources can be accessed at https://hongjielilab.org/afca/. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: jasperh@gene.com (H.J.); steve@quake-lab.org (S.R.Q.); hongjie.li@bcm.edu (H.L.)
†These authors contributed equally to this work.
Cite this article as T.-C. Lu *et al.*, *Science* **380**, eadg0934 (2023). DOI: 10.1126/science.adg0934

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adg0934



The Aging Fly Cell Atlas. The AFCA includes single-nucleus transcriptomes of 868,000 nuclei across the *Drosophila* life span. Males and females were sequenced separately. We characterized 163 distinct cell types, developed aging clock models, and combined four aging features to rank the aging rate of different cell types. [Figure created using Biorender]

RESEARCH ARTICLE

AGING

Aging Fly Cell Atlas identifies exhaustive aging features at cellular resolution

Tzu-Chiao Lu[1,2]†, Maria Brbić[3]†, Ye-Jin Park[1,2,4,5], Tyler Jackson[1,2,6], Jiaye Chen[1,2,7], Sai Saroja Kolluru[8,9,10], Yanyan Qi[1,2], Nadja Sandra Katheder[11], Xiaoyu Tracy Cai[11], Seungjae Lee[12], Yen-Chung Chen[13], Niccole Auld[1,2,6], Chung-Yi Liang[1,2,14], Sophia H. Ding[1,2], Doug Welsch[1,2], Samuel D'Souza[10], Angela Oliveira Pisco[10], Robert C. Jones[8], Jure Leskovec[15], Eric C. Lai[12], Hugo J. Bellen[2,4,5], Liqun Luo[16], Heinrich Jasper[11]*, Stephen R. Quake[8,9,10]*, Hongjie Li[1,2]*

Aging is characterized by a decline in tissue function, but the underlying changes at cellular resolution across the organism remain unclear. Here, we present the Aging Fly Cell Atlas, a single-nucleus transcriptomic map of the whole aging *Drosophila*. We characterized 163 distinct cell types and performed an in-depth analysis of changes in tissue cell composition, gene expression, and cell identities. We further developed aging clock models to predict fly age and show that ribosomal gene expression is a conserved predictive factor for age. Combining all aging features, we find distinctive cell type–specific aging patterns. This atlas provides a valuable resource for studying fundamental principles of aging in complex organisms.

Aging is characterized by the progressive decline in tissue function across the entire body. It is a major risk factor for a wide range of diseases, including cardiovascular diseases, cancers, and neurodegenerative diseases (*1*, *2*). Aging phenotypes have been observed and described for centuries, and a number of different aging hypotheses have been proposed (*3*). However, critical questions remain largely unaddressed in complex organisms: How does aging affect cell composition and the maintenance of specific cell types? Do different cell types age at the same rate? Can we use one cell type's transcriptome to predict age? What genes and signaling pathways drive aging in different cell types?

[1]Huffington Center on Aging, Baylor College of Medicine, Houston, TX 77030, USA. [2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [3]School of Computer and Communication Sciences, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland. [4]Program in Development, Disease Models and Therapeutics, Baylor College of Medicine, Houston, TX 77030, USA. [5]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA. [6]Program in Cancer Cell Biology, Baylor College of Medicine, Houston, TX 77030, USA. [7]Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX 77030, USA. [8]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [9]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. [10]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. [11]Regenerative Medicine, Genentech, Inc., South San Francisco, CA 94080, USA. [12]Developmental Biology Program, Sloan Kettering Institute, New York, NY 10065, USA. [13]Department of Biology, New York University, New York, NY 10013, USA. [14]Institute of Biochemistry and Molecular Biology, National Yang Ming Chiao Tung University (NYCU), Taipei 112304, Taiwan. [15]Department of Computer Science, Stanford University, Stanford, CA 94305, USA. [16]Howard Hughes Medical Institute, Department of Biology, Stanford University, Stanford, CA 94305, USA.
***Corresponding author. Email: jasperh@gene.com (H.J.); steve@quake-lab.org (S.R.Q.); hongjie.li@bcm.edu (H.L.)**
†These authors contributed equally to this work.

The fruit fly, *Drosophila melanogaster*, has been at the basis of many key discoveries in genetics, neurobiology, development, and aging. About 75% of human disease–associated genes have functional homologs in the fly (*4*, *5*). Many of the age-related functional changes in humans are also observed in flies, including a decline in motor activity, learning and memory, cardiac function, and fertility (*6*). Hence, a proper description of the molecular and genetic basis of the age-related decline in flies should provide an important resource for aging studies not only in flies but also in other organisms.

The recent development of single-cell RNA sequencing (scRNA-seq) technologies and the establishment of the Fly Cell Atlas (FCA) (*7*), a single-nucleus transcriptomic atlas of *Drosophila* at the age of 5 days (5d hereafter), have made it possible to investigate aging phenotypes across the whole organism at single-cell resolution. Here, we present the Aging Fly Cell Atlas (AFCA), a single-nucleus transcriptomic map describing age-related changes in most tissues, including those that differ by sex. We performed an in-depth analysis of age-related gene expression and cell composition changes across the entire fly, as well as cell type–specific and common pathways that correlate with aging. Notably, we observed a significant increase of fat body nuclei and a drastic decrease in muscle nuclei with age. Furthermore, we developed aging clock models that predict the animal's biological age from the single-nucleus transcriptomic data. In addition, we found aging variances in expressed gene number and in cell type identity. Our analysis revealed that different cell types are differentially affected by different aging features. The AFCA provides a valuable resource for the *Drosophila* and aging communities as a reference to study aging and age-related diseases and to evaluate the success of antiaging regimens. We developed a website portal for data visualization and custom analyses and made data available at the CELLxGENE portal (figs. S1 and S2). All resources can be accessed at https://hongjielilab.org/afca.

Results

Single-nucleus transcriptomes of the entire fly at different ages

To generate the AFCA, we applied the same single-nucleus RNA sequencing (snRNA-seq) pipeline used for the FCA (5d adults) (*7*) and profiled the whole head and body at three additional ages (30d, 50d, and 70d). These time points were chosen to cover the life-span trajectory of a fly (Fig. 1A) and beyond, up to 70d, the estimated equivalent of 80- to 90-year-old humans. Male and female flies were sequenced separately, allowing the investigation of sexual dimorphism during aging (Fig. 1B). To achieve the most reliable analyses of aging features, we performed preprocessing of this newly generated aging data similarly to that done on the young FCA data (fig. S1A). Consistent with a previous scRNA-seq study of the aging fly brain (*8*), we found that young and old cells have a similar distribution in the *t*-distributed stochastic neighbor embedding (tSNE) space, suggesting that the whole organism largely maintains its cell types during aging (Fig. 1C). Overall, we obtained >868,000 nuclei covering all 17 broad cell type classes (Fig. 1, D and E). The detected numbers of expressed genes and unique molecular identifiers (UMIs) were largely consistent across different ages (fig. S3). The most abundant cell classes were neurons, epithelial cells, muscle cells, and fat cells. Next, we annotated those broad cell classes into detailed cell types.

AFCA cell type annotation and resource for studying cell type–specific aging

Because cell type–specific aging analysis largely depends on accurate cell type annotation, we took multiple approaches to ensure that our AFCA data were annotated with high confidence. We first co-clustered our aging data with the annotated FCA data, either from the head or body or from individual tissues. Then we transferred AFCA annotations using both a cluster-centered method and a supervised machine learning–based method (figs. S4 and S5). Overall, we found that these two approaches agree well, with ~80% overlap (Fig. 2A). The discrepancies of the nonoverlapping annotations were mostly due to uncharacterized cell types in the FCA data or cell types with aging differences. Next, we manually validated each annotation using cell type–specific markers. Marker validation confirmed the accuracy of our automatic annotation procedure, with a few exceptions, such as indirect flight muscles because of age-related loss of specific markers

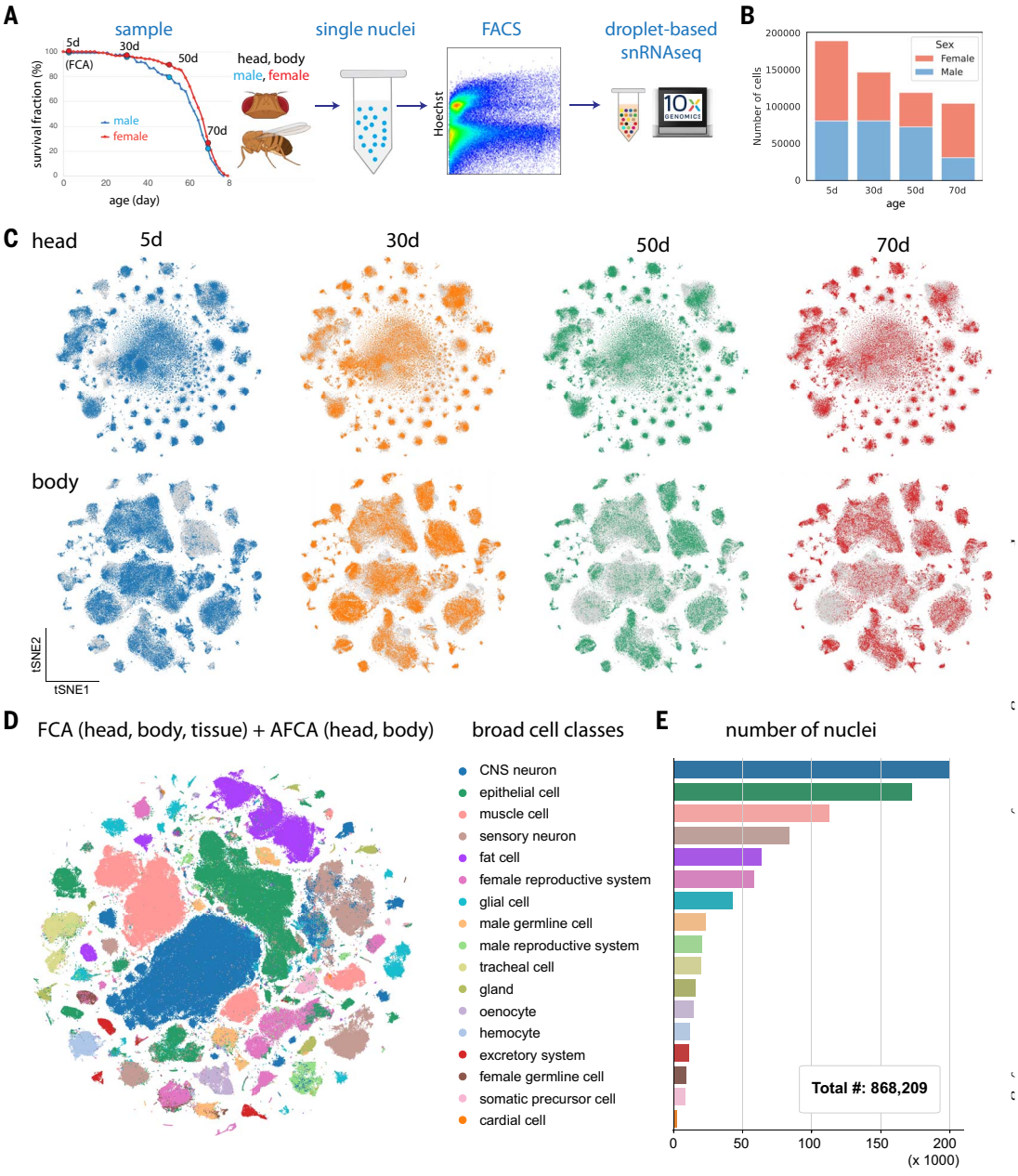


Fig. 1. Overview of the AFCA. (**A**) Flowchart of the snRNA-seq experiment. Flies were collected at 30, 50, and 70 days. The heads and bodies of males and females were processed separately. 5d samples were from the FCA. FACS, fluorescence-activated cell sorting. (**B**) Number of nuclei collected from different ages and sexes. (**C**) tSNE visualizations of the head and body samples from different ages. (**D**) tSNE visualizations showing broad cell classes of datasets integrated across different time points. (**E**) Number of nuclei for each broad cell class shown in (D).

(fig. S6, A and B) and gut cell types because of the high similarity between intestinal stem cells and renal stem cells (fig. S6C). These cell types were then manually added and corrected (figs. S5 and S7). Overall, we characterized 163 distinct cell types, including 91 cell types from the aging head and 72 cell types from the aging body (Fig. 2B and figs. S8 and S9).

For complex tissues such as the brain, more cell types can emerge when substantially more cells are sequenced (*9*–*11*). Indeed, 17 additional neuronal cell types emerged after combining young and old head data (Fig. 2C). Among them, four types are GABAergic neurons (*Gad1*+), two are glutamatergic neurons (*VGlut*+), and the remaining 11 are cholinergic neurons (*VAChT*+) (fig. S10).

Next, we assessed the reliability of AFCA data for investigating age-related changes in specific cell types. As a case study, we focused on the fly gut, for which the somatic stem cell lineage and its aging have been well characterized (*12*). In a healthy young fly gut, intestinal stem cells (ISCs) maintain gut homeostasis through proper proliferation and differentiation. In old flies, ISCs exhibit a high proliferation rate, and their daughter cells, enteroblasts (EBs), do not properly differentiate into mature enterocytes, leading to a dysplasia phenotype (*13*–*15*). We first extracted six major gut cell types and performed pseudotime trajectory analysis (*16*, *17*). There was a significant increase of ISCs and EBs along with a decrease of fully differentiated enterocytes, consistent with previous in vivo studies (Fig. 2D and fig. S11) (*18*), and we could identify genes that showed different dynamic patterns between young and old flies (fig. S11E). As demonstrated by this case study, the detailed annotations in our AFCA data offer a valuable resource to explore cell type– and tissue-specific aging signatures.

Cell composition changes during aging

In complex organisms, aging can affect cellular composition in different ways, such as changing stem cell proliferation or differentiation processes, altering cell identity, or inducing cell death. We assessed whether and how aging affects cellular composition across the whole fly. Note that our measurements are based on nuclei composition. Given that the

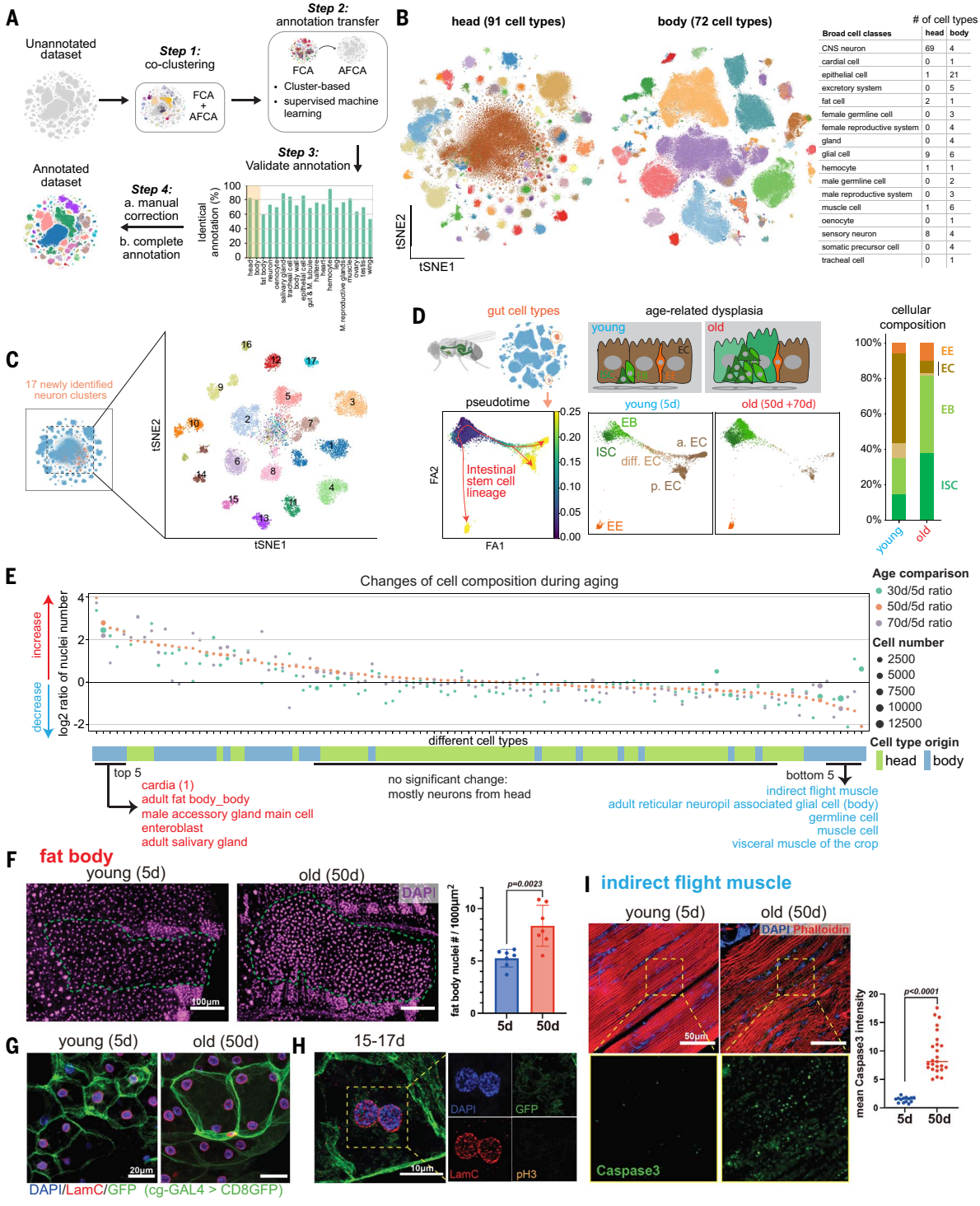


Fig. 2. AFCA resource and changes of cell composition during aging. (**A**) Flowchart of transferring annotations from the FCA to the AFCA. (**B**) Cell types annotated in the AFCA head and body shown on tSNE. The number of annotated cell types corresponding to the broad cell classes is shown in the table. (**C**) Identification of 17 additional neuronal clusters after combining AFCA and FCA head data. (**D**) Pseudotime and cellular composition of ISC and ISC-differentiated cell types. ISC, intestinal stem cell; EB, enteroblast; EC, enterocyte; EE, enteroendocrine cell; a. EC, anterior EC; p. EC, posterior EC; diff. EC, differentiating EC. (**E**) Changes of cellular composition during aging. Each dot represents one cell type. Each color compares one aged sample and the 5d sample. Dot sizes reflect the nuclear numbers of the corresponding cell type from the aged population. Tissue origins are indicated. (**F**) Comparison of the number of nuclei of the fat body from young and old flies. Nuclei are stained by 4′,6-diamidino-2-phenylindole (DAPI) and counted in each fly. The nuclear number is significantly increased in the 50d population (*t*-test, 50d versus 5d, $P = 0.0023$). Error bar, standard deviation (SD). (**G**) Representative confocal images showing nuclei in young and old fat body cells. The membrane is labeled by *cg-GAL4 > UAS-CD8GFP*. Nuclei are stained by DAPI and the LamC antibody. (**H**) Fat body cells with segregating nuclei stained by pH3, DAPI, LamC, and GFP. (**I**) Indirect flight muscle stained with cleaved-Caspase3 antibody, DAPI, and phalloidin. Cleaved-Caspase3 signals are significantly increased in the aged population (*t*-test, 50d versus 5d, $P < 0.0001$). Median numbers are indicated.

nuclei are extracted from the whole head and body with minimal sampling bias, the ratio for each cell type in our sequencing data should largely reflect their composition in vivo.

To perform reliable analyses, we focused on cell types that have >500 nuclei in total (112 cell types after filtering). We then calculated the composition changes by comparing three older ages (30d, 50d, and 70d) to young flies (5d). The top five increased cell types are cardia cells (proventriculus from the gut), fat body cells, male accessory gland main cells, EBs, and adult salivary gland cells (Fig. 2E). Comparing two consecutive ages showed similar results (fig. S12). Age-related increases of proventriculus cells, male accessory gland main cells, and EBs have been reported previously (*19*–*21*), confirming the quality of our data and analysis. Fat body cells are one of the most abundant cell types in *Drosophila*. They are polyploid, filled with lipid droplets, and tightly attached to the abdominal cuticle. These features make it difficult to isolate them or to compare their compositions using traditional methods. To our knowledge, an increase of fat body nuclei in old flies has not been reported. We were able to validate this observation (Fig. 2F).

Fat body cells are postmitotic cells, and no adult stem cells or progenitors have been reported for regeneration (*22*). To examine why fat body nuclei were increased in old flies, we first checked the number of nuclei within single cells using a fat body–specific GAL4 driving cell membrane green fluorescent protein (GFP). Many aged fat body cells exhibited an increase in cell size and contained multiple nuclei per cell (Fig. 2G). The multinucleated phenotype can potentially be caused by cell membrane fusion, as reported in other cell types (*23*), but it cannot explain the increase in the number of nuclei (Fig. 2, E and F). It has been reported that polyploid enterocytes from the fly gut can undergo nuclear cleavage without mitosis (a process called amitosis) (*24*). To test this possibility, we performed immunohistochemistry to detect the nuclear lamina protein, Lamin C (LamC), and a mitosis marker, Phospho-Histone H3 (pH3). We did not detect any mitotic events from >60 flies across different ages, but we did observe many cases where two nuclei were localized very close to each other and were negative for the mitotic marker (Fig. 2H). Three-dimensional reconstruction of confocal images confirmed that these nuclei were present in the same cell without a separating cell membrane (movies S1 to S3). Such events were captured across different ages (fig. S13). Together, these data suggest that fat body cells undergo nuclear division without cytokinesis across different ages, leading to multinucleated cells and an increase in the number of nuclei in older flies.

Among the top five decreased cell types are three types of muscle—indirect flight muscle, visceral muscle, and other muscle cells (mostly skeletal muscle) (Fig. 2E). Loss of muscle mass and strength, known as sarcopenia, is a conserved aging phenotype across different mammals, including humans (*25*). Our in vivo staining data confirmed the age-related degeneration of indirect flight muscles as well as a significant increase of an apoptosis marker, Caspase3, in old flies (Fig. 2I and fig. S14), consistent with a previous study (*26*). Germline cells also showed a significant decrease (Fig. 2E and fig. S12), presumably contributing to the decline of fecundity in old flies. Most cell types from the head, mostly neurons, showed minimal cellular composition changes (Fig. 2E and fig. S12).

Differentially expressed genes

Altered gene expression is another consequence of aging (*27*). To assess such changes, we performed differentially expressed gene (DEG) analysis between young and old flies and ranked cell types on the basis of the number of DEGs (Fig. 3A and fig. S15A). Again, we focused on 112 cell types with >500 cells for a reliable analysis. In the body, the cell type with the highest number of DEGs was the fat body, while the most affected cell type in the head was the outer photoreceptor (Fig. 3A).

To explore the dynamics of cell type–specific changes, we further examined the time window during which cell types change the most by computing DEG numbers between two neighboring ages (fig. S15B) and normalizing their ratios (Fig. 3B and fig. S16). This analysis revealed several notable insights (Fig. 3, B and C). Specifically, ~80% of the cell types showed major changes (>50% of DEGs) during the first time window, suggesting that 30-day-old flies have captured a large portion of age-related gene changes. Some cell types, such as the male accessory gland, showed minimal changes in the last time window, indicating that these cell types reach their maximum transcriptomic changes around 50d. However, 5.3% of cell types, including intestinal stem cells and cardia cells, showed drastic changes (>50% of DEGs) in the last time window, suggesting that they age at a slower rate during the first 50d. Other cell types such as outer photoreceptors and fat body cells showed a steady change. Hence, this analysis indicates that different cell types age at different rates and exhibit distinctive patterns of gene expression changes. We compared DEGs from the AFCA with those from the aging fly brain study (*8*) and found that they are well correlated (fig. S17).

By integrating cellular composition changes and DEG analysis, we determined which cell types were affected by those two parameters (Fig. 3D). Significantly affected cell types ("outliers") fell into four categories: (i) cells showing changes for both, such as fat body cells from the body, pericerebral adult fat mass (fat cells from the head), and male accessory gland main cells; (ii) cells showing high DEGs but minimal composition changes, such as outer photoreceptors and cone cells, consistent with a previous study reporting that age-related fly visual decline is not attributable to the loss of photoreceptors (*28*); (iii) cells showing decreased nuclear number and a moderate number of DEGs, such as indirect flight muscles; and (iv) cells showing increased nuclear number but minimal DEGs, such as cardia cells.

Next, we performed sex-related analysis. We first observed that female marker yolk protein genes (*Yp1*, *Yp2*, and *Yp3*) showed a significant decrease during aging in most female cell types, while the male markers (*roX1* and *roX2*) maintained high expression levels in most male cells (fig. S18, A and B). In addition, some genes, such as *Lsd-2* and *CG45050*, showed different trends between males and females with age (fig. S18C), suggesting that aging affects male and female cells differently. It has been shown previously that different cell types exhibit different DEGs in males and females at a young age (*7*). We next checked how these numbers change during aging. Generally, if one cell type showed a high or low number of DEGs at a young age, it maintained that high or low number during aging (fig. S19A). However, some cell types showed age-specific sex differences. For example, three cell types from the head—pericerebral adult fat mass, skeletal muscle, and hemocyte—all showed few DEGs in young flies but many DEGs in old flies (fig. S19, B and C). In contrast, three glial populations showed a significant decrease of DEGs with age. We also checked how sex affects the DEG number and cell composition and found that these two features highly correlate between male and female flies (fig. S19, D to G).

Analysis of gene pathways

Next, we investigated which genes and pathways are enriched in DEGs. Gene ontology (GO) analysis was performed for both up- and down-regulated genes compared between the 50d and 5d dataset (fig. S20). Most GO terms were cell type–specific (fig. S20A); only 20% of GO terms were shared by more than five cell types for down-regulated genes and 40% for up-regulated genes. We found that one GO term from down-regulated genes, "cytoplasmic translation," was shared by almost all cell types (fig. S20C). Cytoplasmic translation refers to the ribosome-mediated process for protein synthesis. Many transcripts of genes encoding ribosomal proteins (RPs) were decreased across cell types, consistent with previous studies (*29*). There were no globally shared GO terms among the up-regulated genes (fig. S20B). Instead, shared terms were restricted to specific groups of cells, for example, signal transduction seen in neuronal types and protein phosphorylation enriched in different non-neuronal cells.

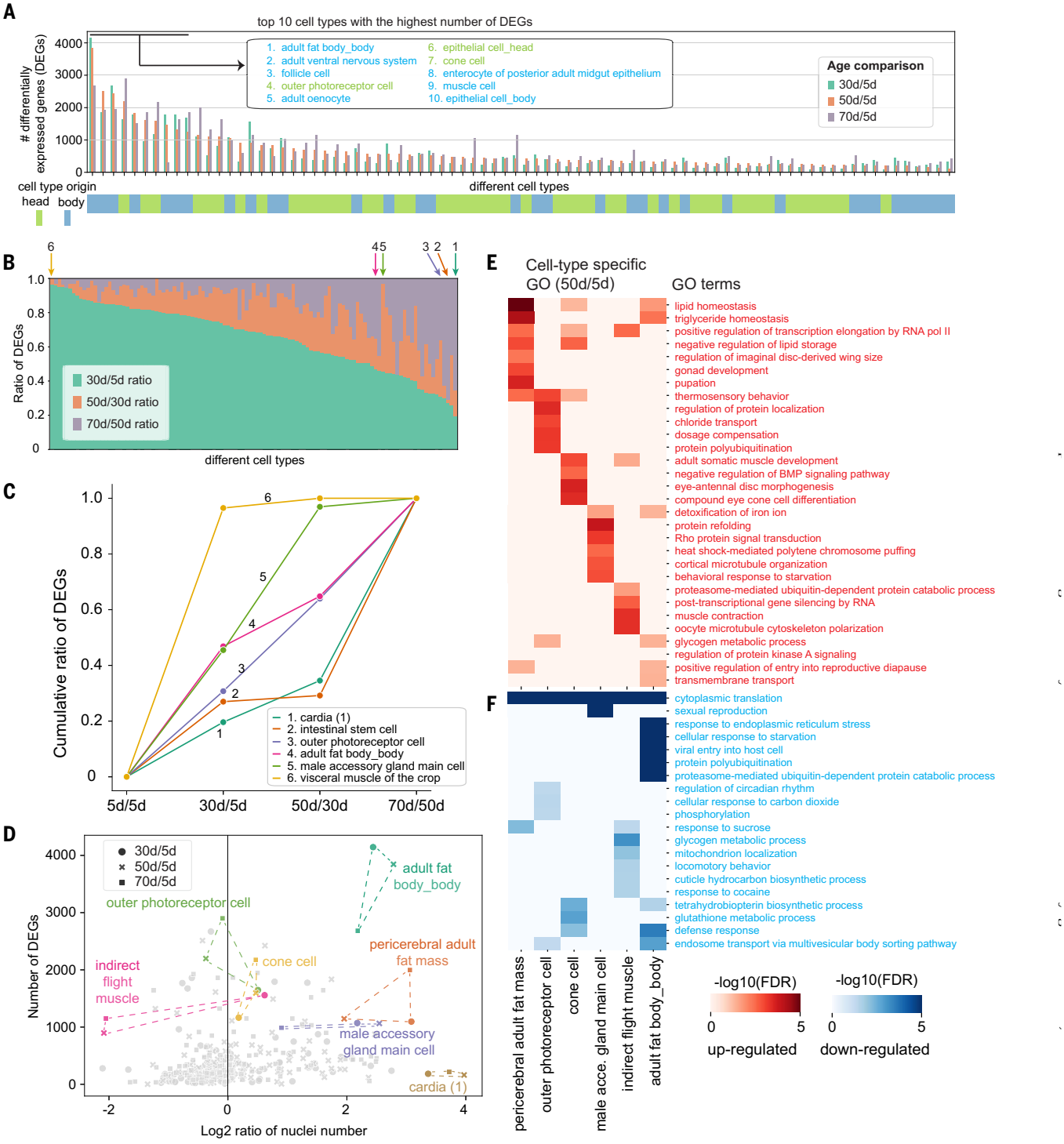


Fig. 3. Differentially expressed genes. (**A**) Number of DEGs from different cell types. Each age group is compared with the 5d population. Each line shows the number of DEGs from the indicated age comparison. Cell types are ranked by DEG numbers from high to low (50d versus 5d). The top 10 cell types are indicated. (**B**) Ratio of DEGs from each age comparison. Arrows point out representative cell types that are further compared in the next panel. (**C**) Cumulative ratios of DEGs from cell types indicated in (B). (**D**) Combination of DEG number and change of nuclear number illustrates different aging patterns. (**E** and **F**) Top five cell type–specific GO terms from selected cell types. (E) GO terms enriched in the selected cell types on the basis of up-regulated DEGs. (F) GO terms enriched in the selected cell types on the basis of down-regulated DEGs.

We next focused on GO terms enriched within a few cell types to understand the cell type–specific regulations (Fig. 3, E and F). Fat body cells from body and head (pericerebral adult fat mass) shared metabolic-related GO terms from up-regulated genes, such as lipid homeostasis and triglyceride homeostasis (Fig. 3E), reflecting common metabolic changes in these tissues. For indirect flight muscles, a reduction of locomotor behavior was observed and is likely to be caused by muscle degeneration (Fig. 3F) (*26*). Also, reproduction-related genes were strongly

decreased in male accessory gland main cells (Fig. 3F), consistent with the decline in reproductive ability in males (*30*). To summarize, we observed that genes involved in "cytoplasmic translation" are commonly decreased during aging in many cell types, whereas most other GO terms showed cell type–specific patterns.

Aging clock to predict the biological age

To predict the biological age of an animal or human, a number of different aging clock models have been recently developed using epigenetic markers and transcriptomic data (*8*, *31*, *32*). We investigated whether our snRNA-seq data can be used to develop aging clocks. To perform a more accurate prediction, we focused on cell types that have >200 cells at each age point (64 cell types). For each cell type, we trained a regression model (*33*) to predict age. We measured predictive performance using the coefficient of determination, R^2 (Fig. 4A). The average performance across all cell types was high (average R^2 = 0.79 for body and 0.84 for head; fig. S21).

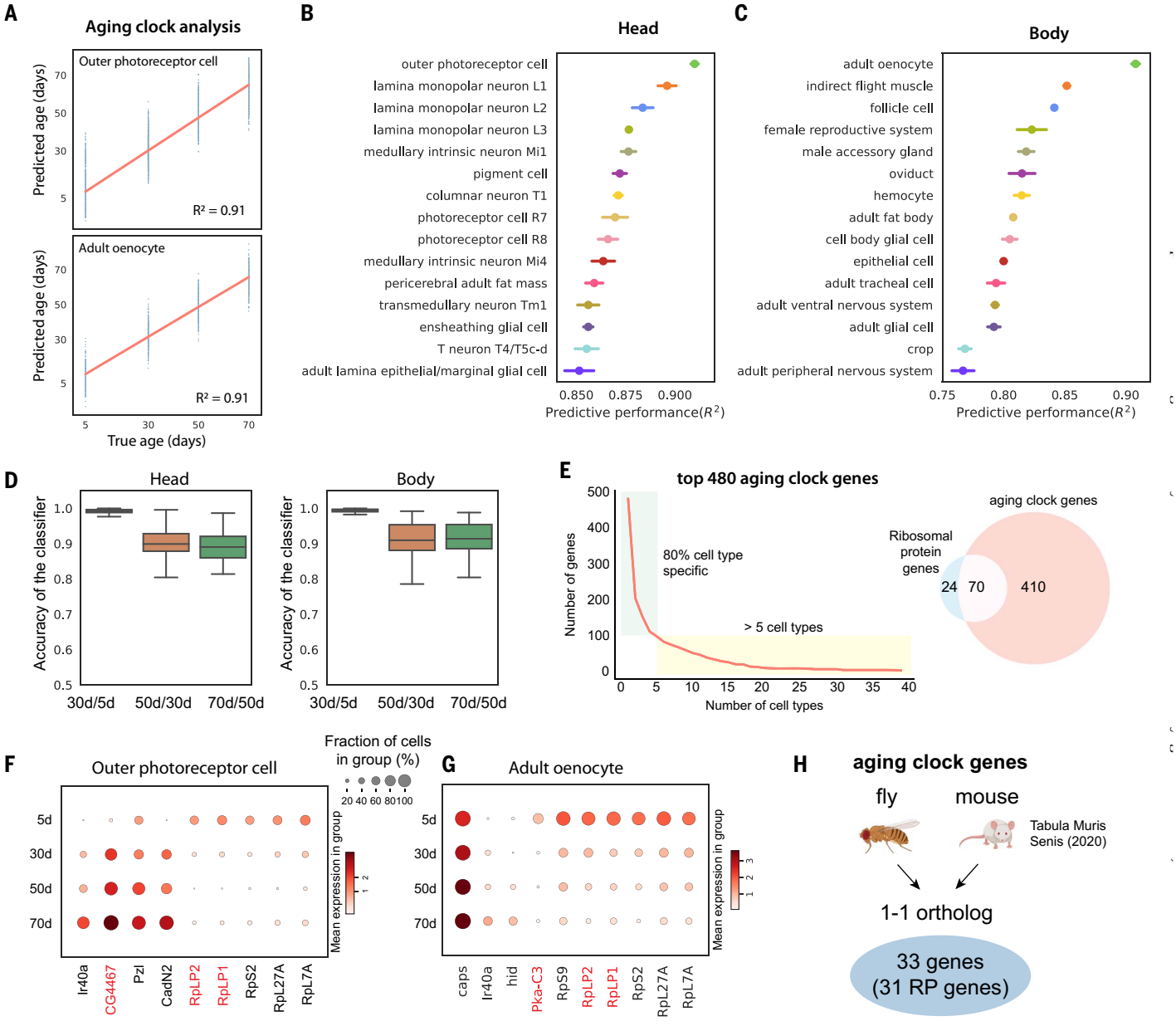


Fig. 4. Aging clock analysis. (**A**) Example of aging clocks for outer photoreceptor cells and adult oenocytes. The red line is the fitted regression line. Blue dots represent individual predictions wherein each dot corresponds to one cell. We measure performance as the proportion of the variance for an age variable that is explained by transcriptome (R^2). (**B** and **C**) Predictive performance of cell type–specific aging clocks for head and body cell types. Fifteen cell types with the highest scores are shown. Error bars are estimated as a SD over five runs. (**D**) Accuracy of logistic regression models trained to distinguish transcriptome between two consecutive time points. Boxplots show the distribution across head cell types (left) and body cell types (right). (**E**) Number of aging clock genes as a function of the number of cell types (left). Eighty percent of genes appear in fewer than five cell types. Out of 480 genes identified as aging clock genes, 70 encode RP genes (right). (**F** and **G**) Examples of aging clock genes for outer photoreceptor cell (F) and adult oenocyte (G). Cell type–specific genes that appear in fewer than five cell types are shown in black, while genes that appear in at least five cell types are shown in red. (**H**) Aging clock genes identified in flies and mice. Thirty-three genes are one-to-one orthologs between the two species, and 31 of these are RP genes.

Outer photoreceptor cells and oenocytes showed the highest predictive scores in head and body, respectively (Fig. 4, B and C). *Drosophila* oenocytes perform liver-like functions including lipid storage and metabolomics functions, similar to fat body cells, and also produce important cuticular pheromones (*34*). We confirmed that high scores were not caused by higher nuclear numbers (fig. S22A). We further asked whether there is a difference in predictive performance between different time windows and found that the largest transcriptomic differences are present between the first two ages (Fig. 4D).

Next, we focused on identifying aging clock genes that are used to predict the age. First, we found that most aging clock genes are used in a cell type–specific manner (Fig. 4E). We found that 70 out of 94 fly RP genes were identified as aging clock genes, showing age-related reductions in different cell types (Fig. 4, E to G). These data are consistent with our previous GO analysis where "cytoplasmic translation" is reduced in almost all cell types (fig. S20C). Decreased protein translation, which can be caused by the reduction of RPs, is a prevalent feature of aging (*35*). To determine what transcription factors might regulate the expression of RPs, we used the regulon information from the FCA (*7*, *36*) and identified several transcription factors regulating RP genes (fig. S23). Our data suggest that the reduction of ribosomal expression contributes to the age-related decrease of protein synthesis. To further investigate the relationship of aging clock genes across species, we identified aging clock genes from the Mouse Aging Cell Atlas (*27*). Among 33 overlapping aging clock genes from the fly and mouse, 31 genes encode RPs (Fig. 4H and figs. S24 and S25).

Comprehensive aging features

Through the above analyses, we noticed that different cell types are sensitive to different aging features. To gain a better understanding of cell type–specific aging, we investigated more aging features, including expressed gene number or transcript number (measured by UMI) changes and decline of cell identity (Fig. 5A).

Similar to the aging clock analysis, we only focused on cell types with at least 200 cells from each age. Gene and UMI numbers were previously found to decrease in the old fly brain (*8*). Consistent with previous observations, we observed decreases in gene and UMI number in central nervous system (CNS) neurons during aging (fig. S26A). To understand whether such a reduction is a general aging feature or not, we examined all other cell types. The overall trend of gene and UMI numbers was largely consistent (fig. S26B). We found that ~80% of cell types exhibited a decrease in expressed gene numbers, but 20% exhibited an increase, suggesting that aging affects expressed gene numbers in a cell type–specific manner (Fig. 5B). Head hemocytes and pericerebral adult fat mass had the greatest increase in expressed gene numbers, while oenocytes, ventral nervous system cells, and fat body cells had the greatest decrease (Fig. 5C and fig. S26C). We confirmed that this was not caused by sequencing depth (fig. S26D). Even though head and body fat body cells both increased in nuclear number (Fig. 3D), their expressed gene numbers showed the opposite trends.

Loss of cell type identity has been shown to occur during aging for certain cell types (*12*, *37*). However, how cell identity changes across the entire organism during aging remains uncharacterized. To assess whether established cellular gene expression programs that define cell identity change during aging, we developed a measurement to combine two ratios—loss of original markers and gain of new markers—by comparing old populations with young populations (Fig. 5D, left panel). We then ranked cell types by their cell identity decline score (Fig. 5D, right panel; fig. S27). The Neuroligin 1 gene, *Nlg1*, which was used as a marker gene to annotate the indirect flight muscle, showed a drastic decrease with age, as did a number of other young marker genes (Fig. 5D and fig. S6A). Meanwhile, many other gene transcripts, such as *Chchd2*, began to appear in this cell type with aging. Other than the indirect flight muscle, pericerebral adult fat mass and epithelial cells from the head were also found to exhibit a large decrease in cell type identity (fig. S28).

Correlation and ranking of aging features

We examined four different aging features: cell composition changes, DEGs, change of expressed gene numbers, and cell identity decline. To understand the overall correlation between them, we ranked each feature from the least age-related change to the most. After integrating four aging features, their correlations were compared using Spearman's correlation and clustered using the correlation scores (Fig. 5E). Among those features, DEG number and decline of cell identity were highly correlated, suggesting that cell types with large numbers of DEGs would usually fluctuate in the expression of marker genes. On the other hand, changes of nuclear number and expressed gene number were more correlated with each other.

Next, we summed different feature ranks and sorted cell types by the total rank sums, with a higher rank indicating a more "aged" cell type (Fig. 5F and fig. S29). Notably, the top three cell types include three adipose cell types—oenocytes, fat body cells, and pericerebral adult fat mass—suggesting that those cells age faster than other cell types. Following them in the ranking are male accessory gland main cells, indirect flight muscle, and enteroblasts. Generally, neurons and glia from the nervous system age slower than other cell types (Fig. 5F). In summary, our analysis provided the first exhaustive analysis of different aging features and revealed the aging rates of different cell types across the entire organism.

Discussion

The cell atlas approach is emerging as a powerful tool to systematically study aging in different organisms, including worms (*38*, *39*), mice (*27*), and humans (*40*). A recent study performed cross-species analysis with scRNA-seq data from three species, including *D. melanogaster* (*41*). Our AFCA provides a complementary dataset and thorough analyses for studying aging features across the whole organism.

One interesting observation is the increase of fat body nuclei during aging. *Drosophila* fat body is a liver-like tissue that stores fat and serves as a detoxifying and immune-responsive organ. Adult fat body cells are postmitotic polyploid cells without a stem cell or progenitor population (*22*). How do they increase their nuclear number? Our observations suggest that these polyploid cells increase their number of nuclei by nuclear cleavage without cytokinesis, forming multinucleated cells. To complete karyokinesis, the nuclear envelope needs to be reassembled or reorganized. Fat body cells have been shown to undergo a decrease in nuclear envelope integrity as a result of the loss of Lamin B during aging (*42*), suggesting that the loss of nuclear envelope integrity may be associated with the multinucleation phenomenon. On the other hand, multinucleated cells have also been observed in the fly male accessory gland (*43*, *44*) and subperineurial glial cells (*45*), suggesting that multinucleated cells play specific roles in *Drosophila*. Multinucleated cells have been observed in species other than *Drosophila*, including mushrooms (*46*), plants (*47*), and the liver cells of humans and mice (*48*, *49*). Understanding the formation and regulation of multinucleated cells in the aging organism may provide new insights into an evolutionarily conserved phenomenon, as well as into the potential roles of multinucleated cells in age-related diseases.

In this study, we focused on four different aging features. Although these features cover several key aspects of age-related changes, the picture is incomplete. Additional aging measurements may reveal more specific aging patterns. For example, we investigated the change of alternative polyadenylation (APA) patterns, which can reflect short or long 3′ untranslated region usage for different isoforms (*50*), and found that neuronal extended 3′ isoforms were progressively depleted during aging. This phenotype is more obvious at 70d and more pronounced in females than in males (fig. S30). These data imply a global change in posttranscriptional regulation in aging neurons. Thus, future analysis may help elucidate additional aging patterns.

One major goal of this study was to characterize how different cell types age across

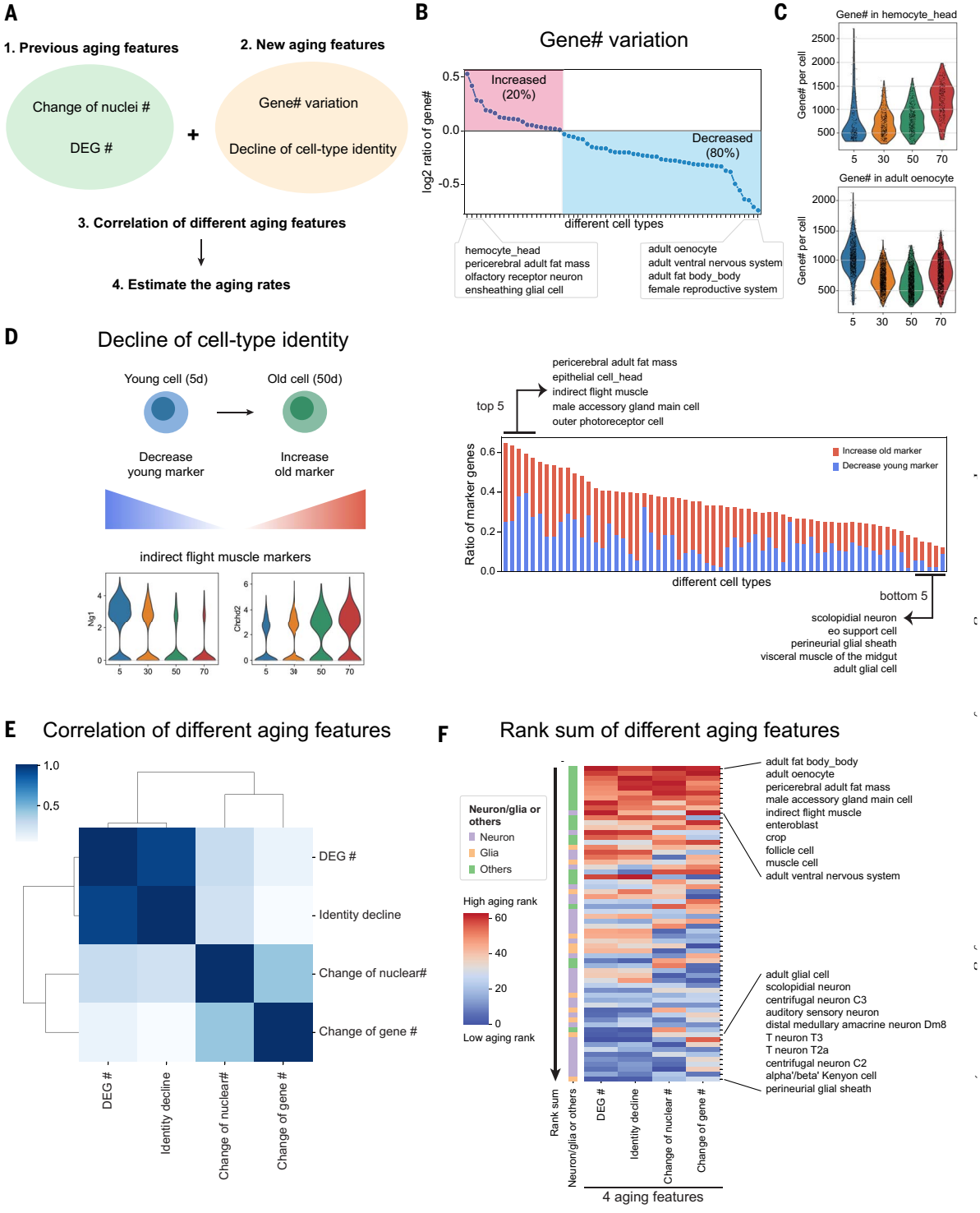


Fig. 5. Systematic comparisons of different aging features. (**A**) Flowchart for the comparison of different aging features. (**B**) Expressed gene numbers per cell from each cell type are compared between 50d and 5d flies. The red block shows cell types with increased expressed gene numbers, while the blue block includes cell types with decreased ones. (**C**) Two cell types, hemocyte from the head and oenocyte, have the highest increase and decrease, respectively, of expressed gene numbers per cell. (**D**) Decline of cell identity during aging. The left panel illustrates two different mechanisms of decreasing cell identity. The right panel shows the ratio of marker genes decreasing cell identity. Each line in the right panel represents one cell type. (**E**) Spearman's correlation of different aging features. (**F**) Rank sums of different aging features. The heatmap shows the overall rank sum scores from different cell types. High aging ranks are shown in red, while low aging ranks are shown in blue. Neuron- or glia-related cell types are indicated beside the heatmap.

the organism. Our analysis using different aging features provides several key insights. First, different cell types have distinct aging patterns. For example, the ventral nervous system showed high ranks for three aging features but a low rank for the change of nuclear number, while scolopidia neurons showed low ranks for three aging features but high ranks for the change in expressed gene number (Fig. 5F). This observation is not unexpected, considering that each cell type carries a specific function. Second, we observed a divergence in the contribution of individual cell types to a tissue's aging. For example, in the female reproductive system, follicle cells were ranked very high (eighth of 64 cell types), but germline cells were ranked near the bottom (41st of 64 cell types) (fig. S29A). This indicates that age-related declines of female fertility may be due to the aging of follicle cells. Third, the top-ranked cell types include all adipose cells. This is surprising, and we do not fully understand the underlying mechanisms. It may be linked to the fact that these cell types play multiple critical roles in different physiological conditions, such as lipid storage and metabolism,

immune responses, and interorgan communication with muscles and gut (*22*, *51*, *52*).

Materials and methods summary

This study involved collecting fly head and body samples from wild-type F1 flies from the cross between female W^{1118} and male Oregon R (OreR). Samples from different ages were dissected and stored at −80°C after flash-freezing with liquid nitrogen. The snRNA-seq was prepared using the FCA protocol (*7*), and each age group had 12 samples with six females and six males. The libraries were sequenced using NovaSeq 6000 (Illumina).

FASTQ files were filtered for index-hooping reads using 10x Genomics index-hopping-filter software. The Cell Ranger (version 4.0.0) index was built using the *D. melanogaster* genome (FlyBase r6.31) and the pre-mRNA GTF established by the FCA. Cell Ranger Count was used to estimate the nuclei number and gene expression from each nucleus. Nuclei from different sexes and ages of flies were integrated using Harmony, and cell type annotations from FCA samples were transferred to AFCA samples using a cluster-centered or machine learning–based method, followed by manual corrections.

For cell composition, age-specific ratios were obtained by dividing the number of nuclei in each cell type from one specific age by the total number of nuclei in the corresponding age. The relative ratios were then compared between the young (5d) and old populations. Wilcoxon rank sum tests were used to compare gene expression between different ages or sexes, and genes with a false discovery rate (FDR) of <0.05 were considered to be differentially expressed. The differential expression of the top 200 marker genes from 5d and 50d samples was used to estimate declines of cell type identity. Aging rates of each cell type were calculated by integrating different aging features and ranking the sums of aging ranks from high to low.

REFERENCES AND NOTES

1. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging. *Cell* **153**, 1194–1217 (2013). doi: 10.1016/j.cell.2013.05.039; pmid: 23746838
2. T. Niccoli, L. Partridge, Ageing as a risk factor for disease. *Curr. Biol.* **22**, R741–R752 (2012). doi: 10.1016/j.cub.2012.07.024; pmid: 22975005
3. D. Harman, The free radical theory of aging. *Antioxid. Redox Signal.* **5**, 557–561 (2003). doi: 10.1089/152308603770310202; pmid: 14580310
4. H. J. Bellen, C. Tong, H. Tsuda, 100 years of *Drosophila* research and its impact on vertebrate neuroscience: A history lesson for the future. *Nat. Rev. Neurosci.* **11**, 514–522 (2010). doi: 10.1038/nrn2839; pmid: 20383202
5. D. Baldridge *et al*., Model organisms contribute to diagnosis and discovery in the undiagnosed diseases network: Current state and a future vision. *Orphanet J. Rare Dis.* **16**, 206 (2021). doi: 10.1186/s13023-021-01839-9; pmid: 33962631
6. M. S. Grotewiel, I. Martin, P. Bhandari, E. Cook-Wiens, Functional senescence in *Drosophila melanogaster*. *Ageing Res. Rev.* **4**, 372–397 (2005). doi: 10.1016/j.arr.2005.04.001; pmid: 16024299
7. H. Li *et al*., Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022). doi: 10.1126/science.abk2432; pmid: 35239393
8. K. Davie *et al*., A single-cell transcriptome atlas of the aging *Drosophila* brain. *Cell* **174**, 982–998.e20 (2018). doi: 10.1016/j.cell.2018.05.057; pmid: 29909982
9. J. Janssens *et al*., Decoding gene regulation in the fly brain. *Nature* **601**, 630–636 (2022). doi: 10.1038/s41586-021-04262-z; pmid: 34987221
10. H. Zeng, What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022). doi: 10.1016/j.cell.2022.06.031; pmid: 35868277
11. G. La Manno *et al*., Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021). doi: 10.1038/s41586-021-03775-x; pmid: 34321664
12. H. Li, H. Jasper, Gastrointestinal stem cells in health and disease: From flies to humans. *Dis. Model. Mech.* **9**, 487–499 (2016). doi: 10.1242/dmm.024232; pmid: 27112333
13. H. Jasper, Intestinal stem cell aging: origins and interventions. *Annu. Rev. Physiol.* **82**, 203–226 (2020). doi: 10.1146/annurev-physiol-021119-034359; pmid: 31610128
14. C. A. Micchelli, N. Perrimon, Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* **439**, 475–479 (2006). doi: 10.1038/nature04371; pmid: 16340959
15. B. Ohlstein, A. Spradling, The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature* **439**, 470–474 (2006). doi: 10.1038/nature04333; pmid: 16340960
16. F. A. Wolf *et al*., PAGA: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019). doi: 10.1186/s13059-019-1663-x; pmid: 30890159
17. M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE* **9**, e98679 (2014). doi: 10.1371/journal.pone.0098679; pmid: 24914678
18. I. A. Rodriguez-Fernandez, H. M. Tauc, H. Jasper, Hallmarks of aging *Drosophila* intestinal stem cells. *Mech. Ageing Dev.* **190**, 111285 (2020). doi: 10.1016/j.mad.2020.111285; pmid: 32544407
19. A. Rezaei, M. S. Krishna, H. T. Santhosh, Male age affects female mate preference, quantity of accessory gland proteins, and sperm traits and female fitness in *D. melanogaster*. *Zoolog. Sci.* **32**, 16–24 (2015). doi: 10.2108/zs140121; pmid: 25660692
20. B. Biteau, C. E. Hochmuth, H. Jasper, JNK activity in somatic stem cells causes loss of tissue homeostasis in the aging *Drosophila* gut. *Cell Stem Cell* **3**, 442–455 (2008). doi: 10.1016/j.stem.2008.07.024; pmid: 18940735
21. J. C. Regan *et al*., Sex difference in pathology of the ageing gut mediates the greater response of female lifespan to dietary restriction. *eLife* **5**, e10956 (2016). doi: 10.7554/eLife.10956; pmid: 26878754
22. E. L. Arrese, J. L. Soulages, Insect fat body: Energy, metabolism, and regulation. *Annu. Rev. Entomol.* **55**, 207–225 (2010). doi: 10.1146/annurev-ento-112408-085356; pmid: 19725772
23. V. P. Losick, D. T. Fox, A. C. Spradling, Polyploidization and cell fusion contribute to wound healing in the adult *Drosophila* epithelium. *Curr. Biol.* **23**, 2224–2232 (2013). doi: 10.1016/j.cub.2013.09.029; pmid: 24184101
24. E. M. Lucchetta, B. Ohlstein, Amitosis of polyploid cells regenerates functional stem cells in the *Drosophila* intestine. *Cell Stem Cell* **20**, 609–620.e6 (2017). doi: 10.1016/j.stem.2017.02.012; pmid: 28343984
25. T. J. Doherty, Invited review: Aging and sarcopenia. *J. Appl. Physiol.* **95**, 1717–1727 (2003). doi: 10.1152/japplphysiol.00347.2003; pmid: 12970377
26. H.-J. Jeon *et al*., Age-related change in γH2AX of *Drosophila* muscle: Its significance as a marker for muscle damage and longevity. *Biogerontology* **16**, 503–516 (2015). doi: 10.1007/s10522-015-9573-0; pmid: 25860864
27. Tabula Muris Consortium, A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020). doi: 10.1038/s41586-020-2496-1; pmid: 32669714
28. H. Hall *et al*., Transcriptome profiling of aging *Drosophila* photoreceptors reveals gene expression trends that correlate with visual senescence. *BMC Genomics* **18**, 894 (2017). doi: 10.1186/s12864-017-4304-3; pmid: 29162050
29. Y. Gonskikh, N. Polacek, Alterations of the translation apparatus during aging and stress response. *Mech. Ageing Dev.* **168**, 30–36 (2017). doi: 10.1016/j.mad.2017.04.003; pmid: 28414025
30. H. Ruhmann, M. Koppik, M. F. Wolfner, C. Fricke, The impact of ageing on male reproductive success in *Drosophila melanogaster*. *Exp. Gerontol.* **103**, 1–10 (2018). doi: 10.1016/j.exger.2017.12.013; pmid: 29258876
31. S. Horvath, K. Raj, DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018). doi: 10.1038/s41576-018-0004-3; pmid: 29643443
32. M. T. Buckley *et al*., Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nat. Aging* **3**, 121–137 (2023). doi: 10.1038/s43587-022-00335-4; pmid: 37118510
33. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.* **67**, 301–320 (2005). doi: 10.1111/j.1467-9868.2005.00503.x
34. A. C. Ghosh *et al*., *Drosophila* PDGF/VEGF signaling from muscles to hepatocyte-like cells protects against obesity. *eLife* **9**, e56969 (2020). doi: 10.7554/eLife.56969; pmid: 33107824
35. N. Basisty, J. G. Meyer, B. Schilling, Protein turnover in aging and longevity. *Proteomics* **18**, e1700108 (2018). doi: 10.1002/pmic.201700108; pmid: 29453826
36. S. Aibar *et al*., SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017). doi: 10.1038/nmeth.4463; pmid: 28991892
37. H. Izgi *et al*., Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. *eLife* **11**, e68048 (2022). doi: 10.7554/eLife.68048; pmid: 35098922
38. S. M. Gao *et al*., Aging atlas reveals cell-type-specific regulation of pro-longevity strategies. bioRxiv 2023.02.28.530490 [Preprint] (2023). https://doi.org/10.1101/2023.02.28.530490.
39. A. E. Roux *et al*., The complete cell atlas of an aging multicellular organism. bioRxiv 2022.06.15.496201 [Preprint] (2022). https://doi.org/10.1101/2022.06.15.496201.
40. A. Sziraki *et al*., A global view of aging and Alzheimer's pathogenesis-associated cell population dynamics and molecular signatures in the human and mouse brains. bioRxiv 2022.09.28.509825 [Preprint] (2022). https://doi.org/10.1101/2022.09.28.509825.
41. R. Wang *et al*., Construction of a cross-species cell landscape at single-cell level. *Nucleic Acids Res.* **51**, 501–516 (2023). doi: 10.1093/nar/gkac633; pmid: 35929025
42. H. Chen, X. Zheng, Y. Zheng, Age-associated loss of lamin-B leads to systemic inflammation and gut hyperplasia. *Cell* **159**, 829–843 (2014). doi: 10.1016/j.cell.2014.10.028; pmid: 25417159
43. M. J. Bertram, G. A. Akerkar, R. L. Ard, C. Gonzalez, M. F. Wolfner, Cell type-specific gene expression in the *Drosophila melanogaster* male accessory gland. *Mech. Dev.* **38**, 33–40 (1992). doi: 10.1016/0925-4773(92)90036-J; pmid: 1525037
44. A. M. Box *et al*., Endocycles support tissue growth and regeneration of the adult *Drosophila* accessory gland. bioRxiv 719013 [Preprint] (2022). https://doi.org/10.1101/719013.
45. Y. Unhavaithaya, T. L. Orr-Weaver, Polyploidization of glia in neural development links tissue growth to blood-brain barrier integrity. *Genes Dev.* **26**, 31–36 (2012). doi: 10.1101/gad.177436.111; pmid: 22215808
46. T. Gehrmann *et al*., Nucleus-specific expression in the multinuclear mushroom-forming fungus *Agaricus bisporus* reveals different nuclear regulatory programs. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4429–4434 (2018). doi: 10.1073/pnas.1721381115; pmid: 29643074
47. P. von Aderkas, G. Rouault, R. Wagner, S. Chiwocha, A. Roques, Multinucleate storage cells in Douglas-fir (*Pseudotsuga menziesii* (Mirbel) Franco) and the effect of seed parasitism by the chalcid *Megastigmus spermotrophus* Wachtl. *Heredity* **94**, 616–622 (2005). doi: 10.1038/sj.hdy.6800670; pmid: 15829985
48. H.-Z. Chen *et al*., Canonical and atypical E2Fs regulate the mammalian endocycle. *Nat. Cell Biol.* **14**, 1192–1202 (2012). doi: 10.1038/ncb2595; pmid: 23064266
49. B. N. Kudryavtsev, M. V. Kudryavtseva, G. A. Sakuta, G. I. Stein, Human hepatocyte polyploidization kinetics in the course of life cycle. *Virchows Arch. B Cell Pathol. Incl. Mol. Pathol.* **64**, 387–393 (1993). doi: 10.1007/BF02915139; pmid: 8148960
50. S. Lee *et al*., Diverse cell-specific patterns of alternative polyadenylation in *Drosophila*. *Nat. Commun.* **13**, 5372 (2022). doi: 10.1038/s41467-022-32305-0; pmid: 36100597
51. X. Zhao, J. Karpac, Glutamate metabolism directs energetic trade-offs to shape host-pathogen susceptibility in *Drosophila*. *Cell Metab.* **33**, 2428–2444.e8 (2021). doi: 10.1016/j.cmet.2021.10.003; pmid: 34710355
52. N. Chatterjee, N. Perrimon, What fuels the fly: Energy metabolism in *Drosophila* and its application to the study of obesity and diabetes. *Sci. Adv.* **7**, eabg4336 (2021). doi: 10.1126/sciadv.abg4336; pmid: 34108216
53. T.-C. Lu, MeshifLu/AFCA: AFCA_v1.0 (AFCA_v1.0), Zenodo (2023); https://doi.org/10.5281/zenodo.7853649.

ACKNOWLEDGMENTS

We thank S. Aerts and J. Janssens for helping preprocess the data and for their constructive advice on annotation, N. Perrimon for suggestions during the initiation of this project, L. Buttitta for comments on polyploid phenotypes, and A. Parkhitko for discussion on metabolic pathways. We thank A.-L. Hsu (NYCU) for supporting C.-Y.L.; C. Desplan for supporting Y.-C.C.; and C. Qi, M. Burns, and all other lab members for comments and feedback during the project. We also thank the FCA Consortium and the fly community for their enthusiastic support. **Funding:** M.B. acknowledges the support of EPFL. T.J. was supported by Baylor College of Medicine Cancer and Cell Biology program grant T32 (GM136560). H.J.B. is supported by NIH/NIA (R01 AG07326), the Huffington foundation, and the endowment of the Chair of the Neurological Research Institute. L.L. is an investigator of the Howard Hughes Medical Institute and supported by NIH (R01-DC005982). L.L. and S.R.Q. are supported by Wu Tsai Neuroscience Institute (Neuro-omics program). This work was supported by CZ Biohub (S.R.Q.). S.L. was supported by a training award from NYSTEM contract C32559GG and the Center for Stem Cell Biology at MSKCC. Work in the E.C.L. group was supported by NIH/NINDS (R01-NS083833) and NIH MSK Core Grant P30-CA008748. Y.-C.C. was supported by the MacCracken Program at New York University, by a NYSTEM institutional training grant (contract C322560GG), and by a Scholarship to Study Abroad from the Ministry of Education, Taiwan. C.-Y.L. is supported by the NSTC grant from Taiwan (110-2311-B-A49A-501-MY3). H.L. is a CPRIT Scholar in Cancer Research (RR200063) and supported by NIH (R00AG062746), the Longevity Impetus Grant, the Welch Foundation, and the Ted Nash Long Life Foundation.

Author contributions: Conceptualization: T.-C.L., L.L., H.J., S.R.Q., and H.L. Sample preparation: N.S.K., X.T.C., and H.L. Sequencing: S.S.K., Y.Q., R.C.J., and H.L. Computational analysis: T.-C.L., M.B., J.L., T.J., and D.W. Data portal and AFCA website: J.C., T.-C.L., S.D., S.H.D., and A.O.P. APA analysis: S.L., Y.-C.C., and E.C.L. Data validation: Y.-J.P., H.J.B., T.J., Y.Q., N.A., and C.-Y.L. Writing: T.-C.L., M.B., and H.L. Review and editing: All authors. Supervision: H.J., S.R.Q., and H.L. Funding acquisition: H.J., S.R.Q., and H.L.
Competing interests: H.J., N.S.K., and X.T.C. are employees of Genentech, Inc. The other authors declare that they have no competing interests. **Data and materials availability:** All data are available for querying at https://hongjielilab.org/afca. Raw FASTQ files and processed h5ad files, including expression matrices and cell type annotations, can be downloaded from NCBI GEO under accession number GSE218661. Count matrices and analysis codes for generating the figures are available in Zenodo (*53*). FCA data were downloaded from https://www.flycellatlas.org/.

SUPPLEMENTARY MATERIALS

View/request a protocol for this paper from *Bio-protocol*.

Submitted 5 December 2022; accepted 4 May 2023
10.1126/science.adg0934

RESEARCH ARTICLE

INORGANIC CHEMISTRY

Diberyllocene, a stable compound of Be(I) with a Be–Be bond

Josef T. Boronski*, Agamemnon E. Crumpton, Lewis L. Wales, Simon Aldridge*

The complex diberyllocene, CpBeBeCp (Cp, cyclopentadienyl anion), has been the subject of numerous chemical investigations over the past five decades yet has eluded experimental characterization. We report the preparation and isolation of the compound by the reduction of beryllocene ($BeCp_2$) with a dimeric magnesium(I) complex and determination of its structure in the solid state by means of x-ray crystallography. Diberyllocene acts as a reductant in reactions that form beryllium-aluminum and beryllium-zinc bonds. Quantum chemical calculations indicate parallels between the electronic structure of diberyllocene and the simple homodiatomic species diberyllium (Be_2).

Because of the extreme biotoxicity of beryllium, its chemistry is the least well-developed of all the nonradioactive elements (*1*–*3*). This toxicity is a manifestation of beryllium's particular chemical properties: It forms the smallest of all metal ions (Be^{2+}; ionic radius, 0.31 Å; compare with that of Li^+, 0.60 Å), with unparalleled charge density (6.45 $Å^{-1}$; compare with that of Li^+, 1.67 $Å^{-1}$). Its bonding interactions often feature considerable covalent character on account of the highly polarizing properties of the dication (*2*). Given this capacity for covalency, the fundamental nature of beryllium-beryllium bonding has been debated for more than a century (*4*–*11*). Basic molecular orbital theory predicts that diberyllium (Be_2) should have a bond order of zero, and only recently has meaningful insight into the true nature of the beryllium-beryllium interaction in gaseous Be_2 been obtained (*12*, *13*). Moreover, despite seminal reports of zinc-zinc and magnesium-magnesium bonds in the early 2000s and a host of quantum chemical investigations into the potential stability of beryllium-containing analogs, the isolation of a compound that features a beryllium-beryllium bond in the condensed phases has not been achieved (*14*–*17*).

Given the dearth of compounds that feature beryllium-beryllium bonds, in recent years, synthetic studies have focused on the preparation of low-valent beryllium species stabilized by redox noninnocent carbene ligands (*18*–*20*). Compounds proposed to contain beryllium in the 0 and +1 oxidation states have been synthesized. However, such descriptions have proven controversial and are a source of ongoing debate within the literature (*21*–*24*). We recently reported the beryllium-aluminyl complex CpBeAl(NON) [**1**; NON, 4,5-bis(2,6-diisopropylanilido)-2,7-di-tert-butyl-9,9-dimethylxanthene; Cp, cyclopentadienyl anion, $[C_5H_5]^-$], which displays the reactivity expected from low-valent, nucleophilic beryllium (*25*–*27*). In this work, we present diberyllocene, CpBeBeCp (**2**), an unambiguous beryllium(I) compound that is stable in the condensed phases and features a Be–Be bond. The experimentally determined properties of diberyllocene (which is stable at 80°C for extended periods) validate past quantum chemical predictions of its stability with respect to disproportionation to Be(0) and $BeCp_2$ (*7*).

Synthesis and characterization

We previously succeeded in using beryllocene $BeCp_2$ for the synthesis of beryllium-aluminyl complex **1** (*25*, *28*). Calculations indicate that compound **1** features a higher partial positive charge at aluminum than beryllium. Thus, the formation of **1** could be viewed as a reduction of beryllium(II) to beryllium(I) by the aluminyl anion, with associated one-electron oxidation of aluminum(I), which forms a covalent Be–Al bond. We therefore set out to examine the possibility of reducing beryllocene to form a covalent Be–Be bond between two beryllium(I) centers. We turned our attention to di-magnesium(I) reagent $[(^{Mes}Nacnac)Mg]_2$ {**3**;^{Mes}Nacnac, $[(MesNCMe)_2CH]^-$; Mes, 2,4,6-trimethylphenyl}, which has been used for the controlled reduction of various organometallic compounds (*6*, *17*). At room temperature, under an inert atmosphere, the reaction of one equivalent of **3** with two equivalents of $BeCp_2$ in toluene leads to quantitative formation of diberyllocene CpBeBeCp (**2**) and (^{Mes}Nacnac)MgCp, as evidenced by multinuclear nuclear magnetic resonance (NMR) spectroscopy (Fig. 1 and fig. S5). Both compounds are highly soluble in alkane solvents, such as pentane and hexane. However, **2** is a volatile solid and can be purified through sublimation at room temperature, which allows for its isolation in 85% yield.

Compound **2** can be crystallized from a concentrated hexane solution, which yields crystals suitable for the elucidation of its structure by means of single-crystal x-ray diffraction (Fig. 1B). Compound **2** is shown to feature two half-sandwich (cyclopentadienyl)beryllium units linked through a beryllium–beryllium bond. Zinc is the only other element for which the dimetallocene structural motif has been reported—for example, in Cp*ZnZnCp* [**4**; Cp*, pentamethylcyclopentadienyl anion $(C_5Me_5)^-$]—and the configuration of **2** closely resembles that of **4** (*14*, *15*). Both **2** and **4** display D_{5h} symmetry in the solid state, with the cyclopentadienyl ligands adopting a parallel, eclipsed configuration. At 2.0545(18) Å, where the number in parentheses is the standard deviation of the least significant digits, the Be1–Be2 distance in **2** is in line with the sum of the single-bond covalent radii for beryllium (2.04 Å) (*29*). This value is extremely close to that predicted by previous computational investigations of this compound (2.041 to 2.077 Å) (*7*–*9*). Examples of hydride-bridged beryllium compounds of the form $XBe(\mu\text{-}H)_2BeX$ all feature wider Be–Be separations [range, 2.098(3) to 2.212(8) Å; mean, 2.136 Å] (*26*, *30*–*32*). The Be–C distances within **2** range from 1.923(2) to 1.938(2) Å (mean, 1.930 Å) and are considerably longer than that of $BeCp_2$ [range, 1.862(9) to 1.936(8) Å; mean, 1.904 Å] (*28*). Consistently, the Be-(η^5-C_5H_5) centroid distances for Be1 and Be2 are both 1.519 Å, which is again longer than that reported for $BeCp_2$ (1.485 Å). These bond metrics reflect the lower oxidation state and greater ionic radius of the beryllium(I) centers in **2** compared with those of the beryllium(II) center in $BeCp_2$.

Compound **2** was investigated by use of multinuclear NMR spectroscopy. The ^{1}H NMR spectrum of **2** consists of a single resonance at 5.73 parts per million (ppm), which corresponds to the five equivalent cyclopentadienyl ligand protons. The ^{1}H NMR resonances of terminal beryllium-bound hydride ligands are generally found between 4 and 5 ppm (*32*, *33*). Thus, **2** shows no signals attributable to Be–H hydrides in its ^{1}H NMR spectrum, only the resonance that corresponds to the cyclopentadienyl ligand. Similarly, the ^{13}C NMR spectrum of **2** features one signal at 102.7 ppm. Compound **2** was also investigated by means of ^{9}Be NMR spectroscopy—which offers a convenient probe of the electron density at the beryllium center—and shows a single high-field resonance at −27.6 ppm (*25*, *34*). The coordination of strongly electron-donating anionic ligands to beryllium generally shifts the ^{9}Be NMR resonances further upfield as the increasingly electron-rich beryllium center becomes more strongly shielded. So, for example, CpBeCl exhibits a $\delta_{^9Be}$ shift of −19.5 ppm,

Chemistry Research Laboratory, Department of Chemistry, Oxford, OX1 3TA, UK.
*Corresponding author. Email: josef.boronski@sjc.ox.ac.uk (J.T.B.); simon.aldridge@chem.ox.ac.uk (S.A.)

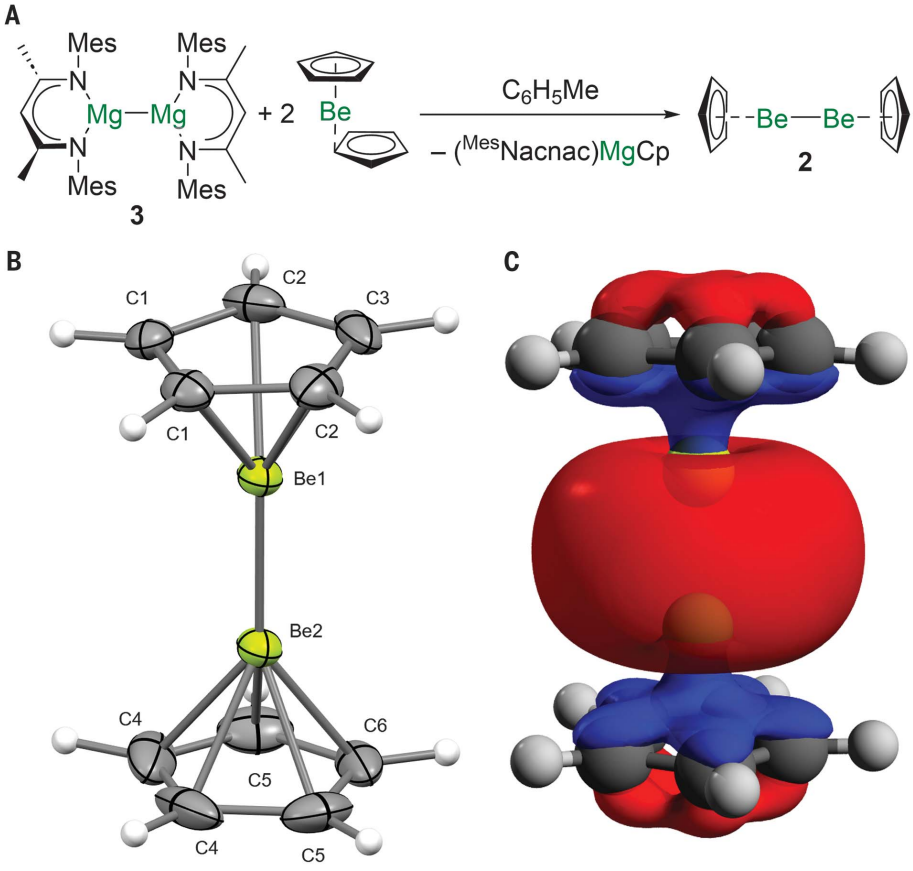


Fig. 1. Synthesis, crystallographic, and electronic structure of diberyllocene (2). (**A**) Synthesis of **2** through reaction of magnesium(I) complex **3** and beryllocene. (**B**) Molecular structure of **2** in the solid state, as determined with x-ray crystallography. Thermal ellipsoids set at 50% probability. (**C**) Calculated HOMO for **2** [isovalue, 0.03 arbitrary units (a.u.)].

whereas the corresponding signals of CpBeMe, CpBeGa(NON), CpBe[Si(CH_3)$_3$], and beryllium–aluminyl compound **1** are found at –20.5, –26.9, –27.7, and –28.8 ppm, respectively (table S2) (*34*). These data suggest that the beryllium center in **2** is similarly as electron rich as that in **1**, which is consistent with the description of **2** as a metal-metal bonded beryllium(I) compound (*25*, *34*). Moreover, the ^{9}Be NMR shift for compound **2** also provides evidence against a bridging hydride formulation for this compound because such a species would be expected to show a much lower field chemical shift (fig. S23) (*34*).

To further argue against the hydride formulation for **2**, we measured attenuated total reflection infrared (ATR IR) spectra for samples of **2** prepared in both protio and perdeutero toluene solvents (figs. S14 and S15). The spectra are identical, and neither features an absorbance band in the 1500 to 2000 cm^{-1} region where a Be–H stretch might be expected to be observed (*35*). The infrared spectra of both **2** and the hypothetical complex CpBe(μ-H)$_2$BeCp were simulated through quantum chemical calculations (figs. S17 and S18). The experimentally derived ATR IR spectrum of **2** closely matches the spectrum calculated for this complex (fig. S19) and differs substantially from that of the hydride-bridged species (fig. S20), which features a very intense band at 1527 cm^{-1} [υ_s(Be–H$_b$)]. The beryllium(II) hydride compound (CpBeH)$_n$ is thermally fragile and has been reported to undergo ligand redistribution to form BeCp$_2$ and BeH$_2$ within minutes at –10°C (*35*, *36*). This reactivity contrasts with the relative stability of **2**, which shows no obvious signs of degradation upon heating in solution at 80°C for 48 hours.

Quantum chemical investigations

Quantum chemical calculations were performed on **2** (B3LYP D3BJ def2-TZVP def2/J), in light of the experimental data that had been unavailable for previous theoretical investigations of this molecule (*7*–*9*). Key structural parameters align well with those determined crystallographically [for example, *d*(Be–Be) = 2.046 Å, as compared with 2.0545(18) Å]. These calculations indicate that the energetic separation between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) is 5.541 eV. The HOMO of **2** is calculated to be the Be–Be σ-bonding orbital, with some Be–Cp σ-bonding character (Fig. 1C); the lower-lying HOMO – 1 to HOMO – 4 correspond to the Cp–Be π-bonding combinations. As with the homodiatomic molecule Be$_2$, we calculated the Be–Be σ* antibonding orbital (LUMO + 1) of diberyllocene to be markedly lower in energy than the Be–Be π-bonding orbitals (LUMO + 4 and LUMO + 5) (figs. S27 and S29) (*12*, *13*). This implies that further reduction of **2** would lead to a weakening of the Be–Be interaction. Indeed, the Be–Be distance in Be$_2$ (2.45 Å) is considerably longer than that measured for **2** [2.0545(18) Å] (*12*, *13*). As such, the core of **2** could be considered to be a (cyclopentadienyl-stabilized) dication of diberyllium, [BeBe]$^{2+}$, which molecular orbital theory would predict to feature a beryllium–beryllium σ-bond and a formal bond order of one, as is found here (*37*, *38*).

We used natural bond orbital (NBO) calculations to examine the bonding and charge distribution within **2**. Natural population analysis (NPA) indicates that the 2s valence orbital of each beryllium atom is substantially populated [0.98 electrons (e$^-$)], with a non-negligible population of the beryllium 2p orbitals (0.17 e$^-$). These data are consistent with a formal beryllium(I) oxidation state. NBO calculations suggest that the Be–Be bond comprises 93% 2s character and 7% 2p character, with each beryllium center contributing equally (50:50). We calculated a Wiberg bond index (WBI) of 0.90 for the Be–Be bonding interaction, which is slightly greater than the WBI calculated for the Al–Be bond of **1** (0.82) (*25*). The charge distribution in **2** was probed with both NPA and quantum theory of atoms in molecules (QTAIM) calculations. QTAIM analysis indicates that a non-nuclear attractor is present for the Be–Be interaction, as has been indicated with previous theoretical studies of **2**, and yields a charge of +1.33 for each beryllium atom (*8*). The NPA charges at each beryllium center are lower (+0.84), but these do not take into account the presence of the non-nuclear attractor. QTAIM calculations previously performed on **1** provide a charge of +1.39 for the beryllium center in this compound, which suggests that the beryllium centers in **1** and **2** are similarly electron rich. This is consistent with the similar ^{9}Be NMR shifts measured for the two compounds (*25*, *34*).

Because of the high quality of the x-ray crystallographic data obtained for **2**, we used quantum-crystallographic methods to gain further evidence that this compound does not feature bridging hydride ligands. Even without using quantum-crystallographic methods, there is insufficient residual electron density to account for the presence of bridging hydride ligands (figs. S34 and S35). We employed

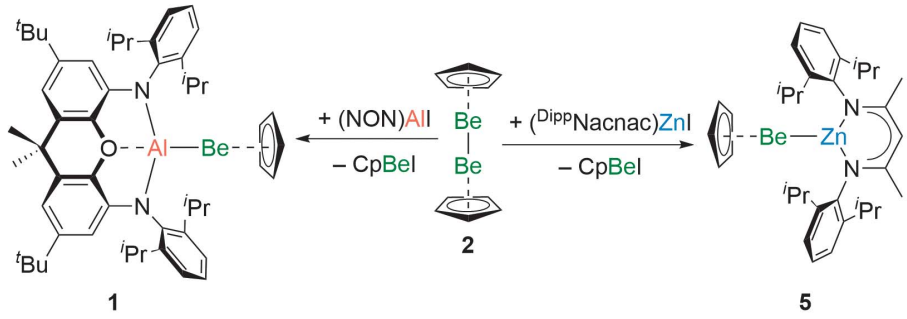


Fig. 2. Reactivity of compound 2 with metal-iodide complexes, which leads to the formation of beryllium-metal bonded complexes 1 and 5. ^{i}Pr, isopropyl; ^{t}Bu, *tert*-butyl.

the NoSpherA2 method; this technique allows for refinement of crystallographic data on the basis of the partitioned wave function of a molecule, rather than the traditional approach of refining against a model in which electron density is considered as point charges associated with specific atoms (independent electron model) (*39*, *40*). By using NoSpherA2, essentially all electron density in the data collected for **2** is accounted for (figs. S31 and S32). The electron density associated with the cyclopentadienyl C–H bonds can be clearly identified (fig. S35), and there is no residual electron density resulting from bridging hydride ligands (figs. S31 and S33).

Reactivity studies

In view of the metal–metal bond in **2**, we probed its reactivity as a source of low-valent beryllium. Complex **2** does not react with H_2 (one atmosphere), even at elevated temperatures—a finding that is consistent with di-magnesium (I) compounds such as **3**, which also do not react with H_2 (*17*). However, compound **2** reduces (NON)AlI to cleanly and quantitatively yield the known beryllium-aluminyl compound **1** and CpBeI (Fig. 2, left), which evidences the usefulness of **2** for the synthesis of beryllium-metal bonds (*25*). This reactivity would not be expected from a beryllium(II) hydride compound. The reaction of **2** with (^{Dipp}Nacnac)ZnI (^{Dipp}Nacnac, $[(DippNCMe)_2CH]^-$; Dipp, 2,6-diisopropylphenyl) was similarly tested to examine whether the formation of a (hitherto unknown) Zn–Be bond could be achieved (Fig. 2, right). In a similar fashion, ^{1}H NMR spectroscopy indicates the complete consumption of **2**, along with quantitative formation of CpBeI and a (^{Dipp}Nacnac)-containing species. Additionally, the formation of a small amount of dark gray metallic precipitate was observed. Crystallization from hexane yielded CpBeZn (^{Dipp}Nacnac) (**5**), which was structurally authenticated by single-crystal x-ray diffraction (fig. S21).

The Zn–Be bond within compound **5** represents a very rare example of a beryllium-metal bonding combination and provides further evidence for the constitution of **2** itself (*25*, *27*, *41*). The Zn–Be bond distance within **5** [2.169(10) Å] is in line with the sum of the covalent radii of Be and Zn (2.20 Å) and is consistent with the presence of a covalent metal-metal bond (*29*). Additionally, **5** exhibits a ^{9}Be NMR resonance (–27.7 ppm) that is consistent with a very electron-rich beryllium metal center, similar to those of **1** and **2** (*25*, *34*). Quantum chemical calculations performed on **5** (B3LYP D3BJ def2-TZVP def2/J) yield a similar Zn–Be distance (2.175 Å) to that determined crystallographically and indicate that the HOMO is a Zn–Be σ-bonding orbital (fig. S30). On the basis of the Pauling electronegativities of beryllium (1.57) and zinc (1.65), compound **5** might be assigned a beryllium(II)/zinc(0) formalism. However, quantum chemical calculations indicate considerable covalent contributions to the Zn–Be bonding in this compound, in similar fashion to the Be–Al bonding in **1** (*25*). For example, NPA calculations imply that the valence orbitals of both zinc and beryllium are substantially populated (Be, 2s, 1.00 e^- and 2p, 0.16; Zn, 4s, 1.04 e^-). Additionally, NBO analysis suggests that beryllium and zinc make almost equal contributions to the Zn–Be bond (Zn:Be, 49.9:50.1). Thus, a beryllium(I)–zinc(I) formulation seems a more appropriate descriptor for **5**.

After half a century, diberyllocene (**2**) has been synthesized. The Be–Be distance in **2** [2.0545(18) Å] is in line with all previous quantum chemical investigations of this compound. Moreover, **2** reacts as a reductant and can be used to synthesize beryllium-metal bonds.

REFERENCES AND NOTES

1. R. Puchta, *Nat. Chem.* **3**, 416 (2011).
2. D. Naglav, M. R. Buchner, G. Bendt, F. Kraus, S. Schulz, *Angew. Chem. Int. Ed.* **55**, 10562–10576 (2016).
3. M. R. Buchner, *Chem. Commun.* **56**, 8895–8907 (2020).
4. C. Jones, *Commun. Chem.* **3**, 159 (2020).
5. L. A. Freeman, J. E. Walley, R. J. Gilliard Jr., *Nat. Synth.* **1**, 439–448 (2022).
6. S. J. Bonyhady *et al.*, *Chemistry* **16**, 938–955 (2010).
7. A. Velazquez, I. Fernández, G. Frenking, G. Merino, *Organometallics* **26**, 4731–4736 (2007).
8. X. Li *et al.*, *Organometallics* **32**, 1060–1066 (2013).
9. Y. Xie, H. F. SchaeferIII, E. D. Jemmis, *Chem. Phys. Lett.* **402**, 414–421 (2005).
10. S. A. Couchman, N. Holzmann, G. Frenking, D. J. D. Wilson, J. L. Dutton, *Dalton Trans.* **42**, 11375–11384 (2013).
11. H. Schmidbaur, *Be Organoberyllium Compounds* (vol. 21, Gmelin Handbook of Inorganic and Organometallic Chemistry, Springer, 1987).
12. J. M. Merritt, V. E. Bondybey, M. C. Heaven, *Science* **324**, 1548–1551 (2009).
13. K. Patkowski, V. Špirko, K. Szalewicz, *Science* **326**, 1382–1384 (2009).
14. I. Resa, E. Carmona, E. Gutierrez-Puebla, A. Monge, *Science* **305**, 1136–1138 (2004).
15. A. Grirrane *et al.*, *J. Am. Chem. Soc.* **129**, 693–703 (2007).
16. S. P. Green, C. Jones, A. Stasch, *Science* **318**, 1754–1757 (2007).
17. C. Jones, *Nat. Rev. Chem.* **1**, 0059 (2017).
18. M. Arrowsmith *et al.*, *Nat. Chem.* **8**, 638–642 (2016).
19. C. Czernetzki *et al.*, *Angew. Chem. Int. Ed.* **60**, 20776–20780 (2021).
20. G. Wang *et al.*, *J. Am. Chem. Soc.* **142**, 4560–4564 (2020).
21. D. Jędrzkiewicz *et al.*, *Angew. Chem. Int. Ed.* **61**, e202200511 (2022).
22. M. Gimferrer *et al.*, *Chem. Sci.* **13**, 6583–6591 (2022).
23. M. Gimferrer *et al.*, *Chem. Sci.* **14**, 384–392 (2023).
24. S. Pan, G. Frenking, *Chem. Sci.* **14**, 379–383 (2022).
25. J. T. Boronski *et al.*, *J. Am. Chem. Soc.* **145**, 4408–4413 (2023).
26. A. Paparo *et al.*, *Dalton Trans.* **50**, 7604–7609 (2021).
27. A. Paparo, C. D. Smith, C. Jones, *Angew. Chem. Int. Ed.* **58**, 11459–11463 (2019).
28. R. Fernández, E. Carmona, *Eur. J. Inorg. Chem.* **2005**, 3197–3206 (2005).
29. P. Pyykkö, *J. Phys. Chem. A* **119**, 2326–2337 (2015).
30. M. Arrowsmith, M. S. Hill, G. Kociok-Köhn, *Organometallics* **34**, 653–662 (2015).
31. G. W. Adamson, H. M. M. Shearer, *Chem. Commun.* **11**, 240 (1965).
32. T. J. Hadlington, T. Szilvási, *Nat. Commun.* **13**, 461 (2022).
33. R. Han, G. Parkin, *Inorg. Chem.* **31**, 983–988 (1992).
34. P. G. Plieger *et al.*, *J. Am. Chem. Soc.* **126**, 14651–14658 (2004).
35. M. del Mar Conejo *et al.*, *Chemistry* **9**, 4462–4471 (2003).
36. T. C. Bartke, A. Bjorseth, A. Haaland, K.-M. Marstokk, H. Møllendal, *J. Organomet. Chem.* **85**, 271–277 (1975).
37. C. A. Nicolaides, P. Valtazanos, *Chem. Phys. Lett.* **174**, 489–493 (1990).
38. A. Kalemos, *J. Chem. Phys.* **145**, 214302 (2016).
39. L. Midgley *et al.*, *Acta Crystallogr. A Found. Adv.* **77**, 519–533 (2021).
40. F. Kleemiss *et al.*, *Chem. Sci.* **12**, 1675–1692 (2020).
41. H. Braunschweig, K. Gruss, K. Radacki, *Angew. Chem. Int. Ed.* **48**, 4239–4241 (2009).
42. J. T. Boronski, A. E. Crumpton, L. L. Wales, S. Aldridge, Diberyllocene: xyz-coordinates of optimized structures. Dryad (2023); https://doi.org/10.5061/dryad.kwh70rz8n.

ACKNOWLEDGMENTS

Funding: J.T.B. thanks St John's College, Oxford, for a Junior Research Fellowship. We thank the John Fell Fund (0011792) for financial support. We thank the EPSRC Centre for Doctoral Training in Inorganic Chemistry for Future Manufacturing (OxICFM, EP/S023828/1 studentship for L.L.W. and A.E.C.). **Author contributions:** J.T.B.: conceptualization, investigation, visualization, writing – original draft, writing – review and editing, funding acquisition, supervision, and project administration. A.E.C.: quantum-crystallographic investigations, writing – review and editing. L.L.W.: collection and processing of x-ray crystallographic data, writing – review and editing. S.A.: supervision, project administration, and writing – review and editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** X-ray data are available free of charge from the Cambridge Crystallographic Data Centre under reference numbers CCDC 2245595 (**2**), 2245596 (**5**), and 2246346 [(^{Mes}Nacnac)MgCp]. All other experimental, spectroscopic, crystallographic, and computational data are included in the supplementary materials. Computational data are also available through Dryad (*42*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh4419
Materials and Methods
Figs. S1 to S35
Tables S1 to S3
References (*43*–*56*)

Submitted 6 March 2023; accepted 25 April 2023
10.1126/science.adh4419

BIOCATALYSIS

An enzyme cascade enables production of therapeutic oligonucleotides in a single operation

E. R. Moody†, R. Obexer†, F. Nickl, R. Spiess, S. L. Lovelock*

Therapeutic oligonucleotides have emerged as a powerful drug modality with the potential to treat a wide range of diseases; however, the rising number of therapies poses a manufacturing challenge. Existing synthetic methods use stepwise extension of sequences immobilized on solid supports and are limited by their scalability and sustainability. We report a biocatalytic approach to efficiently produce oligonucleotides in a single operation where polymerases and endonucleases work in synergy to amplify complementary sequences embedded within catalytic self-priming templates. This approach uses unprotected building blocks and aqueous conditions. We demonstrate the versatility of this methodology through the synthesis of clinically relevant oligonucleotide sequences containing diverse modifications.

Nucleic acid therapeutics (NATs) are short oligonucleotide sequences [typically ~20 nucleotides (nt)] that selectively bind to target mRNAs to modulate the production of disease-related proteins. Drug specificity is defined by the base sequence, and chemical modifications to the phosphate backbone, ribose sugar, and nucleobases confer improved efficacy, metabolic stability, and bioavailability (Fig. 1A) (*1*, *2*). Until recently, marketed NATs were limited to the treatment of rare diseases (*3*, *4*). However, the recent approval of inclisiran as a therapy for atherosclerotic cardiovascular disease, which affects ~30 million people in the US alone, marks a transition in the field from low-volume drugs to oligonucleotide therapies for the treatment of population-based disorders (*5*). This emergence of NATs for common diseases creates a substantial manufacturing challenge because existing methods of chemical synthesis are not suitable for large-scale applications.

Traditional approaches to oligonucleotide synthesis rely on solid-phase phosphoramidite chemistry, which involves iterative coupling, capping, oxidation, and deprotection steps for each nucleotide addition (*6*). Although powerful, this approach suffers from fundamental limitations that make it unsuitable for large-scale manufacturing. The use of solid supports limits production to <10-kg batches (*7*). An excess of densely protected monomer (Fig. 1B) is needed to ensure efficient coupling, which compromises atom efficiency and leads to the formation of by-products that must be separated by washing steps. Moreover, final products require chromatographic purification, which necessitates the use of prohibitively large volumes of acetonitrile (1000 kg per kilogram of oligonucleotide) and contributes to large process mass intensity (PMI) ratios (~4000 kg waste per kilogram of product for a 20-nt sequence) (*8*, *9*). Owing to the limitations of existing synthetic methods, currently marketed therapies are produced in <50% yield with modest purities (~90%) and phosphorothioate (PS)–modified sequences are produced as complex mixtures of stereoisomers. Given these challenges, several elegant approaches to oligonucleotide production have been developed, including the use of nucleoside 3′-oxazaphospholidine derivatives (*10*), phosphorus(V) reagents (*11*–*13*), chiral phosphoric acid catalysts (*14*), and enzymatic approaches that use terminal deoxyribonucleotidyl transferases (*15*–*19*). Although these strategies offer improvements in stereocontrol, step economy, and/or reduced solvent consumption, they all share the same fundamental approach of stepwise chain extension by sequential coupling and deprotection steps. Alternative strategies to construct oligonucleotides in a convergent fashion by ligating smaller fragments have been developed by Biogen, Almac, GSK, and Ajinomoto (*20*–*24*). These methods have great potential for improving the speed of oligonucleotide synthesis and generating products with improved purity, but they currently rely

Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester, UK.
*Corresponding author. Email: sarah.lovelock@manchester.ac.uk
†These authors contributed equally to this work.

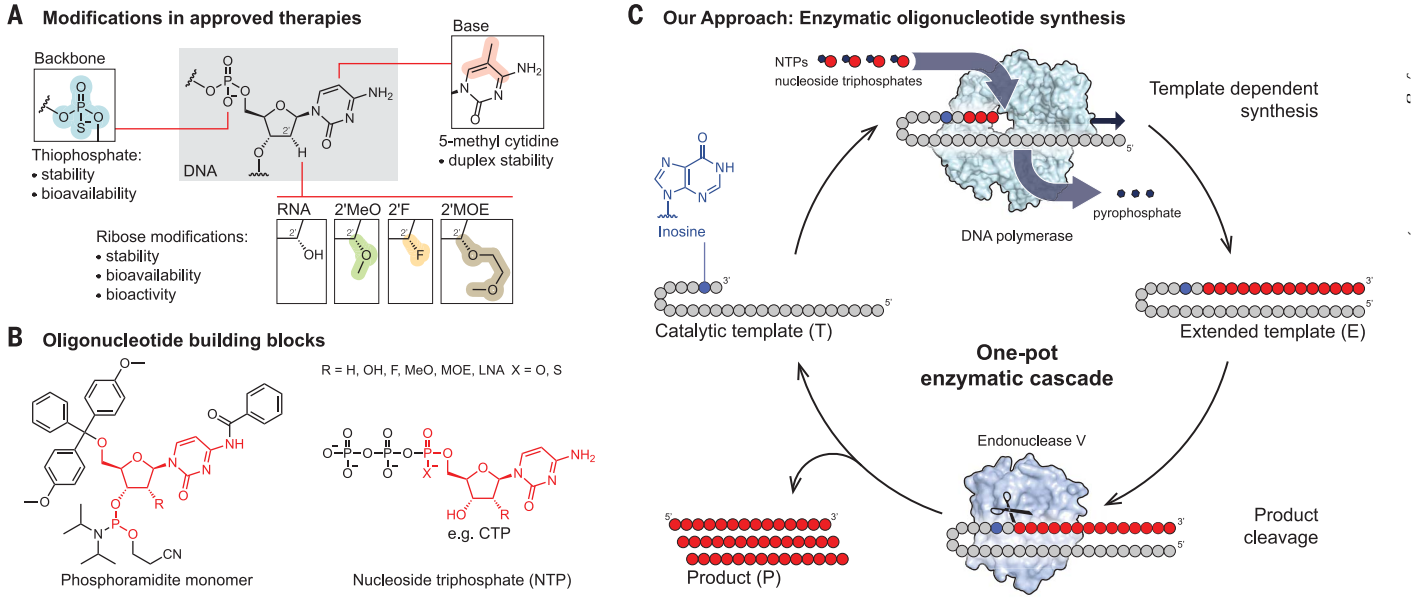


Fig. 1. Nucleic acid therapeutics (NATs). (**A**) Pharmaceutically relevant oligonucleotide modifications and their biological effects. (**B**) Phosphoramidite and nucleoside triphosphate building blocks used in chemical and enzymatic synthesis, respectively. (**C**) This work: a one-pot enzymatic cascade for oligonucleotide synthesis.

on fragments produced by using phosphoramidite chemistry. Here, we report a biocatalytic platform to produce modified oligonucleotides from unprotected nucleoside triphosphate (NTP) building blocks in a single operation. The process operates under aqueous conditions and does not require solid supports or large volumes of acetonitrile, thus addressing scalability and sustainability challenges associated with existing methods.

An enzyme cascade for scalable oligonucleotide synthesis

Our strategy uses a DNA polymerase and NTP building blocks to extend a catalytic self-priming hairpin template (Fig. 1C). An endonuclease V (EndoV) then selectively cleaves a single strand of the resulting duplex DNA downstream of an inosine base built into the hairpin sequence, releasing the product and regenerating the template. Given that the polymerase is larger than the extended DNA, the endonuclease is unable to bind to the template until the polymerase is released, thus avoiding premature cleavage of partially amplified sequences. The process is designed to operate under isothermal conditions above the product melting temperature (T_m) to facilitate scale-up and to ensure effective dissociation of the cleaved product from the template. In this way, repeated cycles of polymerase extension and endonuclease cleavage result in accumulation of the desired oligonucleotide product. Our approach has some similarities to alternative DNA amplification methods, such as those used in biosensing, bioimaging, and genome sequencing (*25*, *26*), but is specifically tailored toward the large-scale production of short, modified oligonucleotides typical of those used as therapeutics. Unlike alternative methods, our system does not require stoichiometric primers (*27*), delivers unscarred products, and does not generate oligonucleotide by-products that complicate purification and compromise process efficiency (*26*, *28*).

To evaluate the potential of the proposed cascade, we first assessed the extension and cleavage steps in isolation. The B-family polymerase from *Thermococcus kodakarensis* KOD1 (KOD) efficiently extended a self-priming template (T1) with deoxynucleotide triphosphates (dNTPs) to afford the double-stranded product (E1) in quantitative conversion (fig. S1A). The cleavage reaction performed with an endonuclease from *Thermotoga maritima* (*Tm*EndoV) also proceeded to completion, providing the 18-nt product (P1) and template (T1) (fig. S1B). With the individual steps working as intended, we performed a one-pot cascade using KOD polymerase, *Tm*EndoV endonuclease, T1 (20 μM), and dNTPs (250 μM each). Encouragingly, under these conditions the target sequence (28 μM) was the sole product, which correlates with 78% consumption of the limiting 3′-deoxythymidine 5′-triphosphate (dTTP) reagent (fig. S1C).

To intensify the process, we next investigated the effect of increased substrate (dNTP) loading on the individual steps. Although KOD polymerase was able to efficiently extend T1 in the presence of 30 mM dNTPs (fig. S2A), *Tm*EndoV endonuclease–mediated hydrolysis was inhibited at dNTP concentrations >2.0 mM (fig. S2B). To address this limitation, we evaluated a panel of thermophilic *Tm*EndoV homologs and identified an endonuclease from *Thermotoga neapolitana* (*Tn*EndoV) that operates effectively at high dNTP concentrations, catalyzing complete hydrolysis of E1 (20 μM) in the presence of 20 mM dNTPs (fig. S2C). After optimization of buffer conditions, hairpin sequences, and reaction temperature (fig. S3), 330 cycles of template extension and product cleavage were achieved in analytical-scale reactions (20 μl) under isothermal conditions, affording 0.33 mM P1 as the sole product (fig. S4). Next, we performed a reaction on a 5-ml scale, which provided 0.44 mM P1 (equivalent to 2.5 g/l), consuming 69% of the available dNTP starting materials. After removal of the proteins and membrane filtration, we isolated P1 in 88% purity (11.5 mg, 2.1 μmol) without chromatographic purification (fig. S5). Although in this instance we did not attempt to separate the catalytic template (<1 mol %) from P1, there are numerous opportunities for template removal and/or recycling (see below) (*29*–*31*). Finally, to demonstrate that our processes can be readily scaled without compromising efficiency, we performed a biotransformation on a 50-ml scale, delivering 123 mg of P1 (22.2 μmol) with 89% purity after protein precipitation and membrane filtration (Fig. 2, A to C, and fig. S6).

Using the metrics of our 50-ml scale reaction, we performed a PMI analysis to evaluate the material efficiency of the process (Fig. 2D). Even at this early stage of development, the enzymatic approach compares favorably to traditional phosphoramidite chemistry [530 kg compared to ~4000 kg of raw material per kg

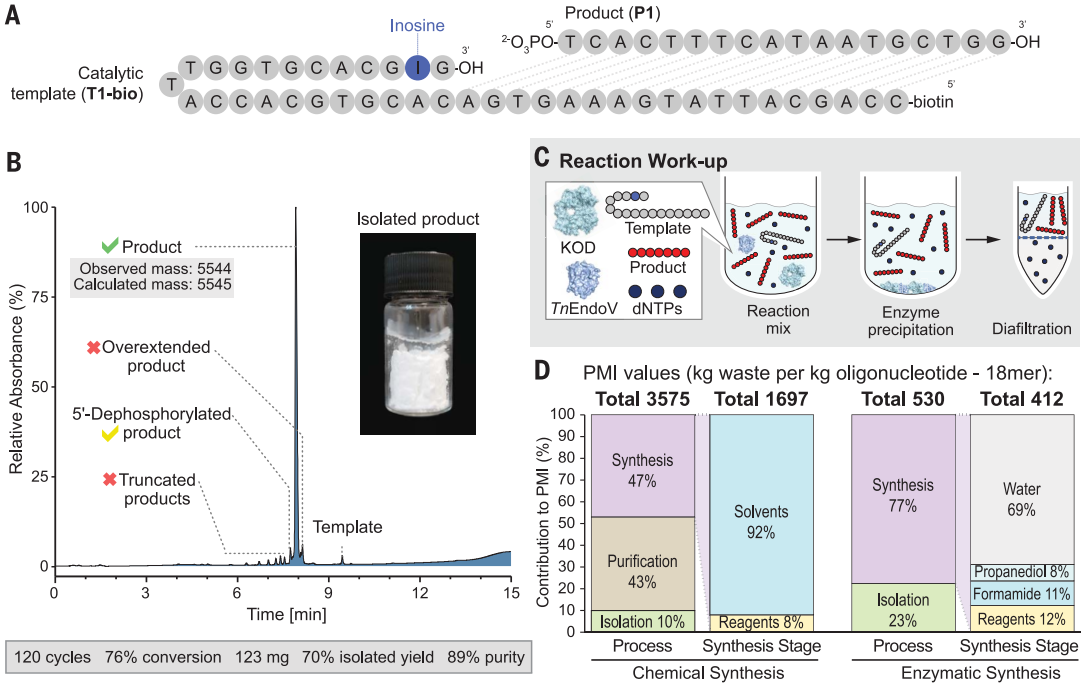


Fig. 2. 123-mg scale synthesis of oligonucleotide P1. (**A**) P1 nucleotide sequence. (**B**) LC-MS trace and image of the isolated product. Minor impurities include overextended product and truncated sequences. (**C**) Scheme showing the product isolation steps. (**D**) PMI analysis (kilogram of waste per kilogram of oligonucleotide product) for the production of an 18-nt sequence by using traditional phosphoramidite chemistry (*8*) versus our enzymatic process. The bar chart shows how each stage of the manufacturing process (synthesis, purification, and isolation) contributes to the overall PMI and provides a breakdown of the raw materials used in the synthesis stage.

of product (*8*)]. This improvement in PMI primarily comes from alleviating the need for multiple washing steps and chromatographic purification. The template that we used in our enzymatic approach is currently produced by using phosphoramidite chemistry, but it is only required in catalytic quantities (<1%) and thus has substantially reduced economic and environmental impacts compared with producing large quantities of DNA products chemically. Moreover, there is the potential to generate catalytic templates by using enzymatic methods in the future, which would provide further improvements in green chemistry metrics.

Biocatalytic synthesis of modified sequences

To manufacture therapeutic oligonucleotides, we require polymerases and endonucleases with activity toward modified nucleic acids. To this end, eight thermophilic polymerases were selected for evaluation and used to extend a hairpin template encoding the sequence TCAGTCAG (T2) with a mixture of three unmodified dNTPs and one modified NTP (Fig. 3A). Polymerases that were able to transcribe the template and incorporate two copies of a modified NTP resulted in full-length extension. We also performed control reactions in the absence of modified NTP. From this preliminary screen, we identified several polymerases with good activity toward NTP substrates containing S_p-PS [(S_p)-**2**]2′-fluoro (**3**), and a combination of S_p-PS and 2′-fluoro (**4**) modifications, including KOD and TfPol* polymerases, as evidenced by polyacrylamide gel electrophoresis (PAGE) analysis (Fig. 3B and fig. S7). We observed more modest activity toward 2′-MeO-NTP substrates (**5**) with several polymerases including SFM4-6 (*32*), TfPol*, and Dpo4*, whereas KOD^{DGLNK} (*33*) and SFM4-6 displayed modest activity with locked nucleic acid (LNA)–modified substrates (**6**). We observed no activity toward R_p-stereoisomers of α-PS–modified dNTPs (R_p-2) with any of the polymerases tested. 2′-O-Methoxyethyl (2′-MOE)–modified NTPs are not commercially available and were not evaluated in this study; however, a variant of Tgo polymerase has recently been reported that has activity toward 2′-MOE-NTPs, making us optimistic that our methodology will be extendable to this class of substrate (*34*). Similarly, therapeutic oligonucleotides containing nucleobase modifications (e.g., 5′-methylated pyrimidines) should also be accessible with our technology, as polymerases have been shown to accept a wide range of base-modified NTPs (*35*–*37*).

Having identified promising polymerases, we explored sequential incorporations of modified NTPs by extending templates (T3 to T6) that encode eight copies of a single nucleobase (Fig. 3C). Consistent with the initial screen,

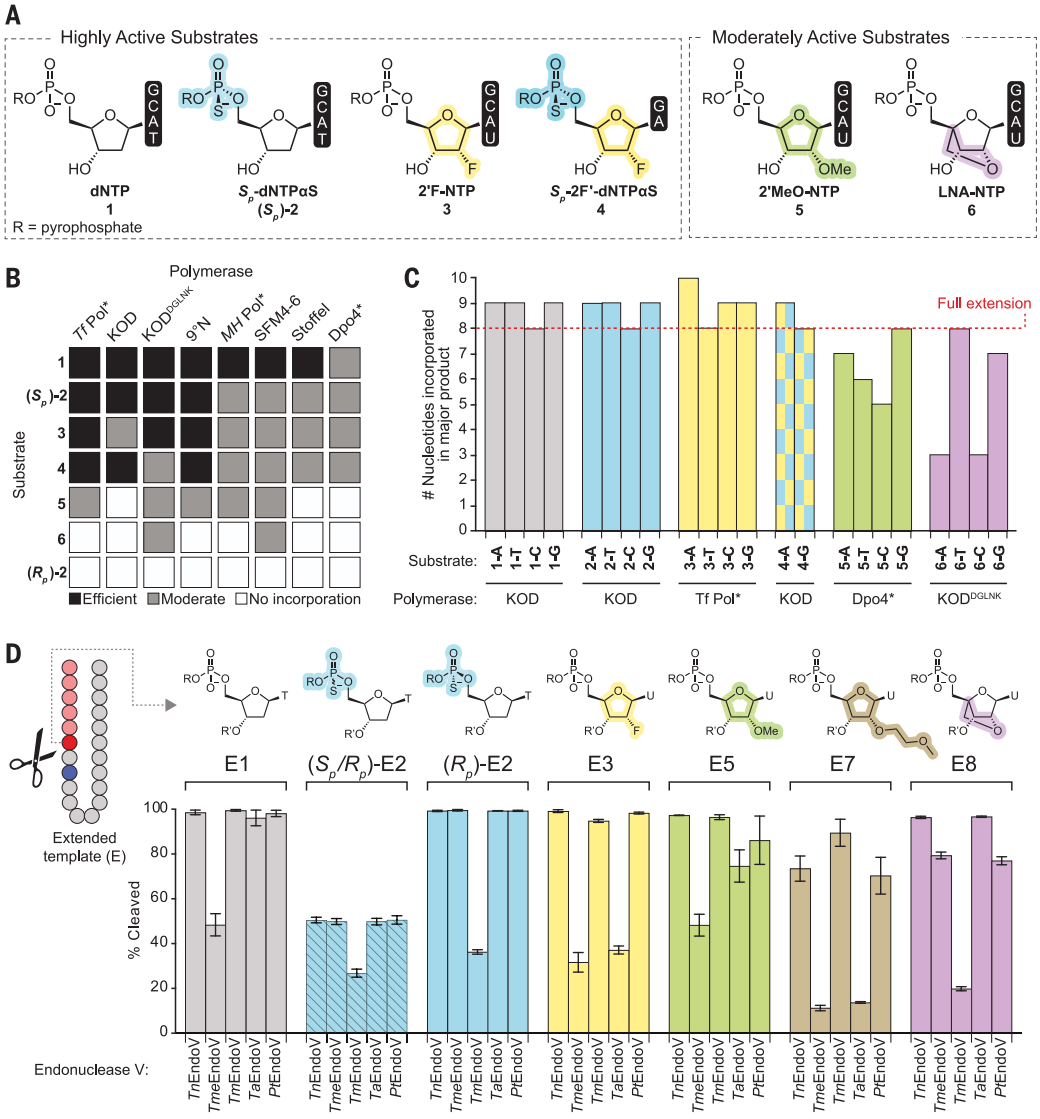
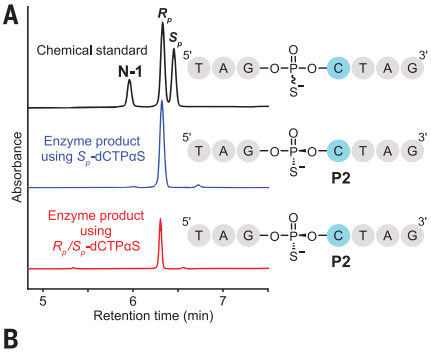


Fig. 3. Substrate profiling of polymerases and endonucleases. (**A**) NTP substrates accepted by polymerases from the panel. (**B**) Heat map showing polymerase activity with a mixture of three unmodified dNTPs and one modified NTP. Polymerases that accept all four nucleobases of a specified NTP and yield the extended product in >50% conversion are shown in black. Polymerases that extend sequences with at least one nucleobase of a specified NTP are shown in gray. (**C**) Polymerase-catalyzed extension of templates that encode eight copies of a single nucleobase. The graph shows the number of nucleotides incorporated in the major product as determined by LC-MS analysis (fig S8). (**D**) Endonuclease-catalyzed cleavage of modified extended templates E1, E2, E5, E7, and E8 after 2 hours of incubation and E3 after 4 hours of incubation. Cleavage of an all-LNA modified 18-mer (E6) gave low conversion and was omitted from the graph (fig. S9). Extended template E8 contains a single LNA 3′ of the cleavage site and was successfully hydrolyzed. Extended template sequences are presented in table S3. Error bars represent the standard deviations of reactions performed in triplicate. Biotransformations were analyzed by PAGE, and product identity was confirmed by LC-MS analysis (figs. S9 and S10).

liquid chromatography–mass spectrometry (LC-MS) analysis shows that selected polymerases were highly active toward substrates **1** to **4**, with reactions proceeding to completion within 1 hour (fig. S8). Although single-nucleotide overextension products are often observed with these substrates, overextension is notably reduced in cascade processes in the presence of EndoV endonucleases (see below). Reactions with each of the four 2′-MeO (**5**)– and LNA (**6**)–modified NTPs gave rise to mixtures of partially extended products (fig. S8), consistent with modest polymerase activity toward these modifications. We anticipate that efficient production of these sequences can be achieved through further enzyme engineering.

We next evaluated the panel of EndoV endonucleases for activity toward extended templates containing 2′-fluoro (E3), 2′-methoxy (E5), and 2′-O-methoxyethyl (E7) ribose modifications; LNAs (E6 and E8); and PS linkages (E2) (figs. S9 and S10). Hydrolysis of E3, E5, and E7 proceeded to >90% conversion, and endonuclease *Tn*EndoV, which tolerates elevated NTP concentrations, performed consistently well across the substrate range (Fig. 3D). Although cleavage of an LNA-modified 18mer (E6) was sluggish, sequences containing a single LNA 3′ of the cleavage site (E8) were efficiently hydrolyzed. The modest activity observed with E6 likely arises from distorted secondary structure of the LNA-DNA duplex (*38*). However, therapeutic oligonucleotides rarely contain more than three sequential LNAs in their sequence. Endonuclease-catalyzed cleavage of stereorandom PS-modified DNA R_p/S_p-(E2) stalled at 50% conversion. The most likely explanation for this behavior is that the endonucleases are sensitive to the stereochemistry at the phosphorothioate center undergoing hydrolysis. To test this hypothesis, we synthesized a stereodefined extended template using KOD polymerase and S_p-dNTPαS. Several polymerases have been shown to incorporate S_p-dNTPαS into oligonucleotides with inversion of stereochemistry (*39*–*42*), meaning that the oligonucleotide product is expected to contain R_p linkages. In this case, endonuclease *Tn*EndoV–mediated cleavage proceeded to completion (Fig. 3D and fig. S11), suggesting that endonucleases are R_p-selective and can be used to cleave enzymatically synthesized products.

Next, we turned our attention to one-pot cascade reactions to produce modified sequences. First, we took advantage of polymerase specificity to produce 6-mers and 7-mers containing a single PS linkage as single diastereomers [>99% diastereomeric excess (d.e.)]. These sequences can be produced by using either pure S_p-dNTPαS or R_p/S_p-dNTPαS mixtures (Fig. 4A and fig. S12). To highlight the versatility of our approach, we successfully amplified a range of 8-nt products (P2 to P16) (Fig. 4B and fig. S13) with different base sequences and a variety of 2′F, 2′-MeO, R_p-PS, R_p-2′F-PS, and LNA modifications. In favorable cases, we achieved up to 238 cycles of template extension and product cleavage (238 moles of product per mole of template), which consumed 76% of the NTP starting materials. The NTP concentrations that we used in these reactions are lower than those used earlier with unmodified dNTPs owing to the lower concentrations of modified NTPs currently available from commercial suppliers.



	Products	Cycles	% conv.	Pol.	[NTPs] (mM)
P3	TCAGTCAG	190	38	KOD	16
P4	TCAGTCAG	42	34	KOD	4
P5	UCAGUCAG	49	39	*Tf*Pol*	4
P6	CAUGCAUG	37	30	KOD^{DGLNK}	4
P7	CCUAGGUA	56	45	KOD^{DGLNK}	4
P8	AUGCAUGC	30	24	Dpo4*	4
P9	TCAGTCAG	238	76	KOD	10
P10	TCAGTCAG	235	75	KOD	10
P11	UCAGUCAG	104	33	*Tf*Pol*	10
P12	TCAGTCAG	138	44	*Tf*Pol*	10
P13	TCAGTCAG	134	43	*Tf*Pol*	10
P14	TCAGTCAG	43	34	KOD	4
P15	UCAGUCAG	76	61	*Tf*Pol*	4
P16	UCAGUCAG	22	18	*Tf*Pol*	4
P17	TCAGTCAG	85	68	*Tf*Pol*	4

Fig. 4. Biocatalytic synthesis of oligonucleotide sequences. (**A**) Comparison of stereopure oligonucleotides produced by our enzyme cascade and either S_p-CTPαS or a mix of R_p/S_p-CTPαS to chemically synthesized standards. Cascade reactions were performed with KOD and *Tn*EndoV. (**B**) Oligonucleotides with varying sequence and chemical modifications were produced in one-pot biotransformations with the most suitable polymerase (Pol.) and *Tn*EndoV. For comparison, under optimized conditions, the unmodified product P1 was produced following 330 cycles (37% conv.). The percentage conversion (% conv.) describes the conversion of NTP starting material to product. Cycles refers to the number of moles of product formed per mole of template. Specific reaction conditions are presented in table S1.

We anticipate that substantially improved turnovers will be achieved by increasing NTP concentrations and through specific optimization of reaction conditions for the target sequence.

Production of clinically relevant oligonucleotides

To further showcase the utility of our methodology, we produced a range of clinically relevant oligonucleotide sequences (Fig. 5A and fig. S14). First, we synthesized prexigebersen, an unmodified liposome-incorporated oligonucleotide currently under evaluation in phase 2 clinical trials for the treatment of acute myeloid leukemia. Prexigebersen was produced in 0.44 mM (equivalent to 2.4 g/l), which is comparable to the product titers achieved in earlier experiments with P1, demonstrating the transferability of the technology to different sequences. We also produced selected fragments of the Food and Drug Administration–approved aptamer pegaptanib (2′-MeO and 2′-F modifications) and gapmers mipomersen (PS modifications) and inotersen (PS modifications). These fragments could potentially be integrated into convergent ligation-based oligonucleotide assembly strategies (*20*–*24*).

Finally, we synthesized the PS-modified oligonucleotide therapies fomivirsen, alicaforsen, trabedersen, and aganirsen as single stereoisomers. Subsequent production of fomivirsen, a 21-nt sequence that was the first approved antisense oligonucleotide therapy (Fig. 5B), on a preparative scale (0.75 ml) demonstrates how reaction intensification can deliver improved conversions. After concentration of commercial S_p-dNTPαS stocks and optimization of reaction parameters, the target sequence was produced (0.26 mM, equivalent to 1.7 g/l, 195 nmol) after 65 cycles of template extension and product cleavage, which correlates to consumption of 90% of the available S_p-dNTPαS starting materials. We achieved product isolation after protein precipitation, deoxyribonuclease (DNase)–catalyzed template degradation, and filtration steps to provide fomivirsen (0.73 mg, 108 nmol) as an all-R_p single stereoisomer in 87% purity (fig. S15). We compared the biophysical and biochemical properties of enzymatically synthesized fomivirsen to stereorandom material produced by chemical synthesis (*10*, *43*). Thermal denaturation studies showed that the all-R_p stereoisomer has higher binding affinity to the target RNA compared with the stereorandom mixture (T_m of 69° ± 0.8°C and 64° ± 0.3°C, respectively). Assays with ribonuclease H (RNase H) nonspecific endonucleases show that *Escherichia coli* RNase H activity is unaffected by oligonucleotide stereochemistry, whereas the human enzyme is about twofold more active with stereorandom fomivirsen compared with the all-R_p stereoisomer (fig. S16). Investigations into fomivirsen stability showed that

phosphodiesterase from *Crotalus adamanteus* preferentially degraded the all-R_p oligonucleotide, whereas stereorandom fomivirsen was more rapidly degraded by S1 nuclease from *Aspergillus oryzae* (fig. S17). A third endonuclease from *Staphylococcus aureus* showed comparable activity with either sample, showing that the rate of oligonucleotide degradation is strongly dependent on the specific nuclease.

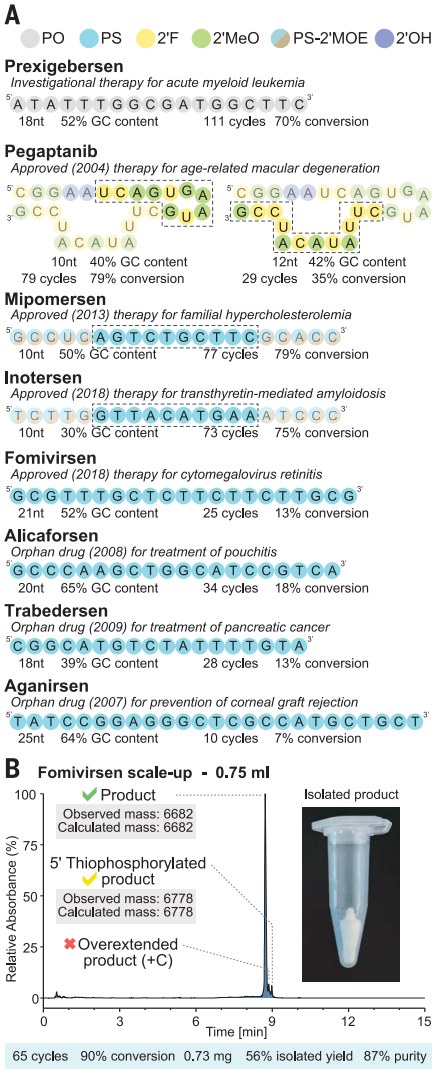


Fig. 5. Enzymatic synthesis of clinically relevant oligonucleotide sequences. (**A**) Oligonucleotide sequences were produced on an analytical scale. Mipomersen and inotersen contain cytosine and not 5′-methylcytosine because the required NTP monomers are not commercially available. The number of nucleotides (nt) in the sequences and percent GC content are provided. The percentage conversion describes the conversion of NTP starting material to product. Cycles refers to the number of moles of product formed per mole of template. The crude HPLC traces are shown in figure S14. Specific reaction conditions are presented in table S2. (**B**) Preparative-scale synthesis of fomivirsen as a single stereoisomer. LC-MS trace and image of the isolated product. Minor impurities include overextended product and 5′-thiophosphorylated product.

Conclusions

This study offers a distinct paradigm in therapeutic oligonucleotide manufacturing: Iterative rounds of chain extension, oxidation, capping, and deprotection are replaced by a single biotransformation. Our approach provides access to a diverse array of oligonucleotide sequences and modifications, which will be further expanded through the discovery and/or engineering of biocatalysts with extended substrate scope (*32*, *44*). For example, the development of polymerases with activity toward R_p-dNTPαS will enable production of libraries of stereochemically defined PS-modified oligonucleotides for biological evaluation, ultimately leading to therapeutics with improved safety and efficacy. Previous studies have shown that the stereopreference of polymerases and cyclic guanosine-adenosine synthase (cGAS) can be altered by exchanging the Mg^{2+} cofactor and protein engineering (*45*, *46*). Similarly, directed evolution will generate more robust biocatalysts specifically engineered to meet target process parameters. Translation of our approach into oligonucleotide manufacturing processes will also be facilitated by the development of scalable and low-cost routes to the required NTP building blocks. These NTPs will likely be accessed from the corresponding nucleoside, a common intermediate in phosphoramidite production, by using enzymatic phosphorylation methods that have been previously demonstrated at scale (*46*). Thus, the biocatalytic platform described here has the potential to transform the way that nucleic acid therapeutics are manufactured, addressing imminent scalability and sustainability challenges associated with the production of this emerging class of therapeutics.

REFERENCES AND NOTES

1. L. K. McKenzie, R. El-Khoury, J. D. Thorpe, M. J. Damha, M. Hollenstein, *Chem. Soc. Rev.* **50**, 5126–5164 (2021).
2. A. Khvorova, J. K. Watts, *Nat. Biotechnol.* **35**, 238–248 (2017).
3. X. Shen, D. R. Corey, *Nucleic Acids Res.* **46**, 1584–1600 (2018).
4. J. A. Kulkarni *et al.*, *Nat. Nanotechnol.* **16**, 630–643 (2021).
5. F. J. Raal *et al.*, *N. Engl. J. Med.* **382**, 1520–1530 (2020).
6. S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **22**, 1859–1862 (1981).
7. J. F. Kim *et al.*, *Org. Process Res. Dev.* **20**, 1439–1452 (2016).
8. B. I. Andrews *et al.*, *J. Org. Chem.* **86**, 49–61 (2021).
9. U. Tedebark, A. Scozzari, O. Werbitzky, D. Capaldi, L. Holmberg, *Methods Mol. Biol.* **683**, 505–524 (2011).
10. N. Iwamoto *et al.*, *Nat. Biotechnol.* **35**, 845–851 (2017).
11. K. W. Knouse *et al.*, *Science* **361**, 1234–1238 (2018).
12. D. Xu *et al.*, *J. Am. Chem. Soc.* **142**, 5785–5792 (2020).
13. Y. Huang *et al.*, *Science* **373**, 1265–1270 (2021).
14. A. L. Featherston *et al.*, *Science* **371**, 702–707 (2021).
15. S. Palluk *et al.*, *Nat. Biotechnol.* **36**, 645–650 (2018).
16. X. Lu *et al.*, *ACS Catal.* **12**, 2988–2997 (2022).
17. M. A. Jensen, R. W. Davis, *Biochemistry* **57**, 1821–1832 (2018).
18. M. Eisenstein, *Nat. Biotechnol.* **38**, 1113–1115 (2020).
19. J. Ashley, I. G. Potts, F. J. Olorunniji, *ChemBioChem* **24**, e202200510 (2023).
20. K. J. D. Van Giesen, M. J. Thompson, Q. Meng, S. L. Lovelock, *JACS Au* **3**, 13–24 (2022).
21. S. Paul *et al*, *ACS Chem. Biol.* 10.1021/acschembio.3c00071 (2023).
22. X. Zhou *et al.*, *J. Org. Chem.* **87**, 2087–2110 (2022).
23. A. Crameri, D. G. Tew, Novel processes for the production of oligonucleotides, International Publication Number WO2019/121500 A1 (2019).
24. D. Takahashi, Y. Hagiwara, S. Kajimoto, M. Konishi, Method for producing modified oligonucleotide including complementary sequence, International Publication Number WO2020/171092 (2020).
25. A. Joneja, X. Huang, *Anal. Biochem.* **414**, 58–69 (2011).
26. Y. Zhao, F. Chen, Q. Li, L. Wang, C. Fan, *Chem. Rev.* **115**, 12491–12545 (2015).
27. T. Chen, F. E. Romesberg, *Biochemistry* **56**, 5227–5228 (2017).
28. X. Wang, D. Gou, S. Y. Xu, *PLOS ONE* **5**, e8430 (2010).
29. E. Bayer, M. Mutter, *Nature* **237**, 512–513 (1972).
30. G. M. Bonora, C. L. Scremin, F. P. Colonna, A. Garbesi, *Nucleic Acids Res.* **18**, 3155–3159 (1990).
31. M. C. de Koning *et al.*, *Org. Process Res. Dev.* **10**, 1238–1245 (2006).
32. T. Chen *et al.*, *Nat. Chem.* **8**, 556–562 (2016).
33. H. Hoshino, Y. Kasahara, M. Kuwahara, S. Obika, *J. Am. Chem. Soc.* **142**, 21530–21537 (2020).
34. N. Freund *et al.*, *Nat. Chem.* **15**, 91–100 (2023).
35. A. P. Mehta *et al.*, *J. Am. Chem. Soc.* **138**, 14230–14233 (2016).
36. A. M. Leconte *et al.*, *J. Am. Chem. Soc.* **130**, 2336–2343 (2008).
37. S. Hoshika *et al.*, *Science* **363**, 884–887 (2019).
38. A. Eichert *et al.*, *Nucleic Acids Res.* **38**, 6729–6736 (2010).
39. P. J. Romaniuk, F. Eckstein, *J. Biol. Chem.* **257**, 7684–7688 (1982).
40. P. M. Burgers, F. Eckstein, *J. Biol. Chem.* **254**, 6889–6893 (1979).
41. R. S. Brody *et al.*, *Biochemistry* **21**, 2570–2572 (1982).
42. F. Eckstein, *Angew. Chem. Int. Ed.* **22**, 423–439 (1983).
43. W. B. Wan *et al.*, *Nucleic Acids Res.* **42**, 13456–13468 (2014).
44. F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4552–4557 (2001).
45. J. Liu, M. D. Tsai, *Biochemistry* **40**, 9014–9022 (2001).
46. J. A. McIntosh *et al.*, *Nature* **603**, 439–444 (2022).

ACKNOWLEDGMENTS

We thank A. Green for helpful discussions and R. Sung (Manchester Institute of Biotechnology) for assistance with ultraperformance liquid chromatography (UPLC) method development. **Funding:** UK Research and Innovation Council (Future Leader Fellowship MR/T041722/1) (S.L.L.); The Faculty of Science and Engineering (University of Manchester) for the award of a Presidential Fellowship (S.L.L.); Engineering and Physical Science Research Council [Integrated Catalysis (iCAT) CDT EP/ S023755/1] (E.R.M.). **Author contributions:** Conceptualization: S.L.L.; Methodology: E.R.M., R.O., and F.N.; Investigation: E.R.M., R.O., F.N., and R.S.; Formal analysis: E.R.M., R.O., and F.N.; Visualization: E.R.M., R.O., and R.S.; Funding acquisition: S.L.L.; Project administration: S.L.L.; Supervision: S.L.L.; Writing – original draft: S.L.L., E.R.M., and R.O.; Writing – review and editing: S.L.L., E.R.M., and R.O. **Competing interests:** S.L.L. has filed an international patent describing the application of the enzyme cascade for production of therapeutic oligonucleotides (WO2023/041931). **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add5892
Materials and Methods
Figs. S1 to S17
Tables S1 to S3
Protein Sequences
References (*47*–*49*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 21 June 2022; resubmitted 1 March 2023
Accepted 12 May 2023
10.1126/science.add5892

CONSERVATION

Widespread diversity deficits of coral reef sharks and rays

Colin A. Simpfendorfer[1,2]*, Michael R. Heithaus[3], Michelle R. Heupel[2,4], M. Aaron MacNeil[5], Mark Meekan[6], Euan Harvey[7], C. Samantha Sherman[1,8], Leanne M. Currey-Randall[4], Jordan S. Goetze[9,10], Jeremy J. Kiszka[3], Matthew J. Rees[6,11], Conrad W. Speed[6], Vinay Udyawer[12], Mark E. Bond[3], Kathryn I. Flowers[3,13], Gina M. Clementi[3], Jasmine Valentin-Albanese[14], M. Shiham Adam[15], Khadeeja Ali[3,16], Jacob Asher[17], Eva Aylagas[17], Océane Beaufort[18], Cecilie Benjamin[19], Anthony T. F. Bernard[20,21], Michael L. Berumen[22], Stacy Bierwagen[4], Chico Birrell[23], Erika Bonnema[3], Rosalind M. K. Bown[24], Edward J. Brooks[25], J. Jed Brown[26], Dayne Buddo[27], Patrick J. Burke[28,29], Camila Cáceres[3], Marta Cambra[30,31], Diego Cardeñosa[3], Jeffrey C. Carrier[32], Sara Casareto[3], Jennifer E. Caselle[33], Venkatesh Charloo[34], Joshua E. Cinner[35], Thomas Claverie[36], Eric E. G. Clua[37,38], Jesse E. M. Cochran[22], Neil Cook[39,40], Jessica E. Cramp[41,42], Brooke M. D'Alberto[1,43], Martin de Graaf[44], Mareike C. Dornhege[45], Mario Espinoza[30,31], Andy Estep[46], Lanya Fanovich[40], Naomi F. Farabaugh[3], Daniel Fernando[24], Carlos E. L. Ferreira[47], Candace Y. A. Fields[3,25], Anna L. Flam[48], Camilla Floros[49,50], Virginia Fourqurean[51,52], Laura Gajdzik[22,53], Laura García Barcia[3], Ricardo Garla[54,55], Kirk Gastrich[3], Lachlan George[2], Tommaso Giarrizzo[56,57], Rory Graham[5], Tristan L. Guttridge[59,60], Valerie Hagan[13], Royale S. Hardenstine[16,22], Stephen M. Heck[13], Aaron C. Henderson[61], Patricia Heithaus[3], Heidi Hertler[61], Mauricio Hoyos Padilla[62,63], Robert E. Hueter[64,65], Rima W. Jabado[1,66], Jean-Christophe Joyeux[67], Vanessa Jaiteh[68,69], Mohini Johnson[70], Stacy D. Jupiter[71], Muslimin Kaimuddin[72,70], Devanshi Kasana[3], Megan Kelley[3], Steven T. Kessel[73], Benedict Kiilu[74], Taratau Kirata[75]†, Baraka Kuguru[76], Fabian Kyne[77], Tim Langlois[78,79], Frida Lara[80,81], Jaedon Lawe[82], Elodie J. I. Lédée[1], Steve Lindfield[83], Andrea Luna-Acosta[84], Jade Q. Maggs[85], B. Mabel Manjaji-Matsumoto[86], Andrea Marshall[87], Lucy Martin[88], Daniel Mateos-Molina[89,90], Philip Matich[60], Erin McCombs[91], Ashlie McIvor[22,92], Dianne McLean[6,93], Llewelyn Meggs[82], Stephen Moore[1], Sushmita Mukherji[1,2], Ryan Murray[94], Stephen J. Newman[95], Josep Nogués[88], Clay Obota[96,97], Domingo Ochavillo[98], Owen O'Shea[99,100], Kennedy E. Osuka[96,101], Yannis P. Papastamatiou[3], Nishan Perera[23], Bradley Peterson[14], Caio R. Pimentel[67,102], Fabián Pina-Amargós[103,104], Hudson T. Pinheiro[105], Alessandro Ponzo[106], Andhika Prasetyo[107], L. M. Sjamsul Quamar[108], Jessica R. Quinlan[3], José Amorim Reis-Filho[109], Hector Ruiz[110], Alexei Ruiz-Abierno[104], Enric Sala[111], Pelayo Salinas-de-León[112,113], Melita A. Samoilys[96,114], William R. Sample[3], Michelle Schärer-Umpierre[110], Audrey M. Schlaff[1], Kurt Schmid[55,115], Sara N. Schoen[3], Nikola Simpson[116], Adam N. H. Smith[117], Julia L. Y. Spaet[118], Lauren Sparks[119], Twan Stoffers[120], Akshay Tanna[24], Rubén Torres[121], Michael J. Travers[95], Maurits van Zinnicq Bergmann[3,58], Laurent Vigliola[122], Juney Ward[123], Joseph D. Warren[14], Alexandra M. Watts[47,124], Colin K. Wen[125], Elizabeth R. Whitman[3], Aaron J. Wirsing[126], Aljoscha Wothke[40], Esteban Zarza-González[127,128], Demian D. Chapman[3,60]

A global survey of coral reefs reveals that overfishing is driving resident shark species toward extinction, causing diversity deficits in reef elasmobranch (shark and ray) assemblages. Our species-level analysis revealed global declines of 60 to 73% for five common resident reef shark species and that individual shark species were not detected at 34 to 47% of surveyed reefs. As reefs become more shark-depleted, rays begin to dominate assemblages. Shark-dominated assemblages persist in wealthy nations with strong governance and in highly protected areas, whereas poverty, weak governance, and a lack of shark management are associated with depauperate assemblages mainly composed of rays. Without action to address these diversity deficits, loss of ecological function and ecosystem services will increasingly affect human communities.

Coral reef ecosystems are under increasing pressure from human activities—including intense fishing, degraded water quality, and climate change (*1*, *2*)—that threaten species supporting a wide range of ecosystem functions (*3*). Sharks and rays (hereafter "elasmobranchs") have diverse roles on coral reefs as predators and prey across multiple trophic levels and in the cycling and movement of nutrients (*3*–*5*). Recent evidence indicates that overfishing has driven sharks toward functional extinction on many reefs. In a global survey, sharks were not observed on nearly 20% of reefs surveyed (*6*). Yet until recently, reef shark species were listed in lower risk extinction categories by the International Union for the Conservation of Nature (IUCN). With ~37% of all elasmobranch species threatened with extinction (*7*), a key question for coral reef ecosystems lies in understanding the global extent of species loss in elasmobranch assemblages. We characterized elasmobranch assemblage structure on coral reefs across a gradient of human pressures to estimate the local depletion and global extinction risk of the most common reef species, revealing the human and environmental factors that influence assemblage structure and that lead to a deficit in predator diversity that could affect reef ecological functioning.

To understand the extent of the reef elasmobranch diversity deficit, we surveyed 391 coral reefs in 67 nations and territories using 22,756 baited remote underwater video stations (BRUVS). We examined reef-level species richness, species composition of elasmobranch assemblages, and species relative abundance (MaxN; the maximum number of each species observed in a single frame of each 60-min deployment then averaged across all deployments on one reef) (*8*). We examined how elasmobranch species assemblages changed in response to human pressures, using unweighted pair group with arithmetic mean (UPGMA) clustering to identify reefs with the most similar assemblages (*8*). We then compared these clusters with estimated depletion of key resident elasmobranch species at the reef level and examined whether socioeconomic, management, or environmental factors could predict cluster membership, using linear discriminant analysis. Reef-level depletion was estimated by dividing the observed mean MaxN of a species at individual reefs by a model-estimated baseline abundance (without human pressures) for each sampling site (a small group of closely associated reefs) and subtracting this value from 1. Baseline abundance (also expressed as MaxN) was estimated from a general linear model relating observed MaxN to sampling site, human pressure [represented by total market gravity, the size and travel time to human markets (*2*)], and marine protected area (MPA) status [closed to all fishing, open to fishing, or restricted (some fishing but with restrictions)]. The baseline was estimated by setting all parameters to those expected at a site with no human pressure (gravity to the minimum for an ocean basin and protection status to closed) (*8*).

Sampling identified 104 distinct elasmobranch species or species complexes (table S1), representing more than 77% of elasmobranch species known to occur on coral reefs at some point during their lives (*9*). More than half (n = 53) of the species were rarely observed, with 10 or fewer sightings. We estimated reef-level depletion for the nine most commonly occurring species of shark [n = 5; Caribbean reef sharks (*Carcharhinus perezi*) and nurse sharks (*Ginglymostoma cirratum*) in the Atlantic; grey reef sharks (*Carcharhinus amblyrhynchos*),

Affiliations are listed at the end of this paper.
*Corresponding author. Email: colin.simpfendorfer@jcu.edu.au
†Deceased.

blacktip reef sharks (*Carcharhinus melanopterus*), and whitetip reef sharks (*Triaenodon obesus*) in the Indo-Pacific] and rays [n = 4; yellow stingrays (*Urobatis jamaicensis*) and southern stingrays (*Hypanus americanus*) in the Atlantic; blue spotted mask rays (*Neotrygon* spp.) and blue spotted ribbontail rays (*Taeniura lymma* and *Taeniura lessoni*) in the Indo-Pacific]. The Galapagos shark was excluded from estimates of global depletion because sampling only covered a relatively small proportion of its range, but the results for this species were broadly similar. The nine key resident species represented 77.7% of all elasmobranchs observed in the study and are those that serve important ecological roles (*10*) and contribute the most to, and underpin, livelihoods through fishing (*11*) and dive tourism (*12*).

We found that mean depletion of five key resident reef sharks on individual reefs ranged from 100% depletion (none observed) to 0% (no depletion), averaging 62.8% (Fig. 1A). Mean depletion of key resident reef sharks followed the overall decline in elasmobranch abundance as measured with MaxN (Fig. 1B), decreased as the fraction of the elasmobranch assemblage comprised of sharks decreased (Fig. 1C), and showed little change across a range of elasmobranch species richness (Fig. 1D); these patterns were generally consistent between ocean basins. Across the range of depletion, five main clusters of reefs were identified in the Atlantic, and eight were identified in the Indo-Pacific (Figs. 2 and 3), including at least one cluster in each ocean basin (cluster 1 in the Atlantic and cluster 2 in the Indo-Pacific) having shark populations in a relatively intact state, with low levels of depletion of the five main resident reef shark species (Caribbean reef and nurse sharks in the Atlantic; grey reef, blacktip reef, and whitetip reef sharks in the Indo-Pacific) (*8*). Remaining clusters represented assemblages with increasing depletion of resident shark species and greater proportions of the overall elasmobranch assemblage represented by rays (Figs. 2C and 3B). Both ocean basins show a similar transition through these assemblages as key resident shark species became depleted. The four key ray species (yellow and southern stingrays in the Atlantic; blue spotted mask and blue spotted ribbontail rays in the Indo-Pacific) increased only with depletion of one or more resident reef shark species, with rays dominating in the most shark-depleted areas. These predictable changes in assemblage provide the ability to infer the status of reef shark populations, and the level of human pressure they are experiencing, in future surveys.

Elasmobranch species assemblage clusters on reefs in both basins were significantly related to certain socioeconomic and management factors, with linear discriminant analysis (LDA) accounting for ~85% of variance between clusters (tables S2 and S3). Important socioeconomic factors included the Human Development Index (an index of a nation's level of education, life expectancy and standard of living) and Voice and Accountability Index (an index of the extent to which people in each nation can participate in governance, free expression, free media, and free association). Important management factors were whether the reef occurred in a marine protected area (MPA) or whether a reef was within a nation where all targeted shark fishing and trade is prohibited, known as a "shark sanctuary." Given that shark sanctuaries have largely been implemented in nations in which fishing for sharks was limited for economic or cultural reasons (*6*), their effectiveness as tools for recovering reef shark populations remains an open question. Total market gravity was more important in the Indo-Pacific than the Atlantic, possibly because remote reefs (>4 hours travel time from human settlements)

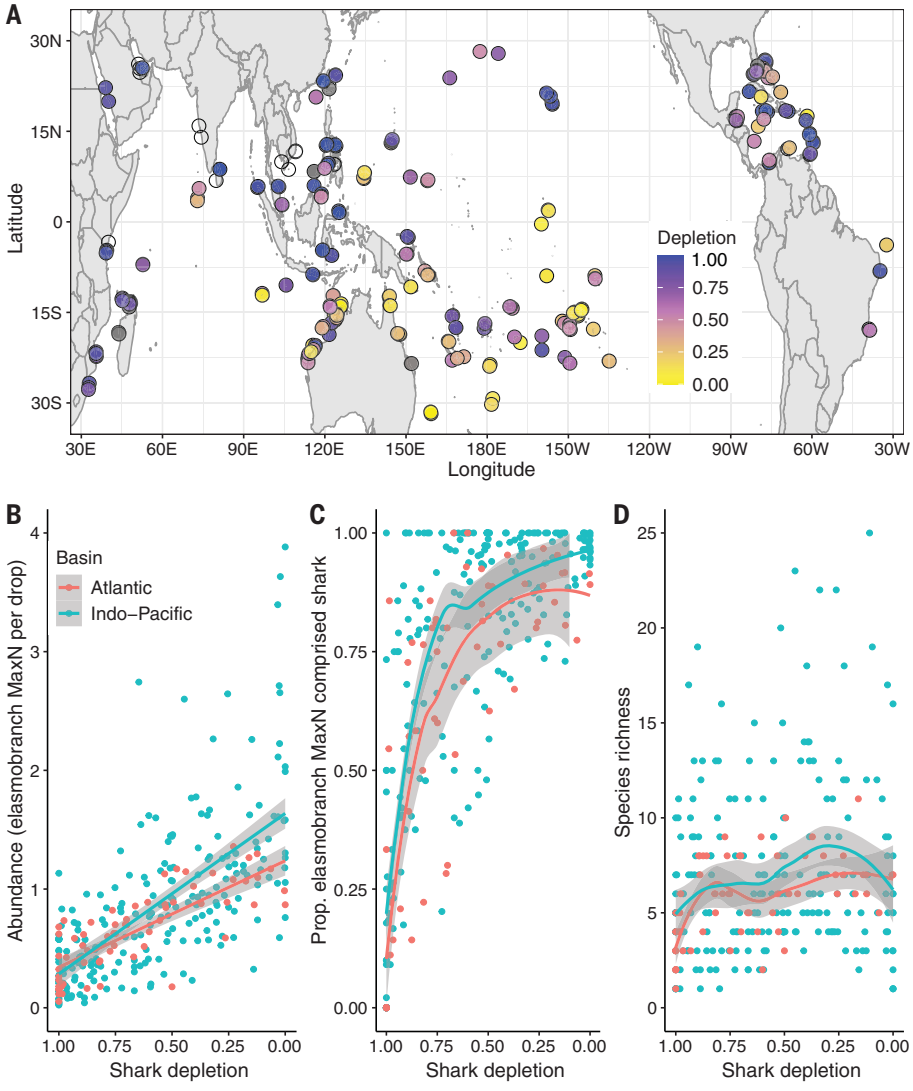


Fig. 1. The global decline of coral reef elasmobranchs. (**A**) Reef-scale estimates of depletion of resident coral reef shark species. Depletion is proportion of unfished population lost, represented as the measured MaxN as a proportion of MaxN in an unfished state (gravity, lowest in basin; MPA status, closed) (*8*). Open circles indicate no sharks or rays were observed; gray circles indicate none of the resident shark species used to calculate mean depletion were present. (**B**) Relationship between depletion of resident shark species and MaxN by ocean basin. (**C**) Relationship between depletion of resident shark species and the proportion of elasmobranch MaxN that comprised shark, demonstrating the transition from shark- to ray-dominated assemblages. (**D**) Relationship between depletion of resident shark species and species richness.

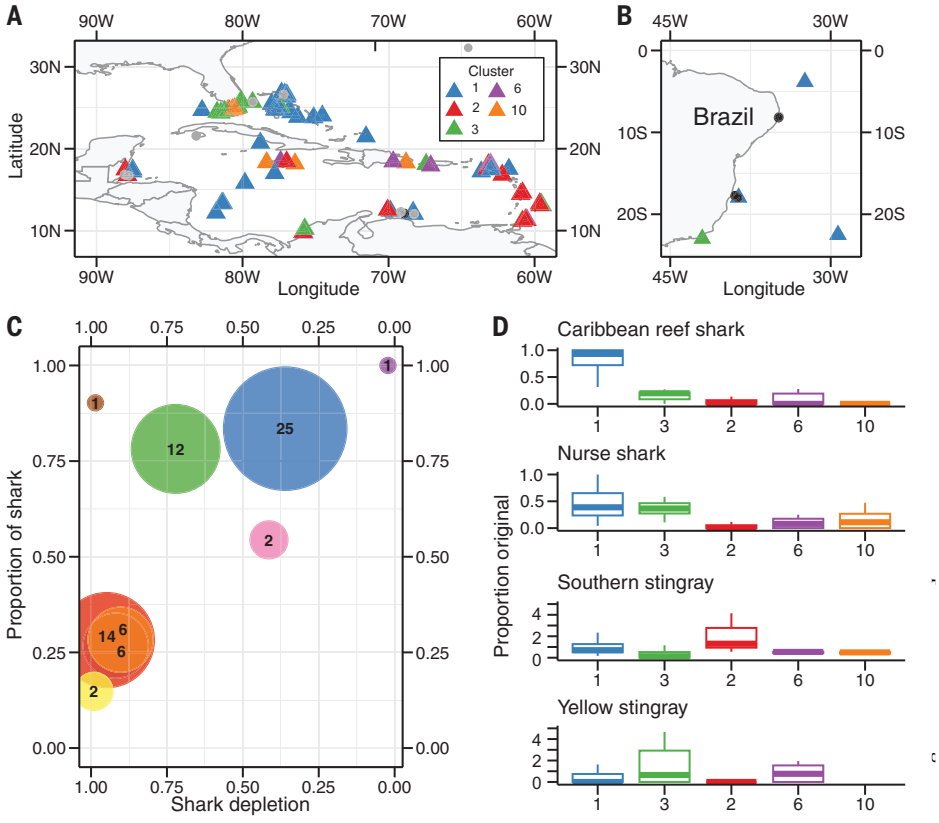


Fig. 2. Structure of shark and ray assemblages on Atlantic coral reefs. (**A** and **B**) Clusters of reefs with similar species composition from UPGMA clustering of 106 reefs in the Atlantic basin based on a global set of 31 coral reef–associated species. Five main clusters, representing 87.0% of reefs, were identified. Their locations are indicated with colored triangles. Reefs with minor clusters are indicated with gray dots (n = 7). Reefs where no elasmobranchs were observed are indicated with black dots (n = 5). (**C**) Regime plot showing all species assemblage clusters as a function of the mean depletion of the resident reef shark species (Caribbean reef and nurse sharks) and the proportion of all observed elasmobranchs that were sharks. Size of points (and numbers) indicate the number of reefs in each cluster, and colors indicate cluster identity as per (A). (**D**) Population level relative to original levels of four resident reef species in each of the five main clusters. Proportion of original level = 1 − depletion. Horizontal lines indicate mean, boxes indicates 25 to 75 percentile, and whiskers indicate 95% confidence interval.

are relatively rare in the Atlantic compared with the Indo-Pacific (fig. S1) (*13*). Environmental factors (coral cover and relief) had little influence in predicting cluster membership. Elasmobranch assemblage structure on coral reefs in both the Atlantic and Indo-Pacific are therefore mainly driven by management and socioeconomic factors, with shark-dominated assemblages more likely to occur in wealthy, well-governed nations and in highly protected areas or shark sanctuaries, whereas poverty, limited governance, and a lack of shark protection are associated with assemblages mainly composed of rays.

To further characterize the diversity deficits that underpin these assemblage differences, we compared species observations in our BRUVS with their historical ranges drawn from published literature, including historical accounts, and found that sharks were not detected at 13.6% of reefs (19 Atlantic and 34 Indo-Pacific), whereas rays were not detected at 21.5% of reefs (10 Atlantic and 74 Indo-Pacific); both groups were not detected at 6.6% of reefs surveyed (5 Atlantic and 19 Indo-Pacific). At the species level, absences were severe. On the basis of their known historic distribution, deficits were 46.9% of reefs (112 of 246) for blacktip reef sharks, 41.3% (31 of 75) for Caribbean reef sharks, 40.8% (102 of 250) for grey reef sharks, 36.2% (89 of 246) for whitetip reef sharks, and 34.7% (n = 26 of 75) for nurse sharks (fig. S2). Among rays, deficits were even more stark: 78.9% (75 of 95) for yellow stingray, 62.8% (81 of 129) for blue spotted ribbontail rays, and 55.6% (79 of 142) for blue spotted maskrays. An exception was the southern stingray, which was not detected at only 19.8% (n = 20 of 101) of expected reefs in the Atlantic. A failure to detect rays may not always indicate absence because they are often cryptic and therefore missed on BRUVS, especially when sharks are present (*14*). Collectively, these diversity deficits show that elasmobranch loss on coral reefs is more extensive than previously demonstrated, with widespread losses of key species across many of the world's coral reefs, especially in Asia, eastern Africa, continental South America, and the central-eastern Caribbean.

Previous estimates of the status of reef shark and ray species have been geographically limited, varying among surveyed reefs from very high abundances (*15*) to local extinction (*16*). This disparity has made it difficult to assess the global status of individual species. Therefore, we used our estimates of reef-level depletion to estimate the global depletion and extinction risk of the most common resident reef sharks (five species) and rays (four species). Mean and standard error reef-level depletion was calculated within jurisdictions (nations or remote territories) and used to produce confidence intervals for jurisdictional depletion levels. To estimate an overall global depletion level by species, we weighted the jurisdictional depletion by the percentage of the world's coral reefs in their waters and produced a weighted global mean depletion (*8*). Extinction risk was estimated by comparing proportional global depletion to the criteria for the IUCN Red List A2 (population decline) category (*17*), assuming that the decline had occurred in the past three generations (29 to 90 years). In IUCN assessments before the availability of this global survey, all reef-resident shark species were considered at lower risk of extinction (Near Threatened) (*18*). Grey reef shark had the highest level of global decline [69.8% ± 1 standard error (SE) 62.6 to 77.1], followed by nurse shark (68.6% ± 49.7 to 87.4), Caribbean reef shark (64.8% ± 42.0 to 87.5), blacktip reef shark (64.5% ± 58.7 to 70.4), and whitetip reef shark (60.4% ± 51.2 to 70.2) (Fig. 4). The estimated declines of resident species of reef sharks met the IUCN Red List criteria for Endangered. Population changes of rays were more variable, with increasing populations in some nations and declines in others (fig. S3), reflecting the compositional changes seen across our gradient of human pressures. When examined at the global level, no ray species examined met criteria for elevated extinction risk, which is consistent with current nonthreatened status of these species on the Red List.

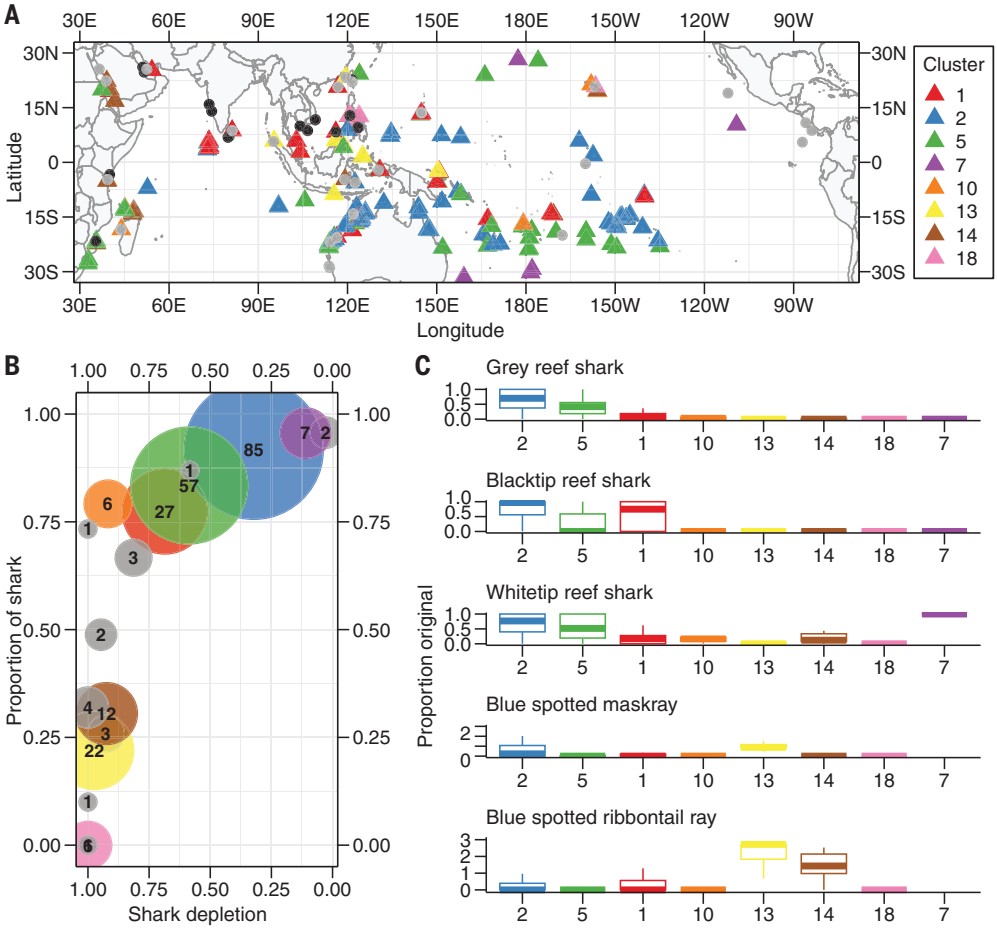


Fig. 3. Structure of shark and ray assemblages on Indo-Pacific coral reefs. (**A**) Clusters of reefs with similar species composition from UPGMA clustering of 285 reefs in the Indo-Pacific basin based on a global set of 31 coral reef-associated species. Eight main clusters, representing 82.1% of reefs, were identified. Their locations are indicated with colored triangles. Reefs with minor clusters are indicated with gray dots (n = 30). Reefs where no elasmobranchs were observed are indicated with black dots (n = 21). (**B**) Regime plot showing all species assemblage clusters as a function of the mean depletion of the resident species of reef shark (grey reef, blacktip reef, whitetip reef, and Galapagos sharks) and the proportion of all observed elasmobranchs that were sharks. Size of points (and numbers) indicate the number of reefs in each cluster, and colors indicate cluster identity as per (A); minor clusters are indicated in gray. (**C**) Population level relative to original levels of five core shark and ray species in each of the eight main species assemblage clusters. Proportion of original level = 1 − depletion. Horizontal lines indicate mean, boxes indicate 25 to 75 percentile, and whiskers indicate 95% confidence interval.

Our study of nations hosting ~90% of global reefs reveals that resident reef shark species are at much higher risk of extinction than previously thought. Local declines, shaped by human pressures that vary across ocean basins, have led to consistent changes in the structure of coral reef elasmobranch assemblages that may have profound effects on the broader ecosystem. The direct and indirect effects of fishing have driven shifts in species composition from shark-dominated to ray-dominated assemblages and ultimately the complete loss of sharks and rays at a small proportion (~ 7%) of reefs surveyed. In addition to changes in the structure of assemblages, all major resident shark species have declined to such levels that they qualify as Endangered by the IUCN Red List Criteria. These changes wrought on coral reef elasmobranch assemblages demonstrate the pervasiveness of fishing on coral reefs (*19*) and the substantial risks to reef-dependent human communities of continued overfishing. Elasmobranch species vary widely in their economic value, with some fished for subsistence, others fished for local or export markets, and others valued alive as tourism resources (*12*, *20*). Thus, understanding threats and conservation options for rebuilding populations at a species level will assist in developing effective management of coral reef elasmobranchs as part of a sustainable social-ecological system.

Although reef sharks are at considerable risk over broad spatial scales, our results show that declines at one reef will have little effect on reefs tens to hundreds of kilometers distant. Thus, despite populations being functionally extinct at the reef level, the potential to rebuild abundances remains relatively high if there are protected areas or strong fisheries management within a region (*6*). These source populations are present among many small oceanic islands where low human populations and the high cultural value of sharks has resulted in fishing levels that are below those seen elsewhere (*21*). MPAs also provide the opportunity to act as source populations; however, their designation alone is insufficient to deliver benefits. As others have observed (*22*), high compliance is required. We show that there are reefs in regions with widespread depletion of reef shark species that had metrics indicating that they are in a relatively healthy state compared with those around them. These

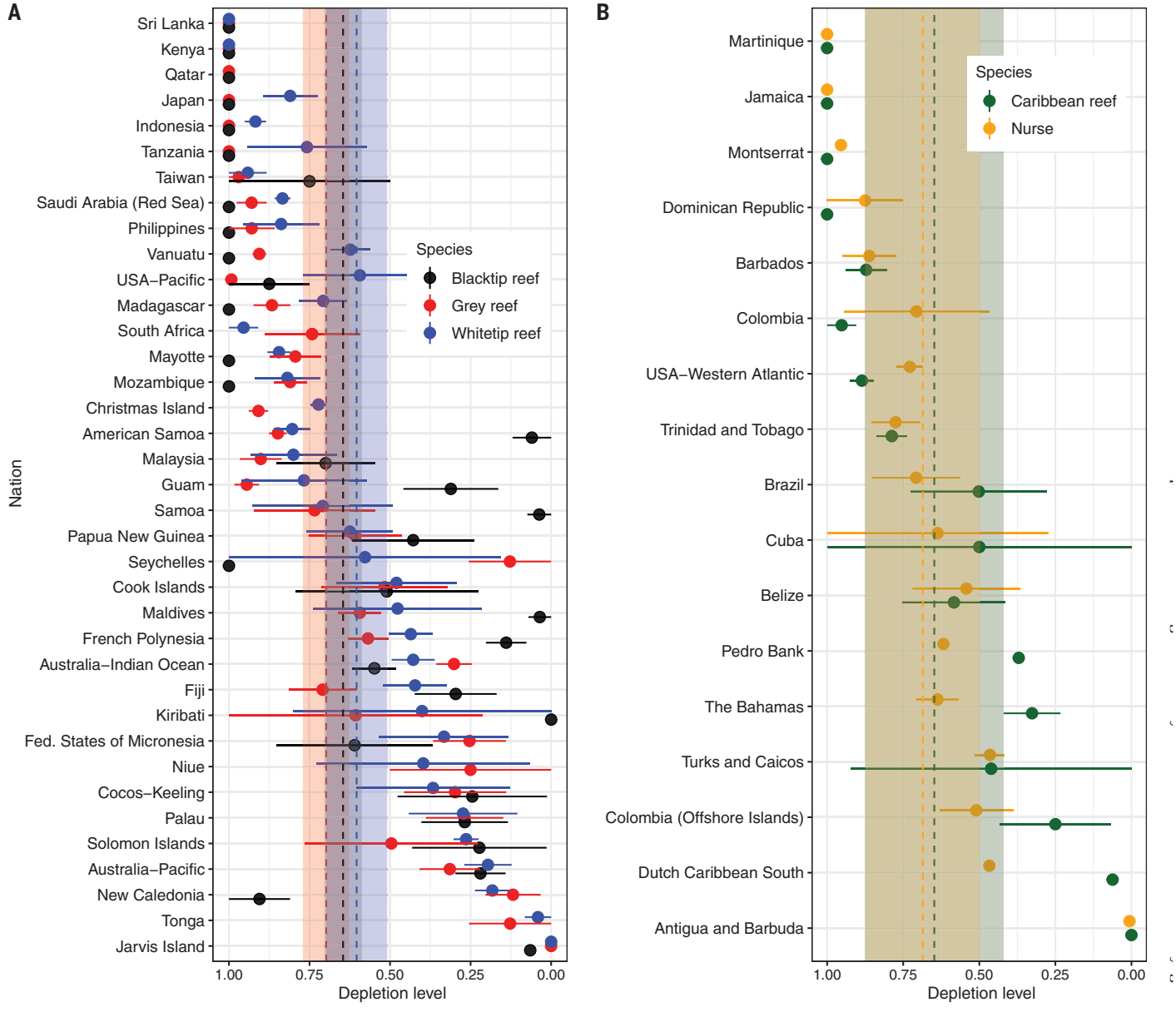


Fig. 4. Depletion of core coral reef shark species in the Indo-Pacific and Atlantic basins at national or near-national scale. (**A**) Indo-Pacific basin. (**B**) Atlantic basin. Depletion was calculated by comparing reef-level species MaxN values to unfished, estimated by using a linear model in which market gravity (a measure of the human pressure from population and access to reefs) was set to the ocean basin minimum and reef protected status was "closed" (no take MPA) (*8*). Reef-level depletion scores were modeled by nation and used to estimate a global level of depletion (vertical dashed lines) ± 1 standard error (shaded area) calculated by weighting national-level depletion by coral reef area (as a percent of global total coral reef area that occurs within the range of each shark species).

included Tubbataha (Philippines), Sipidan (Island Malaysia), Glover's Reef and Lighthouse Reef (Belize), and Misool (Indonesia); in all of these locations, there are programs to actively manage and enforce MPA regulations that are likely to account for these successes (*23*–*25*).

Multiple nations have strong management measures (such as spatial protections and/or fishing restrictions) in place that benefit reef species. This study builds the case that species-specific reef shark management provides the best way forward for conservation and rebuilding of reef sharks in places where they have declined, among nations with the desire and capacity to do so (*7*, *8*). Recent studies show that populations of reef sharks can rebound in under a decade if appropriate management strategies that reduce fishing pressure are in place (*26*). Although direct management is critical, local and national socioeconomic factors that affect the ability of nations to develop, implement, and enforce regulations, and the likelihood that fishers comply with regulations, will be critical to maintaining or rebuilding populations and diverse elasmobranch assemblages. If not addressed, pressures causing the shark and ray diversity deficits we outline will continue to result in a loss of species, ecological functions, and ecosystem services that support sustainable livelihoods for millions of people worldwide.

REFERENCES AND NOTES

1. T. P. Hughes, D. R. Bellwood, S. R. Connolly, H. V. Cornell, R. H. Karlson, *Curr. Biol.* **24**, 2946–2951 (2014).
2. J. E. Cinner *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6116–E6125 (2018).
3. G. Roff *et al*., *Trends Ecol. Evol.* **31**, 395–407 (2016).
4. K. I. Flowers, M. R. Heithaus, Y. P. Papastamatiou, *Fish Fish.* **22**, 105–127 (2021).
5. J. J. Williams, Y. P. Papastamatiou, J. E. Caselle, D. Bradley, D. M. P. Jacoby, Mobile marine predators: an understudied source of nutrients to coral reefs in an unfished atoll. *Proc. R. Soc. B Biol. Sci.* 10.1098/rspb.2017.2456 (2018).
6. M. A. MacNeil *et al*., *Nature* **583**, 801–806 (2020).
7. N. K. Dulvy *et al*., *Curr. Biol.* **31**, 4773–4787.e8 (2021).
8. Materials and methods are available as supplementary materials.
9. C. S. Sherman *et al*., *Nat. Commun.* **14**, 15 (2023).
10. M. R. Heupel, Y. P. Papastamatiou, M. Espinoza, M. E. Green, C. A. Simpfendorfer, *Front. Mar. Sci.* **6**, 12 (2019).
11. S. A. Appleyard, W. T. White, S. Vieira, B. Sabub, *Sci. Rep.* **8**, 6693 (2018).
12. A. M. Cisneros-Montemayor, M. Barnes-Mauthe, D. Al-Abdulrazzak, E. Navarro-Holm, U. R. Sumaila, *Oryx* **47**, 381–388 (2013).
13. T. R. McClanahan, *Mar. Policy* **119**, 104022 (2020).
14. C. S. Sherman, M. R. Heupel, S. K. Moore, A. Chin, C. A. Simpfendorfer, *Mar. Ecol. Prog. Ser.* **641**, 145–157 (2020).
15. J. Mourier *et al*., *Curr. Biol.* **26**, 2011–2016 (2016).
16. M. O. Nadon *et al*., *Conserv. Biol.* **26**, 493–503 (2012).
17. IUCN, *IUCN Red List categories and criteria, version 3.1* (IUCN, ed. 2, 2012).
18. N. K. Dulvy *et al*., *Aquat. Conserv.* **26**, 134–153 (2016).
19. T. D. Eddy *et al*., *One Earth* **4**, 1278–1285 (2021).
20. H. Booth, D. Squires, E. J. Milner-Gulland, *Ocean Coast. Manage.* **182**, 104994 (2019).
21. J. S. Goetze *et al*., *PLOS ONE* **13**, e0200960 (2018).
22. G. J. Edgar *et al*., *Nature* **506**, 216–220 (2014).
23. R. Murray *et al*., *J. Asia-Pac. Biodivers.* **12**, 49–56 (2019).
24. V. F. Jaiteh *et al*., *Front. Mar. Sci.* **3**, (2016).
25. G. Clementi *et al*., *Mar. Ecol. Prog. Ser.* **661**, 175–186 (2021).
26. C. W. Speed, M. Cappo, M. G. Meekan, *Biol. Conserv.* **220**, 308–319 (2018).
27. C. Simpfendorfer, Widespread diversity deficits of coral reef sharks and rays. Dryad (2023); https://doi.org/10.5061/dryad.qbzkh18h0.
28. C. Simpfendorfer, Widespread diversity deficits of coral reef sharks and rays. Zenodo (2023); https://doi.org/10.5281/zenodo.7030578.

ACKNOWLEDGMENTS

We thank our individual funders for country-specific deployments, whose contributions greatly enhanced the sampling coverage of the projects; all of the government permitting agencies that allowed us to work in their waters; and the Global FinPrint volunteers from Stony Brook University, Florida International University, James Cook University, the Aquarium of the Pacific, and Shedd Aquarium who watched the BRUVS footage.
Funding: Core funding for Global FinPrint was provided by the Paul G. Allen Family Foundation (to D.D.C. and M.R.Hei.). **Author contributions:** Conceptualization: D.D.C., M.R.Hei., C.A.S., M.R.Heu., M.A.M., M.M., and E.H. Methodology: D.D.C., M.R.Hei., C.A.S., M.R.Heu., M.A.M., M.M., and E.H. Investigation: All authors. Visualization: C.A.S. Funding acquisition: D.D.C. and M.R.Hei. Project administration: D.D.C., M.R.Hei., C.A.S., M.R.Heu., M.A.M., M.M., and E.H. Writing – original draft: C.A.S., D.D.C., M.R.Hei., M.R.Heu., M.A.M., M.M., E.H., and C.S.S. Writing – review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests **Data and materials availability:** Data files have been deposited in Dryad (*27*), and R script has been deposited in Zenodo (*28*). **License information:**

[1]College of Science and Engineering, James Cook University, Townsville, QLD, Australia. [2]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS, Australia. [3]Institute of Environment, Department of Biological Sciences, Florida International University, North Miami, FL, USA. [4]Australian Institute of Marine Science, Townsville, QLD, Australia. [5]Ocean Frontier Institute, Department of Biology, Dalhousie University, Halifax, NS, Canada. [6]Australian Institute of Marine Science, Perth, WA, Australia. [7]School of Molecular and Life Sciences, Curtin University, Bentley, WA, Australia. [8]Earth to Ocean Group, Biological Sciences, Simon Fraser University, Burnaby, BC, Canada. [9]School of Molecular and Life Sciences, Curtin University, Perth, WA, Australia. [10]Marine Science Program, Biodiversity and Conservation Science, Department of Biodiversity, Conservation and Attractions, Perth, WA, Australia. [11]Centre for Sustainable Ecosystems Solutions, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW, Australia. [12]Australian Institute of Marine Science, Darwin, NT, Australia. [13]Sharks and Rays Conservation Program, Mote Marine Laboratory, Sarasota, FL, USA. [14]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA. [15]International Pole and Line Foundation–Maldives, Malé, Republic of Maldives. [16]Maldives Marine Research Institute, Ministry of Fisheries, Marine Resources and Agriculture, Malé, Republic of Maldives. [17]Red Sea Global, Department of Environmental Protection and Regeneration, AlRaidah Digital City, Riyadh, Saudi Arabia. [18]Kap Natirel NGO, Fort l'Olive, Guadeloupe, France. [19]Mahonia Na Dari Research and Conservation Centre, Kimbe, Papua New Guinea. [20]South African Institute for Aquatic Biodiversity, National Research Foundation, Makhanda, South Africa. [21]Department of Zoology and Entomology, Rhodes University, Makhanda, South Africa. [22]Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [23]Marine Conservation, Madagascar Program, Wildlife Conservation Society, Antananarivo, Madagascar. [24]Blue Resources Trust, Colombo, Sri Lanka. [25]Cape Eleuthera Institute, Cape Eleuthera, Eleuthera, The Bahamas. [26]Center for Sustainable Development, College of Arts and Sciences, Qatar University, Doha, Qatar. [27]Georgia Aquarium–IUCN Center for Species Survival, Atlanta, GA, USA. [28]School of Natural Sciences, Macquarie University, Sydney, NSW, Australia. [29]Bimini Biological Field Station, Bimini, Bahama. [30]Centro de Investigación en Ciencias del Mar y Limnología, Universidad de Costa Rica, San José, Costa Rica. [31]MigraMar, Olema, CA, USA. [32]Department of Biology, Albion College, Albion, MI, USA. [33]Marine Science Institute, University of California Santa Barbara, Santa Barbara, CA, USA. [34]Coastal Impact, Goa, India. [35]College of Arts, Society, and Education, James Cook University, Townsville, QLD, Australia. [36]Centre Universitaire de Formation et de Recherche de Mayotte, Dembeni, France. [37]Paris Sciences Lettres, Centre de Recherche Insulaire et Observatoire de l'Environnement, Opunohu Bay, Papetoai, French Polynesia. [38]Laboratoires d'Excellence Corail, Ecole Pratique des Hautes Etudes, Perpignan, France. [39]School of Biosciences, Cardiff University, Cardiff, UK. [40]Environmental Research Institute Charlotteville, Charlotteville, Trinidad and Tobago. [41]Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD, Australia. [42]Sharks Pacific, Rarotonga, Cook Islands. [43]Oceans and Atmosphere, Commonwealth Scientific and Industrial Research Organization, Hobart, TAS, Australia. [44]Wageningen Marine Research, Wageningen University & Research, IJmuiden, Netherlands. [45]Graduate School for Global Environmental Studies, Sophia University, Tokyo, Japan. [46]Waitt Institute, La Jolla, CA, USA. [47]Reef Systems Ecology and Conservation Lab, Departamento de Biologia Marinha, Universidade Federal Fluminense, Rio de Janeiro, Brazil. [48]Marine Megafauna Foundation, Palm Beach, FL, USA. [49]Oceanographic Research Institute, Durban, South Africa. [50]TRAFFIC International, Cambridge, UK. [51]College of Arts, Science, and Education, Florida International University, North Miami, FL, USA. [52]Science Department, Georgia Jones-Ayers Middle School, Miami, FL, USA. [53]Division of Aquatic Resources, Department of Land and Natural Resources, Honolulu, HI, USA. [54]Centro de Biociências, Departmento de Botânica e Zoologia, Universidade Federal do Rio Grande do Norte, Brazil. [55]Beacon Development Company, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [56]Instituto de Ciencias do Mar, Universidade Federal do Ceará, Fortaleza, Brazil. [57]Grupo de Ecologia Aquática, Espaço Inovação do Parque de Ciência e Tecnologia Guamá, Guamá, Pará, Brazil. [58]Independent consultant, Hull, UK. [59]Bimini Biological Field Station Foundation, South Bimini, The Bahamas. [60]Saving the Blue, Cooper City, FL, USA. [61]The School for Field Studies, Center for Marine Resource Studies, South Caicos, Turks and Caicos Islands. [62]Pelagios Kakunjá, La Paz, Mexico. [63]Fins Attached, Colorado Springs, CO, USA. [64]Center for Shark Research, Mote Marine Laboratory, Sarasota, FL, USA. [65]OCEARCH, Park City, UT, USA. [66]Elasmo Project, Dubai, United Arab Emirates. [67]Departamento de Oceanografia e Ecologia, Universidade Federal do Espírito Santo, Vitória, Espírito Santo, Brazil. [68]Murdoch University, Murdoch, WA, Australia. [69]Centre for Development and Environment, University of Bern, Bern, Switzerland. [70]Operation Wallacea, Spilsby, Lincolnshire, UK. [71]Melanesia Program, Wildlife Conservation Society, Suva, Fiji. [72]Wasage Divers, Wakatobi & Buton, Southeast Sulawesi, Indonesia. [73]Daniel P. Haerther Center for Conservation and Research, John G. Shedd Aquarium, Chicago, IL, USA. [74]Kenya Fisheries Service, Mombasa, Kenya. [75]Ministry of Fisheries and Marine Resources, Kiritimati, Kiribati. [76]Tanzania Fisheries Research Institute, Dar Es Salaam, Tanzania. [77]University of the West Indies, Kingston, Jamaica. [78]School of Biological Sciences, University of Western Australia, Perth, WA, Australia. [79]The UWA Oceans Institute, University of Western Australia, Perth, WA, Australia. [80]Departamento de Pesquerias, Centro Interdisciplinario de Ciencias Marinas del IPN, La Paz, Baja California Sur, Mexico. [81]Pelagios Kakunjá, La Paz, Baja California Sur, Mexico. [82]Yardie Environmental Conservationists Limited, Kingston, Jamaica. [83]Coral Reef Research Foundation, Koror, Palau. [84]Departamento de Ecología y Territorio, Facultad de Estudios Ambientales y Rurales, Pontificia Universidad Javeriana, Bogotá, Colombia. [85]National Institute of Water and Atmospheric Research, Auckland, New Zealand. [86]Borneo Marine Research Institute, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia. [87]Marine Megafauna Foundation, West Palm, FL, USA. [88]Island Conservation Society Seychelles, Victoria, Mahé, Seychelles. [89]Emirates Nature - World Wide Fund for Nature, Dubai, United Arab Emirates. [90]College of Marine Sciences and Aquatic Biology, University of Khorfakkan, Sharjah, UAE. [91]Aquarium of the Pacific, Long Beach, CA, USA. [92]Marine and Environmental Sciences Centre/ Aquatic Research Network, Regional Agency for the Development of Research, Technology and Innovation, Funchal, Madeira, Portugal. [93]Oceans Institute, University of Western Australia, Perth, WA, Australia. [94]Inland Fisheries Ireland, Dublin, Ireland. [95]Western Australian Fisheries and Marine Research Laboratories, Department of Primary Industries and Regional Development, Government of Western Australia, Hillarys, WA, Australia. [96]CORDIO East Africa, Mombasa, Kenya. [97]Blue Ventures, Mombasa, Kenya. [98]American Samoa Department of Marine and Wildlife Resources, Pago Pago, American Samoa. [99]The Centre for Ocean Research and Education, Gregory Town, Eleuthera, The Bahamas. [100]Department of Ocean Science, Memorial University, NL, Canada. [101]Department of Environment and Geography, University of York, York, UK. [102]Departamento de Ciências Agrárias e Biológicas, Universidade Federal do Espírito Santo, São Mateus, Espírito Santo, Brazil. [103]Blue Sanctuary-Avalon, Jardines de la Reina, Cuba. [104]Centro de Investigaciones Marinas, Universidad de La Habana, Habana, Cuba. [105]Center for Marine Biology, University of São Paulo, São Sebastião, São Paulo, Brazil. [106]Large Marine Vertebrates Research Institute Philippines, Puerto Princesa City, Palawan, Philippines. [107]Center for Fisheries Research, Ministry for Marine Affairs and Fisheries, Jakarta Utara, Indonesia. [108]Fisheries Department, Universitas Dayanu Ikhsanuddin, Bau Bau, Southeast Sulawesi, Indonesia. [109]Programa de Pós Graduação em Ecologia: Teoria, Aplicação e Valores, Instituto de Biologia, Universidade Federal da Bahia, Salvador, BA, Brazil. [110]HJR Reefscaping, Boquerón, Puerto Rico. [111]Pristine Seas, National Geographic Society, Washington, DC, USA. [112]Charles Darwin Research Station, Charles Darwin Foundation, Puerto Ayora, Galapagos Islands, Ecuador. [113]Save Our Seas Foundation Shark Research Center and Guy Harvey Research Institute, Nova Southeastern University, Dania Beach, FL, USA. [114]School of Pure and Applied Sciences, Pwani University, Kilifi, Kenya. [115]Thurgau Hunting and Fishing Administration, Frauenfeld, Switzerland. [116]SalvageBlue, Kingstown, Saint Vincent and the Grenadines. [117]School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand. [118]Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Cambridge, UK. [119]Indo Ocean Project, Jln Toyapakeh DESA Toyapakeh, Nusa Penida, Bali, Indonesia. [120]Aquaculture and Fisheries Group, Wageningen University & Research, Wageningen, Netherlands. [121]Reef Check Dominican Republic, Santo Domingo, Dominican Republic. [122]Institut de Recherche pour le Développement, UMR Entropie (IRD-UR-UNC-CNRS-IFREMER), Nouméa, New Caledonia, France. [123]Secretariat of the Pacific Regional Environment Programme, Apia, Samoa. [124]Department of Natural Sciences, Faculty of Science Engineering, Manchester Metropolitan University, Manchester, UK. [125]Department of Life Science, Tunghai University, Taichung, Taiwan. [126]School of Environmental and Forest Sciences, University of Washington, Seattle, WA, USA. [127]GIBEAM Research Group, Universidad del Sinú, Cartagena, Colombia. [128]Corales del Rosario and San Bernardo National Natural Park, Colombia.

SUPPLEMENTARY MATERIALS

Submitted 31 August 2022; accepted 27 April 2023
10.1126/science.ade4884

CHEMICAL PHYSICS

Solvated dielectrons from optical excitation: An effective source of low-energy electrons

Sebastian Hartweg[1,2]*, Jonathan Barnes[3], Bruce L. Yoder[3], Gustavo A. Garcia[1], Laurent Nahon[1], Evangelos Miliordos[3,4], Ruth Signorell[3]*

Low-energy electrons dissolved in liquid ammonia or aqueous media are powerful reducing agents that promote challenging reduction reactions but can also cause radiation damage to biological tissue. Knowledge of the underlying mechanistic processes remains incomplete, particularly with respect to the details and energetics of the electron transfer steps. In this work, we show how ultraviolet (UV) photoexcitation of metal-ammonia clusters could be used to generate tunable low-energy electrons in situ. Specifically, we identified UV light–induced generation of spin-paired solvated dielectrons and their subsequent relaxation by an unconventional electron transfer–mediated decay as an efficient, low-energy electron source. The process is robust and straightforward to induce with the prospect of improving our understanding of radiation damage and fostering mechanistic studies of solvated electron reduction reactions.

Unbound low-energy electrons are produced in liquids by ionizing radiation or arise as intermediates in chemical reactions. If their kinetic energy lies below the first electronic excitation threshold of the liquid, they are also referred to as subexcitation electrons. Being potent chemical reduction reagents, slow electrons are able to drive reactions in molecular solvents, such as liquid water and ammonia (*1*, *2*). Subexcitation electrons have been identified as important components in the chain of radiation chemistry in aqueous environments—for example, causing severe damage to DNA molecules (*1*, *3*). Our understanding of these electron-induced reactions is still far from complete, in part owing to the difficulty of performing controlled low-energy electron studies in liquids. Solvated electrons have been widely used as strong reducing agents in chemical synthesis. The most famous example is the Birch reduction, which relies on the high reducing power of solvated electrons of alkali metals dissolved in liquid ammonia (*2*, *4*). The conditions, however, are technically demanding and hazardous, which has prompted the search for alternative approaches. Even though substantial progress has been made in this direction (*5*–*7*), it has proved challenging to find alternative approaches that can match the high reduction capability of the Birch reaction under milder conditions. Various Birch-type photoreduction pathways have been explored in this context (*5*–*7*), culminating in the discovery of a consecutive light-induced electron transfer catalytic cycle capable of generating highly reducing solvated electrons (*7*). Detailed mechanistic studies of such complex multicomponent systems remain challenging, which highlights the demand for simpler model systems to investigate individual steps. Reduction of aqueous carbon dioxide and nitrogen by solvated electrons has been demonstrated using ultraviolet (UV) illumination of diamond as the electron source (*8*, *9*). Comparable photoinduced electron transfer processes have not yet been reported for the conventional Birch system—i.e., alkali metal–ammonia solutions.

Here, we report a hitherto unknown light-induced electron transfer mechanism in alkali metal–ammonia solutions that efficiently produces low-energy electrons through the intermediate formation of a solvated dielectron. Composed of only a single solvent and solute, this system is exceptionally simple and thus amenable to fundamental studies of photoinduced electron transfer processes. In the mechanism that we discovered, UV light promotes an electron from a valence orbital of the solvent molecules to the singly occupied solvated electron orbital, generating a solvated spin-paired dielectron. The dielectron subsequently decays through an electron transfer–mediated decay (ETMD) process. As key intermediates, dielectrons (also referred to as bipolarons) play the central part in this charge transfer process. Although there has been much progress in unraveling the nature of single (unpaired) solvated electrons (*10*–*21*), the same is not true for the dielectron, whose direct experimental observation has proven difficult. To date, insights into the characteristics of this elusive species mostly rely on theoretical studies (*10*, *21*–*25*). The diamagnetism observed experimentally in metal-ammonia solutions at higher metal concentration has been explained by electron spin-pairing, although no definitive identification of the spin-paired species was possible (*10*, *15*, *26*). Experimental evidence for hydrated dielectrons has also been found in radiolysis experiments (*27*) and in mass spectrometry studies of water cluster anions (*23*). However, recent photoelectron studies on metal-ammonia solutions covering a wide range of metal concentrations have failed to detect signatures specific to the dielectron in the experimental spectra (*24*, *28*).

On the basis of quantum chemical calculations, Larsen and Schwartz have suggested possible experimental procedures to create detectable amounts of metastable dielectrons by capturing a second electron in the cavity of a preexisting single solvated electron (*21*). Pulsed radiolysis experiments and time-delayed multipulse laser schemes have been proposed. Inspired by their work, we developed an alternative UV photoexcitation and photoelectron detection method that can not only generate high yields of metastable dielectrons but also detect them through a specific photoelectron signature. The key to selective detection of dielectrons is the special ETMD process, which effectively replaces the pump-probe detection steps in the schemes suggested by Larsen and Schwartz (*21*). ETMD and intermolecular coulombic decay (ICD)—i.e., nonlocal decay processes of electronically excited species—were first predicted theoretically (*29*–*31*) before being observed experimentally in the soft x-ray range (*32*–*34*). ETMD processes accessible by UV light excitation offer new perspectives for the practical use of such nonlocal decay processes. We show how slow electrons were produced in a straightforward way by this unexpected UV-pumped dielectron-ETMD process. The discovered slow electron formation pathway could play a role in mechanistic studies of radiation chemistry and solvated electron reduction reactions.

Photoelectron spectroscopy

UV light–induced generation of low-energy electrons was directly probed in sodium-doped ammonia clusters $[Na(NH_3)_n]$ using double-imaging photoelectron-photoion coincidence spectroscopy at the DESIRS VUV beamline of the Synchrotron SOLEIL [see supplementary materials, sections S1 and S2, and (*35*)]. Coincidence spectroscopy provided cluster size–resolved photoelectron spectra (PES) and images (Fig. 1). The experiments were complemented by quantum chemical calculations of the ground and electronically excited states of the clusters, treating electron correlation in terms of Møller-Plesset and coupled cluster theory [(*36*, *37*); sections S4 to S8]. The combined approach allowed us to identify the UV light–induced generation of solvated, spin-paired dielectrons and their subsequent ETMD as the source of low-energy electrons. The mechanism that we discovered also explains the dominance of this process over a broad UV wavelength range.

[1]Synchrotron SOLEIL, L'Orme des Merisiers, 91190 Saint-Aubin, France. [2]Institute of Physics, University of Freiburg, 79104 Freiburg, Germany. [3]Department of Chemistry and Applied Biosciences, ETH Zurich, 8093 Zurich, Switzerland. [4]Department of Chemistry and Biochemistry, Auburn University, Auburn, AL, USA.
*Corresponding author. Email: sebastian.hartweg@physik.uni-freiburg.de (S.H.); ruth.signorell@phys.chem.ethz.ch (R.S.)

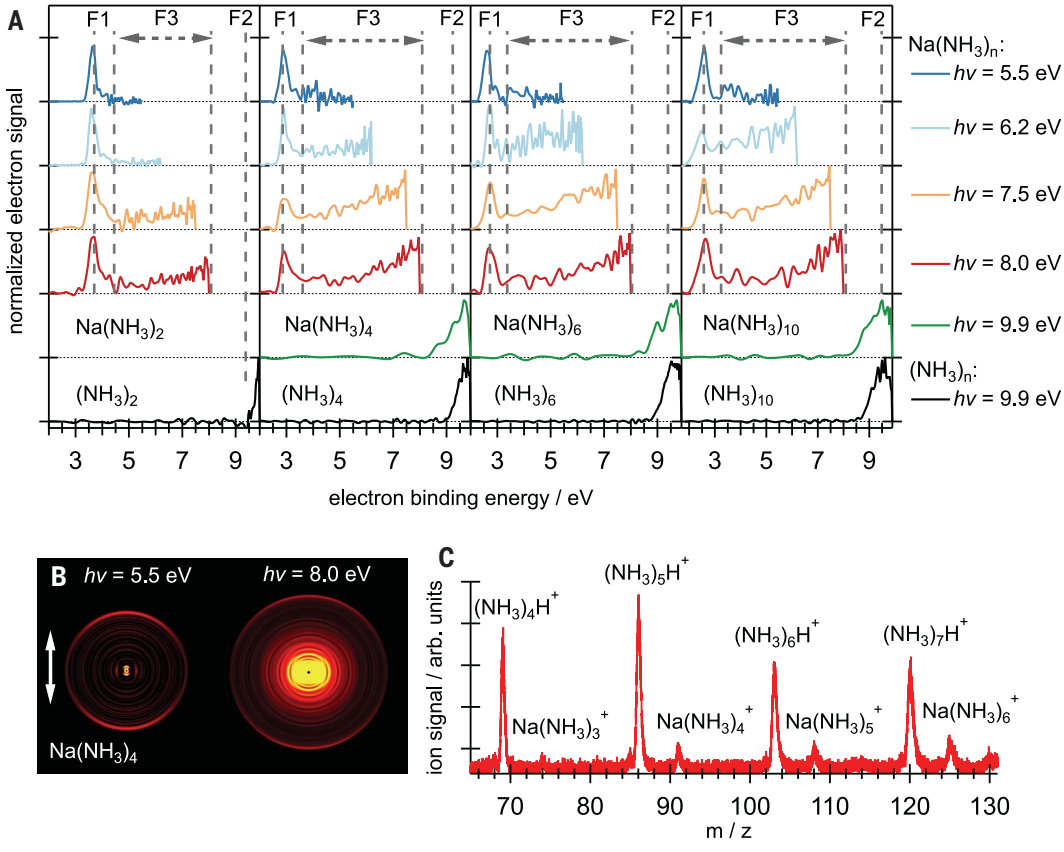


Fig. 1. Experimental results from photoelectron-photoion coincidence spectroscopy of sodium-doped ammonia clusters $Na(NH_3)_n$. (**A**) Cluster size–resolved PES of selected $Na(NH_3)_n$ clusters recorded after photoexcitation at different photon energies $h\nu$ (dark blue to green lines). Black lines are PES of pure ammonia clusters $(NH_3)_n$ recorded at $h\nu$ = 9.9 eV. Feature F1: Photoelectrons from the direct photoionization of single solvated electrons in $Na(NH_3)_n$ (e^-_{ph} from Φ_{solv}, Fig. 2A). Feature F2: Photoelectrons from the direct photoionization of ammonia solvent molecules in $Na(NH_3)_n$ (e^-_{ph} from Φ_L, Fig. 2B) or pure $(NH_3)_n$ clusters (black line). Feature F3: Low-energy electrons from the ETMD of the dielectrons (e^-_{ETMD} from Φ_{solv}, Fig. 2). (**B**) Velocity map images for $Na(NH_3)_4$ recorded at $h\nu$ = 5.5 eV (left) and $h\nu$ = 8.0 eV (right). Anisotropic photoelectron angular distribution (left) result from direct photoionization (e^-_{ph}), and isotropic distributions centered at low kinetic energies (right) result from ETMD of the dielectrons (e^-_{ETMD}). (**C**) Section of a mass spectrum recorded at $h\nu$ = 9.9 eV composed of intact $Na(NH_3)_n^+$ cluster signals from the pathway shown in Fig. 3B and protonated $(NH_3)_nH^+$ signals from ionization of pure $(NH_3)_n$ clusters. arb. units, arbitrary units; m/z, mass/charge ratio. Further information is presented in sections S1 to S3.

Figure 1A (blue to green traces) shows representative PES of $Na(NH_3)_n$ clusters acquired at different UV photon energies $h\nu$ (see section S3 for PES of larger clusters). Higher electron binding energies (eBEs) correspond to lower electron kinetic energies (E_{kin} = $h\nu$-eBE; fig. S2). PES recorded at photon energies below 5.5 eV (dark blue traces in Fig. 1A) consist of a single, narrow low-eBE band (feature F1) (*17*), and PES recorded at 9.9 eV (green traces) consist of a single high-eBE band (feature F2). At intermediate photon energies (5.5 eV < $h\nu$ < 9.9 eV), the PES are dominated by an unexpectedly broad distribution of low–kinetic energy electrons (E_{kin} < 6 eV; fig. S2), with increasing electron abundance toward zero kinetic energy (feature F3 in the light blue, orange, and red traces). Supported by quantum chemical calculations (sections S4 to S8), the observation of the three distinct features F1, F2, and F3 points to three distinct pathways of electron formation (Fig. 2), which can be turned on and off selectively by the choice of UV photon energy.

Direct photoionization of solvated electrons and solute molecules

The sharp F1 band results from direct photoionization of the solvated sodium 3s electron (Φ_{solv} in Fig. 2D) (*17*–*19*, *38*), corresponding to the transition from the electronic ground state ^2X of the neutral cluster to the electronic ground state $^1x^+$ of the cationic cluster (Fig. 2A). The high symmetry of Φ_{solv} is reflected in the pronounced anisotropic photoelectron distributions (Fig. 1B at $h\nu$ = 5.5 eV) (*17*). The coincidence study directly provided cluster size–resolved vertical eBEs in good agreement with ab initio calculations (tables S1 and S2). The vertical eBE decreases systematically from 4.3 eV for $Na(NH_3)$ to ~2.6 eV for $Na(NH_3)_{10}$, where it converges with respect to cluster size (figs. S3 and S4). The cluster size–resolved eBE values improve and extend previous data obtained from PES of cluster ensembles and photoion studies (*17*, *18*).

A different pathway of electron formation dominates at photon energies above the lowest ionization energy of the NH_3 solvent molecules (Fig. 1A, green trace at $h\nu$ = 9.9 eV) (*39*). In these cases, ionization occurs mainly through the removal of an electron from the highest occupied molecular orbital (HOMO) of the $(NH_3)_n$ solvent shell, Φ_L (Fig. 2E), instead of ionization from the HOMO of the entire $Na(NH_3)_n$ cluster, Φ_{solv}. The assignment of the F2 band to ionization from Φ_L is corroborated by the good agreement with the PES of pure liquid ammonia (*40*) and ammonia clusters (*39*) and by the fact that the PES of $Na(NH_3)_n$ clusters and bare $(NH_3)_n$ clusters are very similar (Fig. 1A, green and black traces, respectively). The F1 band is missing at $h\nu$ = 9.9 eV because of the higher ionization probability from Φ_L compared with that from Φ_{solv} combined with a large increase of the total number of photoelectrons (direct ionization of NH_3 molecules) and the limited dynamic range of the coincidence detection. The almost identical PES for $Na(NH_3)_n$ and bare $(NH_3)_n$ clusters imply that the electronic structure of the NH_3 solvent molecules is only slightly affected by the presence of the Na^+ cation and the solvated electron. In stark contrast to the PES, the relaxation dynamics after direct photoionization critically depends on the absence [$(NH_3)_n$] or presence [$Na(NH_3)_n$] of a solvated electron. Relaxation of bare $(NH_3)_n^+$ cluster ions occurs through proton transfer between neighboring NH_3 molecules and the release of an $NH_2^•$ radical from the cluster: $(NH_3)_n + h\nu = (NH_3)_{n-1}H^+ + NH_2^• + e^-_{ph}$ (*41*). Thus, mass spectra of undoped ammonia clusters consist of protonated cluster signals $(NH_3)_{n-1}H^+$ (major mass peaks in Fig. 1C). This ultrafast relaxation

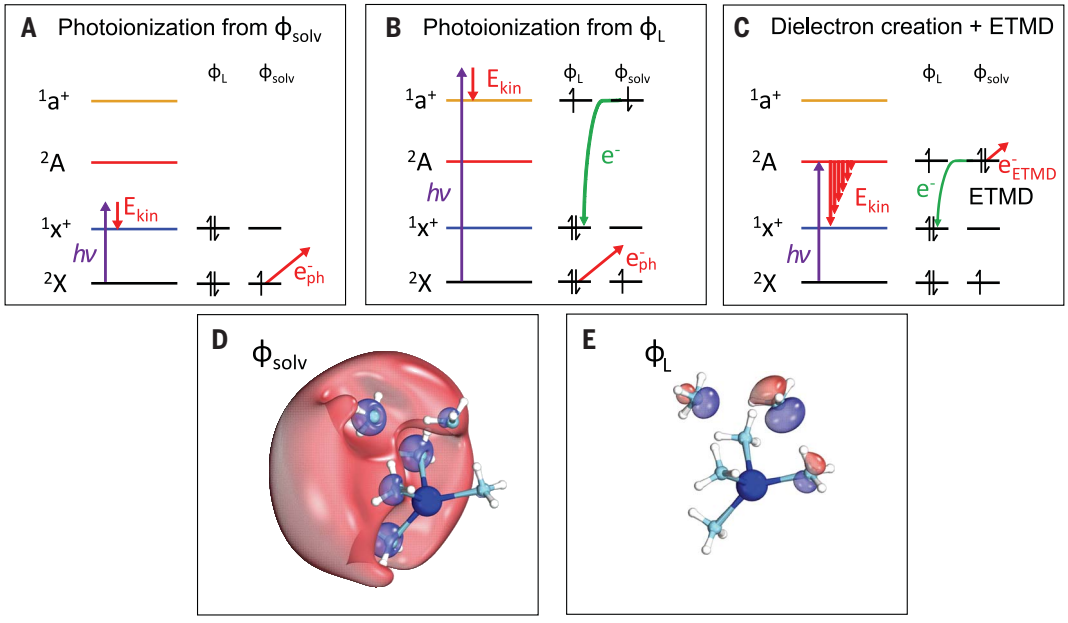


Fig. 2. The three distinct pathways of electron formation observed after UV photoexcitation of $Na(NH_3)_n$ clusters. (**A** to **C**) The three pathways. Horizontal lines on the left indicate qualitative energy levels of the four electronic states labeled ^2X, $^1x^+$, ^2A, and $^1a^+$. Orbital occupations are indicated on the right. Φ_L is the HOMO of the bare ammonia solvent shell, and Φ_{solv} is the HOMO of the entire $Na(NH_3)_n$, which corresponds to the delocalized solvated electron orbital. $h\nu$ is the UV photon energy, and E_{kin} is the kinetic energy of the ejected electron. e^-_{ph} (red), e^- (green), and e^-_{ETMD} (red) are electrons from direct photoionization, internally transferred electrons, and electrons from ETMD processes, respectively. Red arrows indicate the ejection of electrons, and green arrows indicate an internal electron transfer. (A) Direct photoionization of an unpaired solvated electron from Φ_{solv} in $Na(NH_3)_n$ (feature F1 in Fig. 1A), producing a cation cluster in the ground state $^1x^+$. (B) Direct photoionization of an electron from Φ_L of the ammonia solvent molecules (e^-_{ph}, feature F2 in Fig. 1A), producing a cation cluster in the excited state $^1a^+$ and internal electron hopping from Φ_{solv} to the Φ_L vacancy (e^-). (C) Photoexcitation of an electron from Φ_L of the ammonia solvent molecules to the singly occupied Φ_{solv}, forming a dielectron (^2A state). Subsequent decay of the dielectron through ETMD, producing a cation cluster in the ground state $^1x^+$. ETMD consists of internal hopping of one electron from Φ_{solv} to the Φ_L vacancy (e^-) and ejection of the second electron from Φ_{solv} from the cluster (e^-_{ETMD}, feature F3 in Fig. 1A). The electron is ejected with a broad distribution of different kinetic energies (E_{kin}, see Fig. 3). (**D** and **E**) Representation of the Φ_{solv} and the Φ_L orbitals, respectively, of a $Na(NH_3)_6$ cluster (section S4).

pathway through proton transfer is well known to occur in various hydrogen-bonded systems (*41*). The proton transfer is completely suppressed in the presence of a solvated electron by the fast electron transfer from Φ_{solv} to Φ_L (Fig. 3B and section S8). The mass spectra show intact cluster ion signals $Na(NH_3)_n^+$ (minor mass peaks in Fig. 1C) instead of ion signals of protonated clusters. Albeit ionization occurs from the same ammonia orbital (Φ_L) in neat and Na-doped clusters, the presence of the solvated electron opens an unexpected relaxation pathway in $Na(NH_3)_n$ that is faster than the proton transfer pathway. We propose that relaxation of $Na(NH_3)_n^+$ clusters occurs through initial ultrafast electron transfer from the delocalized Φ_{solv} orbital to the localized Φ_L vacancy formed by photoionization (Fig. 2B), with subsequent energy dissipation through internal conversion. The electron hopping between Φ_{solv} and the Φ_L vacancy corresponds to a transition from the first electronically excited $^1a^+$ state of the cluster cation to its $^1x^+$ ground state. This electron transfer is analogous to the first step of a Birch reduction, in which a solvated electron transfers to an unoccupied molecular orbital.

Low-energy electron formation through the decay of dielectrons

Even more intriguing with respect to the formation of reducing electrons is the mechanism that occurs after photoexcitation with photons of ~5.5 to 8 eV energy (wavelength ~225 to 155 nm). PES are dominated by the broad F3 feature (Fig. 1A). This band corresponds to a broad distribution of low–kinetic energy electrons, with increasing electron abundance toward zero kinetic energy (maximum eBE; fig. S2) and isotropic angular distribution (Fig. 1B, left). This behavior is fully consistent with the excitation of a dielectron state that autoionizes by ETMD (Fig. 2C). Conceivable alternative mechanisms to explain the origin of the F3 feature are assessed in section S8 in detail. Because they prove to be inconsistent with our experimental observations, we do not discuss them further here. The dielectron-ETMD mechanism producing the F3 electrons is visualized in Fig. 2C. Direct ionization from the ammonia HOMO Φ_L, responsible for the F2 band, is not possible at these lower photon energies. Instead, photoexcitation promotes an electron from Φ_L to the singly occupied, delocalized solvated electron orbital Φ_{solv} to form a metastable, solvated spin-paired dielectron. This unusual electronically excited dielectron state, ^2A, is characterized by a doubly occupied Φ_{solv} orbital and a vacancy in the Φ_L orbital of the solvation shell, in agreement with ab initio calculations (Fig. 3 and sections S4 and S5). Regarding the difference of this dielectron compared with the ground state dielectron species postulated in metal-ammonia

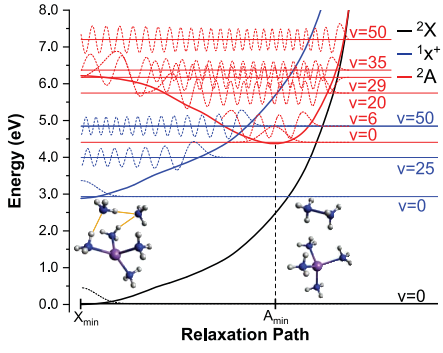


Fig. 3. Calculated PEPs along the one-dimensional relaxation path of the dielectronic state for $Na(NH_3)_6$. Vertical photoexcitation from the minimum geometry (X_{min}) of the electronic ground state ^2X (black curve) creates a dielectron state in the saddle point region of the electronically excited ^2A state (red curve), which then relaxes along the path. The PEPs of the ^2A state and the ground state of the cation, $^1x^+$ (blue curve), intersect before the minimum structure of the ^2A state, A_{min}, is reached. A_{min} can be described as an ammonia dimer cation $(NH_3)_2^+$ bound to a smaller sodium-doped ammonia cluster cation $Na(NH_3)_{n-2}^+$ by a solvated electron pair. Selected vibrational wave functions shown for the ^2X (black dashed), $^1x^+$ (blue dashed), and ^2A (red dashed) states illustrate the many possibilities (and hence the high probability) for ETMD processes to occur along the relaxation path. Further information is presented in sections S4 and S5.

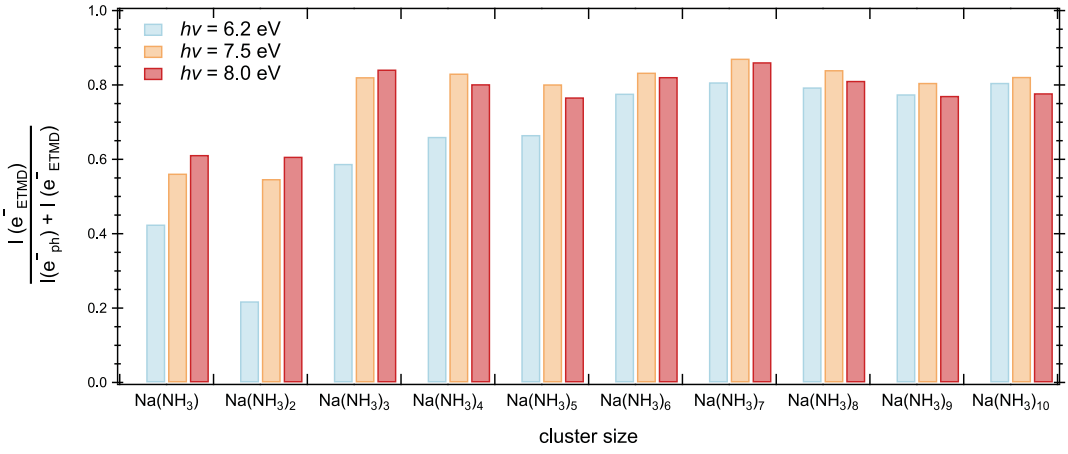


Fig. 4. Relative yield of ETMD electrons at different UV excitation energies as a function of the cluster size. Ratio $I(e^-_{ETMD})$ / [$I(e^-_{ETMD})$ + $I(e^-_{ph})$] retrieved from the experimental PES by integrating over the F3 [$I(e^-_{ETMD})$] and F1 [$I(e^-_{ph})$] bands (Fig. 1A and figs. S3 and S4). The figure illustrates that the UV light–induced dielectron ETMD (Fig. 2C) is largely insensitive to system size and UV energy. The high relative yield of e^-_{ETMD} can be explained by the high probabilities of forming solvated dielectrons by UV photoexcitation and their subsequent autoionization through ETMD (Fig. 2C and Fig. 3).

solutions (*10*), we refer the reader to section S8. The ^{2}A state lies above the $^1x^+$ electronic ground state of the cation and autoionizes through a specific version of ETMD (Fig. 2C): One of the two delocalized electrons in Φ_{solv} transfers to the solvation shell, where it localizes in the ammonia Φ_L vacancy. Concurrent with this transfer, the second electron is ejected with low kinetic energy (e^-_{ETMD}) and is detected in the PES as part of the F3 band (Fig. 1A). The broad kinetic energy distribution of the resulting photoelectrons reflects the concerted progress of photoelectron ejection and electron transfer during the ETMD (see discussion of Fig. 3). This UV light–induced ETMD process offers a straightforward way to generate high yields of dielectrons and to distinguish them (F3 band) from the always accompanying single solvated electrons (F1 band) (*21*, *24*, *28*). Compared with previously reported ETMD processes (*30*–*32*, *34*, *42*, *43*), this ETMD is distinctive in two aspects: The electron transfer goes hand in hand with a considerable spatial contraction of the electron wave function, and it is initiated by exceptionally low photon energies in the UV range. The former might enhance the efficiency of the electron transfer, whereas the latter makes ETMD more amenable to applications.

Figure 3 shows computational results of the potential energy profiles (PEPs) along the one-dimensional relaxation path of the neutral dielectron state (red) after vertical photoexcitation from the neutral ground state (black). The large difference between the equilibrium geometries of the electronic ground and dielectron state imply a broad Franck-Condon progression in the photoexcitation spectrum, which makes the dielectron state accessible over a broad range of UV photon energies. The PEP of the cationic state (blue) intersects the PEP of the dielectron state just before its minimum, A_{min}, has been reached. The selected vibrational wave functions (dashed curves) of the dielectron state and the cationic state highlight the many possibilities along the relaxation path for ETMD processes to occur (section S4 and S5), which are reflected in the broad kinetic energy distribution of the $e^-_{ETMD,}$ constituting the F3 band in Fig. 1A. In terms of a time-dependent wave packet picture, an increased hopping probability near the intersection with the cationic ground state PEP offers a plausible explanation for the observed preference of the e^-_{ETMD} toward zero kinetic energies (see discussion of fig. S2 in sections S3 and S8). For all photon energies, the sharp cutoff of the F3 feature (Fig. 1A) is therefore located at a binding energy corresponding to the photon energy. The observed high yield of e^-_{ETMD} is consistent with a high photoexcitation probability from the neutral ground state to delocalized dielectron states and a high autoionization probability through ETMD (section S9).

Figure 4 shows the relative electron yields of the F3 band for selected cluster sizes at three different photoexcitation energies. Several essential insights emerge from these results. ETMD dominates electron formation over a broad range of photoexcitation energies, extending from the near to the far UV. It is largely insensitive to the specific UV photon energy. Except for the smallest clusters ($n \leq 2$), ETMD does not depend on the cluster size either (see figs. S3 and S4 for larger clusters). The stability of the ETMD process with respect to changes of the system size was also confirmed by calculations, so that autoionization through solvated dielectrons should persist in bulk solutions (section S6). Moreover, the phenomenon exhibits some robustness to chemical changes. For example, replacing sodium with lithium had little effect (section S6). Theoretical assessment of solvent exchange proved to be more challenging because of competing relaxation channels (section S7). For solvents hosting solvated electrons, the formation of slow electrons from a transient dielectron state through ETMD generally appears to be an important channel.

Conclusions

What are the possible implications of this hitherto unknown slow electron formation pathway? ICD and ETMD have been discussed as in situ sources of low-energy electrons, with the caveat that high-energy photon sources are generally required to trigger these processes. The present ETMD process offers the possibility to use conventional UV lamps in the near to far UV range. By controlling the photon energy, the maximum electron kinetic energy can be tuned across the subexcitation range. This could be exploited for systematic energy-dependent mechanistic studies relevant to radiation chemistry and radiation damage in biological tissues. Given that broadband UV lamps can be used for in situ generation of high yields of low-energy electrons, the discovery of the dielectron-ETMD mechanism might also be of practical relevance to electron reduction reactions. It is conceivable that reduction reactions with solvated electrons are photo-enhanced by the dielectron-ETMD mechanism, as the more highly energized ETMD electrons could overcome reaction barriers unsurmountable to common solvated electrons. It would be intriguing to study the energy-dependent activity of the solvated electron as a reducing agent by tuning the energy of the ETMD electron. Future investigations of the dynamics of the dielectron-ETMD process after UV photoexcitation might even hold the prospect of discovering previously unknown reaction pathways.

REFERENCES AND NOTES

1. B. C. Garrett *et al.*, *Chem. Rev.* **105**, 355–390 (2005).
2. P. W. Rabideau, Z. Marcinow, in *Organic Reactions*, L. A. Paquette, Ed. (Wiley, 1992), pp. 1–334.
3. E. Alizadeh, L. Sanche, *Chem. Rev.* **112**, 5578–5602 (2012).
4. A. J. Birch, *J. Chem. Soc.* **1944**, 430–436 (1944).

5. B. K. Peters *et al*., *Science* **363**, 838–845 (2019).
6. A. Chatterjee, B. König, *Angew. Chem. Int. Ed.* **58**, 14289–14294 (2019).
7. J. P. Cole *et al*., *J. Am. Chem. Soc.* **142**, 13573–13581 (2020).
8. L. Zhang, D. Zhu, G. M. Nathanson, R. J. Hamers, *Angew. Chem. Int. Ed.* **53**, 9746–9750 (2014).
9. D. Zhu, L. Zhang, R. E. Ruther, R. J. Hamers, *Nat. Mater.* **12**, 836–841 (2013).
10. E. Zurek, P. P. Edwards, R. Hoffmann, *Angew. Chem. Int. Ed.* **48**, 8198–8232 (2009).
11. R. M. Young, D. M. Neumark, *Chem. Rev.* **112**, 5553–5577 (2012).
12. J. M. Herbert, M. P. Coons, *Annu. Rev. Phys. Chem.* **68**, 447–472 (2017).
13. P. Vöhringer, *Annu. Rev. Phys. Chem.* **66**, 97–118 (2015).
14. O. Marsalek, F. Uhlig, J. VandeVondele, P. Jungwirth, *Acc. Chem. Res.* **45**, 23–32 (2012).
15. P. P. Edwards, *J. Phys. Chem.* **88**, 3772–3780 (1984).
16. D. Luckhaus, Y. I. Yamamoto, T. Suzuki, R. Signorell, *Sci. Adv.* **3**, e1603224 (2017).
17. A. H. C. West *et al*., *J. Phys. Chem. Lett.* **6**, 1487–1492 (2015).
18. T. Zeuch, U. Buck, *Chem. Phys. Lett.* **579**, 1–10 (2013).
19. A. O. Gunina, A. I. Krylov, *J. Phys. Chem. A* **120**, 9841–9856 (2016).
20. C. P. Schulz, A. Scholz, I. V. Hertel, *Isr. J. Chem.* **44**, 19–25 (2004).
21. R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. B* **110**, 1006–1014 (2006).
22. G. J. Martyna, Z. Deng, M. L. Klein, *J. Chem. Phys.* **98**, 555–563 (1993).
23. R. N. Barnett, R. Giniger, O. Cheshnovsky, U. Landman, *J. Phys. Chem. A* **115**, 7378–7391 (2011).
24. T. Buttersack *et al*., *Science* **368**, 1086–1091 (2020).
25. M. Mauksch, S. B. Tsogoeva, *Phys. Chem. Chem. Phys.* **20**, 27740–27744 (2018).
26. U. Schindewolf, M. Werner, *J. Phys. Chem.* **84**, 1123–1129 (1980).
27. K. H. Schmidt, D. M. Bartels, *Chem. Phys.* **190**, 145–152 (1995).
28. S. Hartweg, A. H. C. West, B. L. Yoder, R. Signorell, *Angew. Chem. Int. Ed.* **55**, 12347–12350 (2016).
29. L. S. Cederbaum, J. Zobeley, F. Tarantelli, *Phys. Rev. Lett.* **79**, 4778–4781 (1997).
30. J. Zobeley, R. Santra, L. S. Cederbaum, *J. Chem. Phys.* **115**, 5076–5088 (2001).
31. A. Ghosh, L. S. Cederbaum, K. Gokhberg, *Chem. Sci.* **12**, 9379–9385 (2021).
32. K. Sakai *et al*., *Phys. Rev. Lett.* **106**, 033401 (2011).
33. T. Jahnke *et al*., *Phys. Rev. Lett.* **93**, 163401 (2004).
34. T. Jahnke *et al*., *Chem. Rev.* **120**, 11295–11369 (2020).
35. X. Tang, G. A. Garcia, J.-F. Gil, L. Nahon, *Rev. Sci. Instrum.* **86**, 123108 (2015).
36. R. J. Bartlett, J. D. Watts, S. A. Kucharski, J. Noga, *Chem. Phys. Lett.* **165**, 513–522 (1990).
37. K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479–483 (1989).
38. I. R. Ariyarathna, F. Pawłowski, J. V. Ortiz, E. Miliordos, *Phys. Chem. Chem. Phys.* **20**, 24186–24191 (2018).
39. A. H. C. West, B. L. Yoder, R. Signorell, *J. Phys. Chem. A* **117**, 13326–13335 (2013).
40. T. Buttersack *et al*., *J. Am. Chem. Soc.* **141**, 1838–1841 (2019).
41. F. Dong, S. Heinbuch, J. J. Rocca, E. R. Bernstein, *J. Chem. Phys.* **124**, 224319 (2006).
42. I. Unger *et al*., *Nat. Chem.* **9**, 708–714 (2017).
43. L. Ben Ltaief *et al*., *Phys. Chem. Chem. Phys.* **22**, 8557–8564 (2020).
44. S. Hartweg *et al*., Solvated dielectrons from optical excitation: An effective source of low-energy electrons, data collection, ETH Zürich Research Collection (2023); https://doi.org/10.3929/ethz-b-000611963.

ACKNOWLEDGMENTS

We thank D. Stapfer, M. Steger, and J.-F. Gil for technical support and the general staff of SOLEIL for providing beamtime under projects 20200627 and 99210131. E.M. acknowledges ETH Zurich, the US National Science Foundation for a guest professorship, the donors of the James E. Land endowment, and computing resources provided by the Auburn University Hopper and Easley Clusters. S.H. acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, RTG 2717). **Funding:** This project has received funding from the European Union's Horizon 2020 research and innovation program from the European Research Council under grant agreement 786636, the Swiss National Science Foundation (project 200020_200306), and the US National Science Foundation (grant no. CHE-1940456). **Author contributions:** Conceptualization: S.H. and R.S. Methodology: S.H., B.L.Y., G.A.G., L.N., E.M., and R.S. Investigation: S.H., J.B., B.L.Y., G.A.G., L.N., and E.M. Visualization: S.H., E.M., and R.S. Funding acquisition: E.M. and R.S. Project administration: S.H. and R.S. Supervision: L.N. and R.S. Writing – original draft: S.H., E.M., and R.S. Writing – review & editing: S.H., J.B., B.L.Y., G.A.G., L.N., E.M., and R.S. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data that reproduce the analyses are available in a data repository (*44*). All (other) data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh0184
Supplementary Text
Figs. S1 to S11
Tables S1 to S8
References (*45*–*62*)

Submitted 6 February 2023; accepted 12 May 2023
Published online 25 May 2023
10.1126/science.adh0184

CHEMISTRY

Modification of ground-state chemical reactivity via light–matter coherence in infrared cavities

Wonmi Ahn[1], Johan F. Triana[2], Felipe Recabal[2], Felipe Herrera[2,3]*, Blake S. Simpkins[4]*

Reaction-rate modifications for chemical processes due to strong coupling between reactant molecular vibrations and the cavity vacuum have been reported; however, no currently accepted mechanisms explain these observations. In this work, reaction-rate constants were extracted from evolving cavity transmission spectra, revealing resonant suppression of the intracavity reaction rate for alcoholysis of phenyl isocyanate with cyclohexanol. We observed up to an 80% suppression of the rate by tuning cavity modes to be resonant with the reactant isocyanate (NCO) stretch, the product carbonyl (CO) stretch, and cooperative reactant-solvent modes (CH). These results were interpreted using an open quantum system model that predicted resonant modifications of the vibrational distribution of reactants from canonical statistics as a result of light–matter quantum coherences, suggesting links to explore between chemistry and quantum science.

Controlling chemical reactions with electromagnetic fields is a long-standing goal in chemistry and physics (*1*, *2*). Femtosecond laser pulses can transiently excite vibrational modes of reactant molecules to selectively promote breaking or forming of chemical bonds (*3*–*5*). However, fast energy redistribution in polyatomic molecules severely limits this approach, despite efforts to overcome this obstacle using laser pulse shaping (*6*, *7*).

Chemical control without lasers has been recently demonstrated using cavities (*8*–*12*). In this approach, hybrid light–matter polariton states arise from strong interactions of dipole-allowed molecular transitions with the cavity vacuum at optical (*13*) and infrared frequencies (*14*–*16*). Experiments show inhibition of excited-state processes such as photoisomerization (*17*) and photobleaching (*18*) in visible cavities and also modification of bond formation and cleavage rates in infrared cavities as a result of vibrational strong coupling (VSC) (*8*–*10*, *19*). VSC is characterized by collective molecular response in transmission (*20*, *21*), a spatial dependence of the interaction that follows the mode profile (*22*, *23*), and reversible modulation of the system using ultrafast lasers (*24*, *25*) or electrochemistry (*26*, *27*).

Achieving coupling-induced selective chemistry would enable chemical catalysis by design, but challenges to its reproducibility (*28*, *29*) and lack of mechanistic explanation have stifled progress. Here, we report robust experimental evidence of cavity-modified chemistry and describe a theory consistent with measurements. We studied the alcoholysis of phenyl isocyanate (PHI) with cyclohexanol (CHol) in tetrahydrofuran (THF) to give urethane [cyclohexyl carbanilate (CC)]. The reaction is exothermal (*30*), has a low activation energy (*31*), and resonant cavity modes can be tuned to reactant, product, or solvent vibrational modes. We measured a strong cavity-tuning dependence of the reaction kinetics, with rate constants reduced by 30 to 80%, and developed a quantum model that qualitatively agrees with observations and provides mechanistic understanding for intracavity reaction kinetics. Our theory proposes that the intracavity reactivity depends on stationary light–matter coherences, and we discuss the importance of energy disorder in preserving coherence over chemical time scales.

[1]UNAM — National Nanotechnology Research Center and Institute of Materials Science and Nanotechnology, Bilkent University, Ankara, Turkey. [2]Department of Physics, Universidad de Santiago de Chile, Santiago, Chile. [3]Millennium Institute for Research in Optics (MIRO), Concepción, Chile. [4]Chemistry Division, US Naval Research Laboratory, Washington, DC, USA.
*Corresponding author. Email: blake.simpkins@nrl.navy.mil (B.S.S.); felipe.herrera.u@usach.cl (F.H.)

Results and discussion

The alcoholysis of isocyanates is well understood (*31*–*33*) and proceeds through concerted nucleophilic addition at the NC bond in isocyanate (*31*, *33*, *34*). The geometry of the PHI-CHol complex (Fig. 1A) involves an NHO hydrogen bond that evolves into a cyclic NHOC structure in the transition state (*3*). Cleavage of the HO bond results in ring opening and exothermic formation of urethane ($\Delta H_{rxn} \approx -20.5$ kcal/mol) with activation energy of 6.7 kcal/mol (2343 cm^{-1}) in THF. The second-order rate constant at room temperature is $k_0 = 0.59 \times 10^{-5}$ $M^{-1}{\cdot}s^{-1}$ (*3*). Back reactions are negligible.

We injected the reactant solution into a thin-layer cell bounded by transparent CaF_2 windows, for out-of-cavity control measurements, or by Au-coated CaF_2 windows, for cavity-coupled measurements [Fig. 1B; additional details in section 1 of the supplementary materials (SM)]. Figure 1C shows two sets of transmission spectra. For control measurements (upper curves, red), reactant bands decreased (NCO stretch of PHI at ~2260 cm^{-1} and OH band of CHol at ~3470 cm^{-1}), and product bands grew (CO at 1730 cm^{-1} and [illegible] at 3293 cm^{-1}; see detailed spectra in figs. S1 and S2). The NCO band absorption was converted to reactant concentration through direct proportionality (see procedure in section 1 of the SM and fig. S3 for extinction coefficient calibration), then inverted and plotted against time (Fig. 1D), yielding a line whose slope equaled the second-order rate constant (*35*). The average of six such measurements gave a control rate constant $k_0 = (2.34 \pm 0.2) \times 10^{-5}$ $M^{-1}{\cdot}s^{-1}$ (datasets in fig. S4). This rate was higher than in previous reports (*32*), which involved lower reactant concentrations. The rates measured under our conditions were consistent in independent measurements performed in a period of 24 months, with all reactions (control and cavity-coupled) carried out using the same initial reactant concentrations.

Typical transmission spectra for a cavity-coupled sample are shown in Fig. 1C (lower curves, blue). These example data exhibited multiple resonant peaks, with one coupled to the NCO band of PHI (2260 cm^{-1}), giving a splitting at normal incidence of 112 cm^{-1} (cavity Q ~ 100, cavity linewidth $\kappa \approx 38$ cm^{-1}; see table S2). These evolving transmission spectra were fit to a function that accounted for the absorbance of the intracavity medium (Fig. 1E), which, again, was directly proportional to reactant concentration. We inverted and plotted this data (Fig. 1F) to extract a rate constant $k = (1.48 \pm 0.2) \times 10^{-5}$ $M^{-1}{\cdot}s^{-1}$ for this sample, ~37% lower than uncoupled controls. Collection of the entire cavity dispersion allowed identification and fitting of spectra showing strong interaction with the mode of interest regardless of tuning at normal incidence [see the model in section 3.3 of the SM and (*14*, *21*, *23*, *25*, *27*)].

Reaction rates were extracted for different cavities. The resulting "action spectrum" (*9*) is shown in Fig. 2A. The initial (blue) and final (orange) transmission spectra of the control solution are shown to identify relevant vibrational modes. There was a strong dependence of the reaction rate on the cavity mode tuning, with rate suppression due to VSC on reactant (NCO), product (CO), and cooperative reactant-solvent (CH) modes (full dataset in fig. S8 and table S1). Cavity-induced suppression spanned 30 to 80%, relative to uncoupled controls. The largest suppression was found for cavities tuned to the NCO reactant mode, with a frequency dependence that closely followed the shape of the NCO absorption band. The rate constants for far-detuned cavities (squares) were close to the out-of-cavity rates (k/k_0 ~ 0.91; see table S1). Figure 2B shows representative inverse concentration plots and linear fits for cavities tuned to the reactant NCO and reactant-solvent CH bands, highlighting the lower slopes (rates), relative to out-of-cavity controls. Our mechanistic discussion below

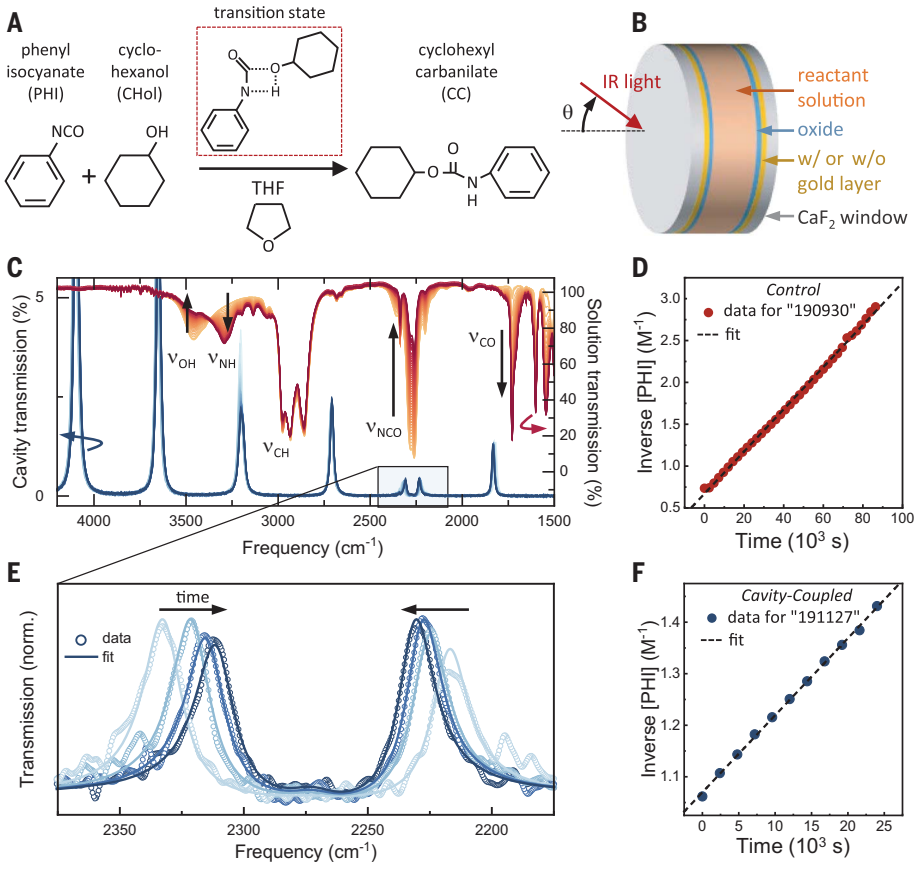


Fig. 1. Description of urethane monomer formation and reaction monitoring. (**A**) The reactants phenyl isocyanate (PHI) and cyclohexanol (CHol) were combined in tetrahydrofuran (THF) to form cyclohexyl carbanilate (CC). (**B**) Solution was contained between two CaF_2 windows that were either transparent (for control measurements) or coated with Au/SiO_2 (for cavity-coupled experiments). (**C**) Time-dependent Fourier transform infrared transmission spectra for out-of-cavity control measurements (red hues) showed reactant absorptions, ν_{NCO} of PHI at 2260 cm^{-1} and ν_{OH} of CHol at 3470 cm^{-1}, diminished as the reaction proceeded, while product features, ν_{CO} at 1730 cm^{-1} and ν_{NH} at 3293 cm^{-1}, increased. The ν_{NCO} absorption was converted to PHI concentration, inverted, and plotted versus time to extract the second-order reaction rate constant as shown in (**D**). The blue curves in (C) correspond to a time series of cavity-coupled transmission spectra showing strong coupling between the cavity and NCO vibrational mode of the PHI reactant. These spectra were fit, as shown in (**E**), to yield time-dependent PHI concentration, which was inverted, plotted, and fit to yield the reaction rate constant under cavity-coupled conditions. One typical cavity-coupled dataset is shown in (**F**).

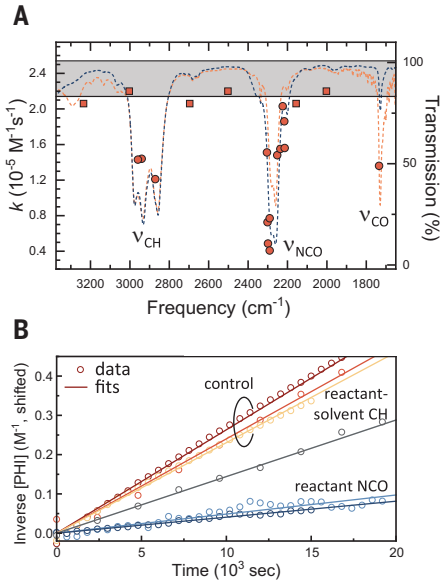


Fig. 2. Cavity-modified chemical reactivity. The action spectrum presented in (**A**) shows the extracted reaction rate constants (orange symbols) as a function of cavity tuning (i.e., Fabry-Perot mode position at normal incidence). The gray horizontal band represents the average out-of-cavity control rate with its width equal to the standard deviation of six measurements. Blue and orange dashed curves correspond, respectively, to initial and final transmission spectra. Reaction suppression was observed when the cavity was tuned to prominent vibrational modes. Several example linear fits, from which reaction rate constants were extracted, are shown in (**B**).

focuses on the NCO band because it plays a prominent role in the reaction, however, modification of one mode can influence others (intramolecular vibrational relaxation, Fermi coupling, etc.). Further, we note that the product-coupled cavity also supported a higher-order mode that weakly coupled to the reactant OH mode (~3500 cm^{-1}), however, we have only highlighted modes under strong coupling. The role of weak coupling in chemical reactivity has yet to be fully understood.

Our mechanistic description of VSC-modified reactivity started by modeling the vibrational structure of the PHI molecule in the frequency region of the NCO band, which includes a fundamental NCO stretch, ν_6, and a Fermi resonance between ν_6 and a combination of low-frequency CH bending modes (analysis in section 4 of the SM). The NCO fundamental at 2260 cm^{-1} was inhomogeneously broadened (full width at half maximum ≈ 47 cm^{-1}; see fig. S13). We modeled an ensemble of N reactant NCO vibrations under VSC at 300 K, considering vibrational relaxation, cavity decay, thermalization, and many-body correlations, using an open quantum system approach (see sections 5 and 6 of the SM). Field-dependent dipole self-energy terms (*36*) were not included. The theory suggested that although the coupled vibration-cavity system was at thermal equilibrium with its environment, as confirmed by experiments (*37*), when tracing out the photonic degrees of freedom, the stationary vibrational population of reactants could deviate from canonical Boltzmann statistics. This phenomenon has been shown to occur for other strongly coupled subsystems (*38*), but its potential consequences in cavity chemistry have yet to be fully explored. In this picture, chemical bonds can break and form through local two-body processes with the vibrational level statistics modified by the strongly interacting photonic environment. This insight can be complemented by other approaches wherein the cavity photon quadrature is treated as another classical coordinate that contributes to a static polaritonic potential energy surface (*39*, *40*).

In Fig. 3A, we show the deviation of the most-probable ν_6 occupation from its cavity-free canonical Boltzmann value, $\Delta n_6/n_6^{th}$, as a function of cavity frequency, for an ensemble of PHI molecules (N = 50) with a Gaussian distribution of ν_6 mode frequencies (variance = σ^2), coupled to a single cavity mode. Cavity-modified rate measurements (orange points reproduced from Fig. 2A) and the PHI transmission spectrum (orange dashed curve) are also shown. This single-mode theory predicted a narrow vibrational depopulation feature (blue

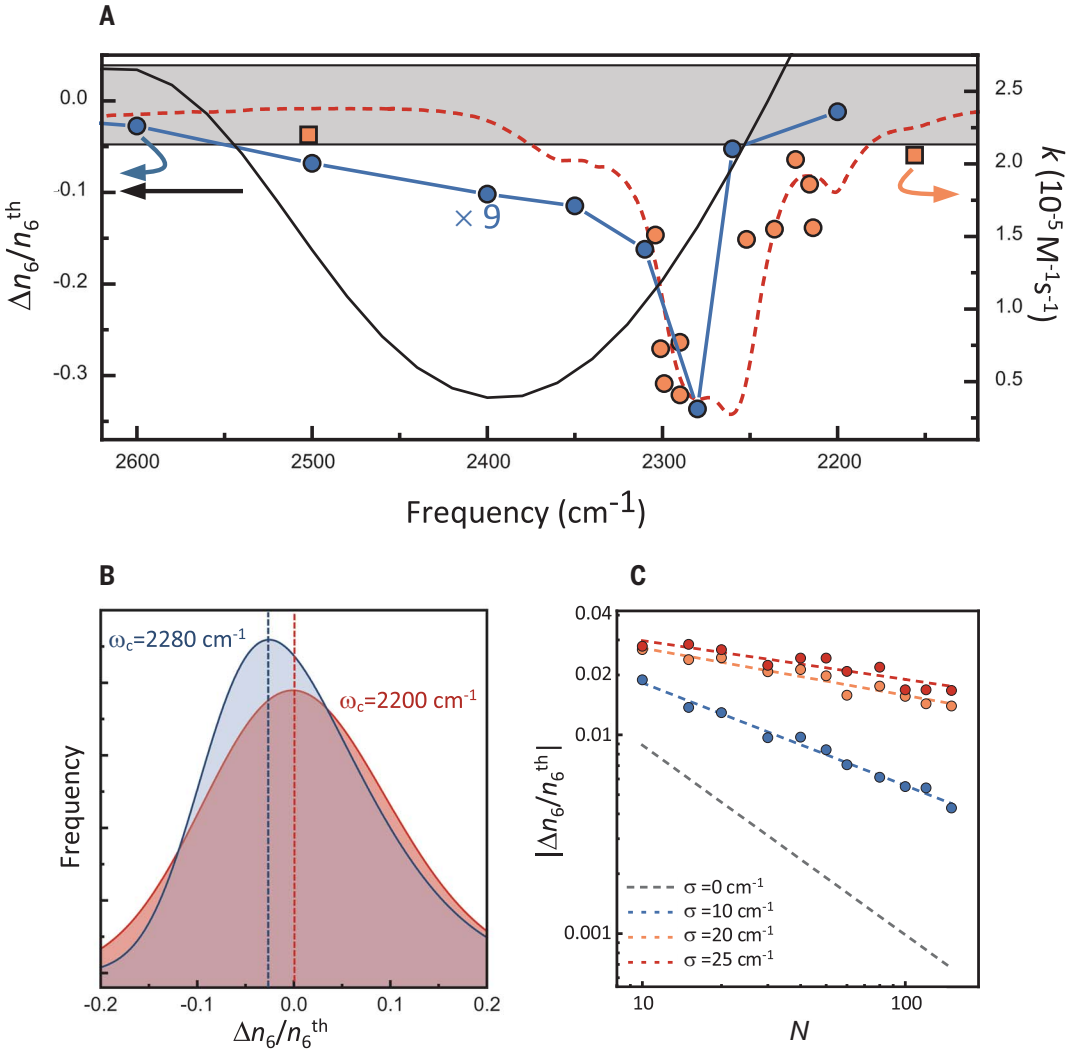


Fig. 3. Comparison of cavity-modified reactivity with theoretical prediction. (**A**) The NCO region of the action spectrum of Fig. 2 is reproduced here. Single-mode theory predicted a resonant depopulation of the ν_6 mode (blue points) that qualitatively followed the experimental data well. Example cavity-induced population redistribution is shown in (**B**). A multimode single-molecule treatment, solid black curve in (A), yielded a resonant depopulation effect that was considerably stronger in magnitude (note the 9× scaling applied to the single-mode analysis) but was broader and blue-shifted relative to the experimental results. Although the cavity-induced effect was predicted to diminish with increased oscillator number, N, the scaling power strongly depended on molecular disorder (**C**). Molecular disorder, σ, was defined as a Gaussian broadening of the Lorentzian linewidth.

circles, $\Delta n_6/n_6^{th} < 0$) which closely followed the transmission lineshape and was most prominent at the frequency of the combination band (2280 cm^{-1}), in qualitative agreement with the measured action spectrum. Figure 3B shows that at large detuning, the vibrational occupation of ν_6 was symmetrically centered at the canonical Boltzmann average, but near-resonant cavities gave a skewed distribution of $\Delta n_6/n_6^{th}$, whose most-probable value corresponded to net vibrational depopulation (additional histograms in fig. S14). The formal connection between $\Delta n_6/n_6^{th}$ and the reaction rate has yet to be developed (see ansatz in section 6 in the SM).

The single-mode many-particle analysis qualitatively agreed with the experimental data and improved our understanding of cavity-suppressed reactivity. However, the predicted values of $\Delta n_6/n_6^{th}$ were relatively small. This result was due to the many-particle model not accounting for the full dispersion of the cavity field. Treating large N and a continuum of cavity modes is prohibitive, but we could gain insight by treating a single PHI molecule in a multimode Fabry-Perot cavity with a quasi-continuous spectrum ω_k, where k denotes the in-plane wave number (details in section 6.2 of the SM). We showed that $\Delta n_6/n_6^{th}$ in this case is proportional to $\sum_k (g_{6,k}/\gamma_6)\mathrm{Im}\left\langle \hat{a}_k \hat{b}_6^\dagger \right\rangle_{ss}$, where $\left\langle \hat{a}_k \hat{b}_6^\dagger \right\rangle_{ss}$ is the stationary light–matter coherence between the vibrational mode and the kth cavity mode, $g_{6,k}$ is the Rabi frequency, and γ_6 is the homogeneous vibrational linewidth. Solving for the steady-state coherence gave

$$\Delta n_6/n_6^{th} = \sum_k P(\omega_k)\left(e^{-\Delta_{k,6}/k_B T} - 1\right) \quad (1)$$

where $\Delta_{k,6} = \omega_k - \omega_{\nu_6}$ is the detuning of ν_6 from the kth mode, and $P(\omega_k)$ is a normalized distribution function scaling with $\left(g_{6,k}/\Delta_{k,6}\right)^2$ for $|\Delta_{k,6}| \gg \gamma_6$. Figure 3A (solid black curve) shows that the frequency response from Eq. 1 was much broader and blue-shifted relative to experiments. However, integrating over the cavity dispersion gave values of $\Delta n_6/n_6^{th}$ considerably larger than the single cavity mode approach, suggesting that the entire photon spectrum contributed to cavity chemistry phenomena.

Finally, we addressed the N-scaling of the vibrational depopulation effect. Because the total vibration-cavity system was in thermal equilibrium (*37*), deviation of the photonic occupation from a canonical distribution by δn would correspond to a redistribution of vibrational occupation per molecule of $-\delta n/N$. Therefore, for typical values of $N \sim 10^6$ in Fabry-Perot cavities, the population redistribution on individual molecules should be negligible. However, in Fig. 3C we showed that this many-body dilution behavior did not hold in general for ensembles with frequency disorder, by plotting $\Delta n_6/n_6^{th}$ as a function of N for different values of the disorder width σ (see also fig. S15). We found that inhomogeneous broadening could protect resonant population redistribution from the homogeneous $1/N$ scaling, possibly due to partial delocalization of molecular states (*41*, *42*).

Conclusions

In this work, we suppressed a ground-state addition reaction through strong coupling between molecular vibrational modes and cavity vacuum fields. This suppression reached 80%, and we found a strong cavity frequency dependence that closely followed the reactant infrared absorption spectrum. The strongest effect was for cavities resonant with an NCO mode that participated in the transition state of the reaction. We described the mechanism quantum mechanically as the emergence of stationary noncanonical vibrational populations as a result of strong vibration–cavity coupling but noted that the composite vibration-cavity polaritonic state remained in a Boltzmann thermal state. Deviations of the vibrational occupations from canonical statistics were due to stationary light–matter coherences that depended on the details of the cavity spectrum and dispersion. For molecular ensembles, we showed evidence that inhomogeneous spectral broadening could protect the light–matter coherences that influenced vibrational reactivity, suggesting fundamental links between chemistry and quantum science that have yet to be fully developed.

REFERENCES AND NOTES

1. A. H. Zewail, *Phys. Today* **33**, 27–33 (1980).
2. W. S. Warren, H. Rabitz, M. Dahleh, *Science* **259**, 1581–1589 (1993).
3. T. Stensitzki *et al*., *Nat. Chem.* **10**, 126–131 (2018).
4. M. Delor *et al*., *Science* **346**, 1492–1495 (2014).
5. Z. Lin *et al*., *J. Am. Chem. Soc.* **131**, 18060–18062 (2009).
6. A. Assion *et al*., *Science* **282**, 919–922 (1998).
7. M. P. A. Branderhorst *et al*., *Science* **320**, 638–643 (2008).
8. A. Thomas *et al*., *Angew. Chem. Int. Ed.* **55**, 11462–11466 (2016).
9. A. Thomas *et al*., *Science* **363**, 615–619 (2019).
10. F. J. Garcia-Vidal, C. Ciuti, T. W. Ebbesen, *Science* **373**, eabd0336 (2021).
11. B. S. Simpkins, A. D. Dunkelberger, J. C. Owrutsky, *J. Phys. Chem. C* **125**, 19081–19087 (2021).
12. F. Herrera, J. Owrutsky, *J. Chem. Phys.* **152**, 100902 (2020).
13. G. Khitrova, H. M. Gibbs, M. Kira, S. W. Koch, A. Scherer, *Nat. Phys.* **2**, 81–90 (2006).
14. J. P. Long, B. S. Simpkins, *ACS Photonics* **2**, 130–136 (2015).
15. A. Shalabney *et al*., *Nat. Commun.* **6**, 5981 (2015).
16. J. Del Pino, J. Feist, F. J. Garcia-Vidal, *New J. Phys.* **17**, 053040 (2015).
17. J. A. Hutchison, T. Schwartz, C. Genet, E. Devaux, T. W. Ebbesen, *Angew. Chem. Int. Ed.* **51**, 1592–1596 (2012).
18. B. Munkhbat, M. Wersäll, D. G. Baranov, T. J. Antosiewicz, T. Shegai, *Sci. Adv.* **4**, eaas9552 (2018).
19. J. Lather, P. Bhatt, A. Thomas, T. W. Ebbesen, J. George, *Angew. Chem. Int. Ed.* **58**, 10635–10638 (2019).
20. B. S. Simpkins *et al*., *ACS Photonics* **2**, 1460–1467 (2015).
21. C. Weisbuch, M. Nishioka, A. Ishikawa, Y. Arakawa, *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
22. W. Ahn, I. Vurgaftman, A. D. Dunkelberger, J. C. Owrutsky, B. S. Simpkins, *ACS Photonics* **5**, 158–166 (2018).
23. S. Wang *et al*., *J. Phys. Chem. Lett.* **5**, 1433–1439 (2014).
24. A. D. Dunkelberger *et al*., *ACS Photonics* **6**, 2719–2725 (2019).
25. R. Houdré *et al*., *Phys. Rev. B* **52**, 7810–7813 (1995).
26. J. J. Pietron, K. P. Fears, J. C. Owrutsky, B. S. Simpkins, *ACS Photonics* **7**, 165–173 (2020).
27. W. Ahn, B. S. Simpkins, *APL Photonics* **5**, 076107 (2020).
28. M. V. Imperatore, J. B. Asbury, N. C. Giebink, *J. Chem. Phys.* **154**, 191103 (2021).
29. G. D. Wiesehan, W. Xiong, *J. Chem. Phys.* **155**, 241103 (2021).
30. E. Delebecq, J.-P. Pascault, B. Boutevin, F. Ganachaud, *Chem. Rev.* **113**, 80–118 (2013).
31. F. Kössl, M. Lisaj, V. Kozich, K. Heyne, O. Kühn, *Chem. Phys. Lett.* **621**, 41–45 (2015).
32. G. Raspoet, M. T. Nguyen, M. McGarraghy, A. F. Hegarty, *J. Org. Chem.* **63**, 6878–6885 (1998).
33. A. A. Caraculacu, S. Coseri, *Prog. Polym. Sci.* **26**, 799–851 (2001).
34. M. Çoban, F. A. S. Konuklar, *Comput. Theor. Chem.* **963**, 168–175 (2011).
35. K. J. Laidler, *Chemical Kinetics* (Pearson Education Inc., 1987).
36. C. Schäfer, M. Ruggenthaler, A. Rubio, *Phys. Rev. A* **98**, 043801 (2018).
37. M. Seidel *et al*., *ACS Photonics* **6**, 1823–1825 (2019).
38. D. Xu, J. Cao, *Front. Phys. (Beijing)* **11**, 110308 (2016).
39. T. E. Li, J. E. Subotnik, A. Nitzan, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18324–18331 (2020).
40. C. Schäfer, J. Flick, E. Ronca, P. Narang, A. Rubio, *Nat. Commun.* **13**, 7817 (2022).
41. D. Wellnitz, G. Pupillo, J. Schachenmayer, *Commun. Phys.* **5**, 120 (2022).
42. M. Du, J. Yuen-Zhou, *Phys. Rev. Lett.* **128**, 096001 (2022).
43. J. F. Triana, F. Recabal, F. Herrera, B. S. Simpkins, W. Ahn, Data: Modification of ground state chemical reactivity via light-matter coherence in infrared cavities, version 2, Zenodo (2023); https://doi.org/10.5281/zenodo.7915873.

ACKNOWLEDGMENTS

We thank J. Owrutsky, A. Dunkelberger, I. Vurgaftman, and J. Schachenmayer for discussions. **Funding:** This work was supported by US Naval Research Laboratory Nanoscience Institute, grant WU 1J03 (W.A. and B.S.S.); ANID Fondecyt Regular grant 1221420 (F.H.); ANID Fondecyt Doctorado grant 21221970 (F.R.); ANID Fondecyt Iniciación grant 11230679 (J.T.); Millennium Science Initiative Program grant ICN17_012 (F.H. and J.T.); and Programa de Cooperación Científica ECOS-ANID ECOS grant 200028 (F.H.). **Author contributions:** Conceptualization: B.S.S. and F.H. Methodology: B.S.S., F.H., and W.A. Investigation: W.A., J.F.T., F.R., F.H., and B.S.S. Visualization: B.S.S., J.F.T., and F.R. Funding acquisition: B.S.S. and F.H. Writing – original draft: W.A. and B.S.S. Writing – review & editing: B.S.S. and F.H. **Competing interests:** None declared. **Data and materials availability:** All data needed to support the conclusions of the main text and supplementary materials have been uploaded to Zenodo (*43*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade7147
Materials and Methods
Supplementary Text
Figs. S1 to S19
Tables S1 to S4
References (*44*–*77*)

Submitted 2 September 2022; resubmitted 7 March 2023
Accepted 12 May 2023
10.1126/science.ade7147

OPTOELECTRONICS

Metamaterial graphene photodetector with bandwidth exceeding 500 gigahertz

Stefan M. Koepfli*, Michael Baumann, Yesim Koyaz†, Robin Gadola, Arif Güngör, Killian Keller, Yannik Horst, Shadi Nashashibi, Raphael Schwanninger, Michael Doderer, Elias Passerini, Yuriy Fedoryshyn, Juerg Leuthold*

Although graphene has met many of its initially predicted optoelectronic, thermal, and mechanical properties, photodetectors with large spectral bandwidths and extremely high frequency responses remain outstanding. In this work, we demonstrate a >500 gigahertz, flat–frequency response, graphene-based photodetector that operates under ambient conditions across a 200-nanometer-wide spectral band with center wavelengths adaptable from <1400 to >4200 nanometers. Our detector combines graphene with metamaterial perfect absorbers with direct illumination from a single-mode fiber, which breaks with the conventional miniaturization of photodetectors on an integrated photonic platform. This design allows for much higher optical powers while still allowing record-high bandwidths and data rates. Our results demonstrate that graphene photodetectors can outperform conventional technologies in terms of speed, bandwidth, and operation across a large spectral range.

Graphene has met many of its initially predicted properties (*1*, *2*) and is pushing toward the market (*3*). However, graphene-based, high-performance electronics and photonics are lagging behind, even though the predicted market impact is large (*4*, *5*). Nevertheless, impressive optoelectronic device demonstrations related to modulators (*6*, *7*), mixers (*8*), and photodetectors (PDs) (*9*–*12*) have been reported. In particular, graphene-based PDs that make use of graphene's high carrier mobility, tunable electrical properties, and relatively easy integration have been demonstrated, showing, for example, high responsivities that exploit photo gain effects (*13*, *14*) or bandwidths exceeding 100 GHz (*15*–*17*). Although graphene offers a nearly uniform absorption from the ultraviolet all the way to beyond the far infrared (*18*), a major challenge lies in overcoming its relatively low absorption of roughly 2.3% (*19*). Most of the fastest and highest-performing detectors have therefore been demonstrated on photonic integrated circuit (PIC) platforms such as silicon (*10*–*12*, *15*, *20*) or silicon nitride (*21*). A parallel propagation of the electric field with graphene provides longer interaction lengths, resulting in a higher absorption. By using plasmonic enhancement, even shorter and more-responsive detectors have been demonstrated (*16*, *17*, *22*, *23*). And although the use of graphene on PICs has already shown a multitude of functional applications (*24*, *25*), PIC integration also comes at a price. PIC integration limits the accessible wavelength range, be it due to limitations in the transparency of the waveguide material (silicon and others) or the limited bandwidth of the integrated optical circuit elements (grating couplers, splitters, and so on). Further, PIC integration puts constraints on the polarization dependence and the footprint because of access waveguides. The small modes of the PIC and the plasmonic enhancement also mean that all of the light interacts with the graphene sheet in a very limited volume, leading to an early onset of saturation, which effectively limits the maximum extractable photocurrents to a level in the microampere range (see table S1). As an alternative, graphene may be directly illuminated by vertical incidence from free space. This approach could leverage the full capabilities of graphene for photodetection without being masked by the limitations of the chosen photonic platform. Yet this requires a structure to efficiently enhance the absorption in graphene. Furthermore, additional considerations regarding the overall device geometry and its contacting scheme are more cri… because of the larger size of the device. Ne… theless, it has been shown that even free space–coupled graphene detectors can reach bandwidths exceeding 40 GHz (*9*). Without some of the constraints of the PIC, the overall efficiency does not suffer from coupling schemes, and, furthermore, other properties such as different wavelengths and polarization are now freely accessible. For example, access to any polarization was recently used to demonstrate directional photocurrents for polarization-resolved detection in the mid-infrared regime (*26*).

Graphene offers a large variety of physical detection effects: Unlike conventional PDs, such as PIN photodiodes or bolometers that work with one specific detection mechanism, graphene detectors have a variety of different detection mechanisms (*27*–*29*), such as carrier-based mechanisms [photoconductive (PC) (*17*) and photovoltaic (PV) (*30*)], thermal mechanisms [bolometric (BOL) (*16*) and photothermoelectric (PTE) (*31*)], or gain medium–assisted mechanisms (*13*, *14*). Recent device demonstrations have pushed toward PTE operation to overcome high dark currents when relying on biased detection mechanisms (*23*, *32*, *33*). For the design of a high-speed, efficient graphene PD, it remains unclear which of the direct detection mechanisms (PV, PC, BOL, or PTE) can achieve the highest bandwidth, and many of these effects can coexist at the same time in one device, making dedicated design difficult. Time-resolved spectroscopy measurements of graphene indicate that the carrier dynamics could allow for both thermal and carrier-based graphene PDs that operate beyond 300 GHz (*34*–*36*).

We report on a zero-bias, graphene PD with a >500-GHz electro-optic bandwidth. Our device operates under ambient conditions across a large spectral range of more than 200 nm and may be adapted to a large variety of center wavelengths, from <1400 to >4200 nm. A metamaterial perfect absorber (*37*) layer stack

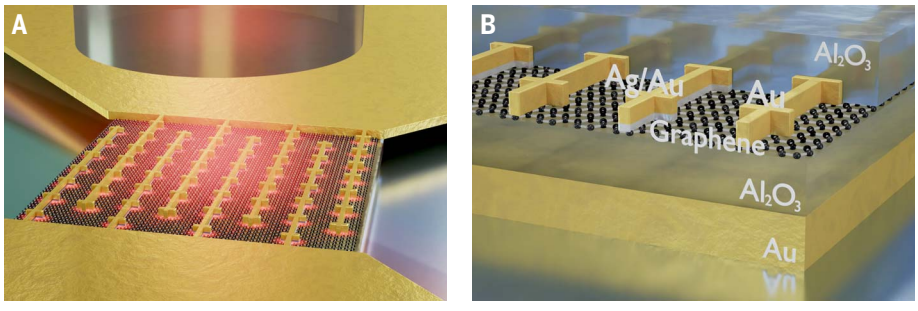


Fig. 1. Artist's view of the interdigitated graphene metamaterial PD. (**A**) Artistic visualization of the device directly illuminated by a single-mode fiber from the top. (**B**) Visualization of the device architecture. The PD consists of a gold reflector backplane, an aluminum oxide spacer layer, monolayer graphene, and connected dipole resonators. The metal lines have alternating contact metals made from silver or gold. The structure is then capped by an aluminum oxide passivation layer. [Illustration credit: S.M.K. and M.D.]

Institute of Electromagnetic Fields (IEF), ETH Zürich, 8092 Zürich, Switzerland.
***Corresponding author. Email: koepflis@ethz.ch (S.M.K.); leuthold@ethz.ch (J.L.)**
†Present address: Photonic Systems Laboratory (PHOSL), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

provides the resonant enhancement, acts simultaneously as an electrical contact, and induces a pn-doping that enables efficient and fast carrier extraction. The light can be directly coupled to the detector by means of a standard single-mode fiber. The direct free-space coupling enables a distribution of the optical power, which leads to high saturation powers above 100 mW and damage thresholds exceeding 1 W. The detector has been tested for high-speed operation up to 132 Gbit/s with a two-level pulse-amplitude modulation format (PAM-2). The layer stack allows for almost substrate-independent processing, laying a foundation for a cost-effective technology that could allow for a dense monolithic co-integration with electronics. We further show the versatility of the approach by tuning the metamaterial geometry to operate in the mid-infrared wavelength regime, thereby offering high-speed and cost-efficient detectors in a range that otherwise lacks such detectors. The new sensor thereby offers opportunities for applications in communications and sensing.

The operation principle

The graphene PD consists of a modified metamaterial perfect absorber (Fig. 1, A and B). It comprises a gold reflector backplane, an aluminum oxide spacer layer, a chemical vapor deposition (CVD)–grown monolayer of graphene, gold dipole antennas, and an aluminum oxide passivation layer. The dipole resonators are connected with interdigitated electrical contact lines. This architecture combines absorption enhancement by means of a metamaterial perfect absorber stack (*37*), built-in driving fields at the graphene-metal contact (*9*), electrostatic gating capabilities, and a passivation for stable ambient operation (*38*). The resonators, in combination with the appropriately spaced backplane (*39*), allow the architecture to reach almost perfect absorption. Compared with a normal PIC approach, this layer stack is processed on top of a generic, inexpensive silicon substrate with silicon dioxide.

Scanning electron micrographs of the fabricated structure are shown in various magnifications in Fig. 2, A to D. The total active device area is 10 μm by 10 μm, consisting of 10 unit cells by 10 unit cells of the metamaterial. To understand the operation principle better, we looked at four of these 1-μm–by–1-μm unit cells, as illustrated in Fig. 2D (see materials and methods for design dimensions). In the simulated electric-field distribution under plane-wave illumination at 1550 nm (Fig. 2E), the bright spots show the dipole antenna–like behavior. The resonators are connected by metallic contact lines (Fig. 2C), which do not disturb the field pattern. The bottom panel of Fig. 2E shows the cross section at the marked (black dashed) center line, showing the high field close to the resonators. From the field simulations, the resulting spatial absorption distribution is calculated by

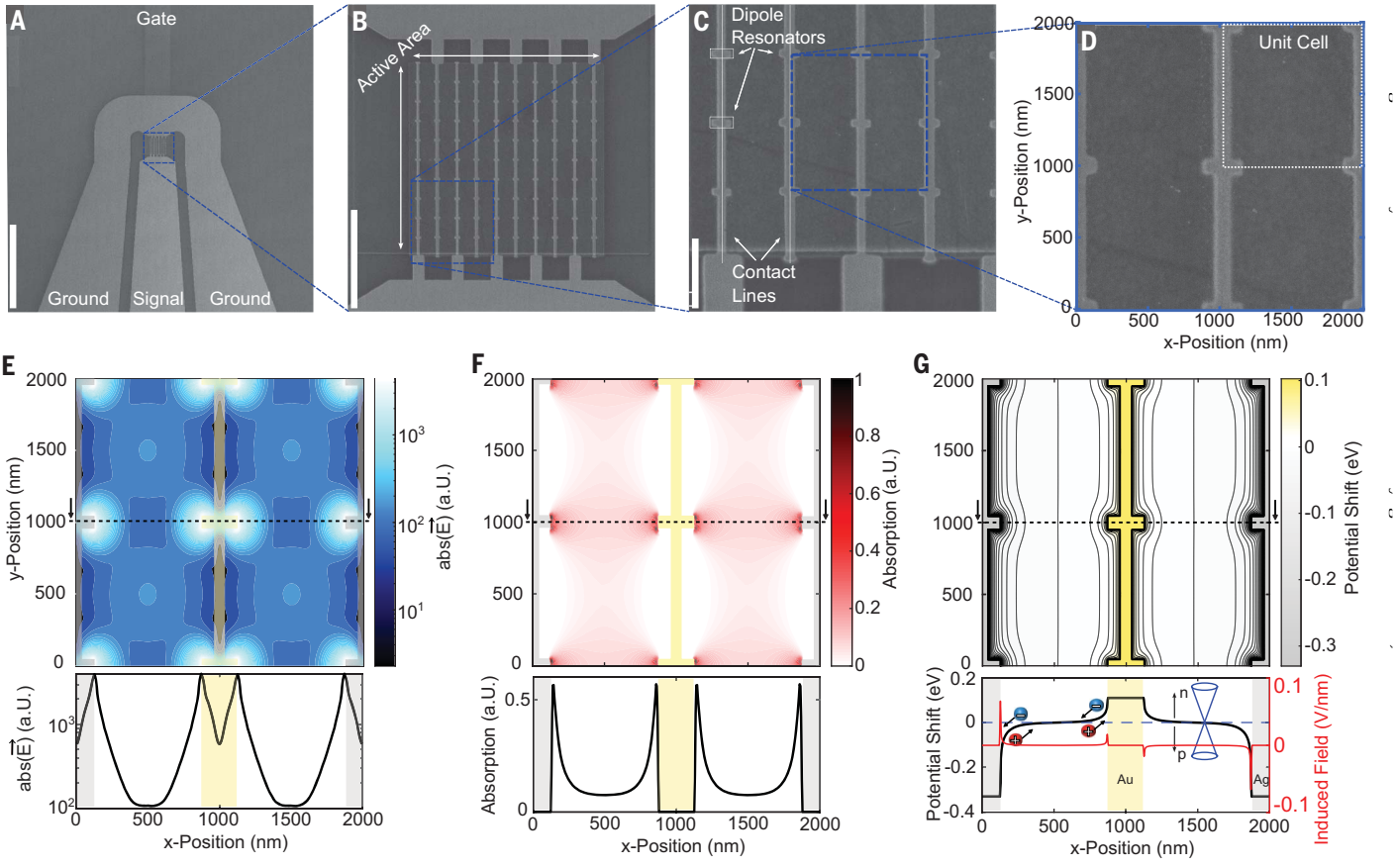


Fig. 2. Fabricated devices and simulated optical and electronic behavior. (**A** to **D**) Scanning electron micrographs of the proposed metamaterial graphene PD (prepassivation) at various magnifications. The micrographs illustrate the device architecture from the electrical signal lines to the active area and down to the resonator elements. Shown in (D) are four unit cells (each 1 μm by 1 μm) in an *x*- and *y*-coordinate system. Scale bars are 50 μm (A), 5 μm (B), and 1 μm (C). (**E** to **G**) The simulated optical and electrostatic behavior for the same four unit cells. The dipole antenna behavior with the electromagnetic field distribution is illustrated in (E), and the resulting spatial absorption distribution is illustrated in (F). Most of the absorption is confined close to the dipole resonators. The simulated contact metal–induced potential shift (G) shows the induced p-n doping due to the alternating contact metals. The cross sections along the center line (y = 1000 nm) for each simulation type [(E) to (G)] are presented at the bottom of each panel, showing that the optical signal and the doping are strongest close to the contact region. a.U., arbitrary units; abs(E), absolute electric field.

using Poynting's theorem (Fig. 2F). The absorption is strongly localized close to the interconnected resonators that are schematically added, as can be clearly seen in the cross-sectional view shown in Fig. 2F.

To be able to extract the resulting locally photoexcited carriers efficiently, we leveraged the alternating contact metals on the resonators (Fig. 1B). A thin silver contact layer was added below a thicker gold layer on every other line. The remaining lines are made purely from gold. This way, n- and p-type contact doping is achieved (*40*). The simulated potential shift is visualized with the contour in Fig. 2G, which shows the strong doping underneath and close to the contacts. In addition, we calculated the induced field by taking the gradient of the potential map, as shown in the bottom panel of Fig. 2G (red solid line, right axis). The induced field allows for bias-free extraction of the photogenerated carriers as depicted. Overall, the plots show that there is a very good overlap of locations where absorption is strongest (Fig. 2F) and where the driving field is largest (Fig. 2G). The design therefore promises short extraction pathways and efficient and fast carrier extraction.

PD under illumination

An optical microscope image of the device (top view) is provided in Fig. 3A; the bottom panel shows a side view of the device in contact with the electrical probes and the direct optical coupling with a single-mode fiber. The spectral behavior of the metamaterial graphene PD is shown in Fig. 3B, where we measured the normalized photoresponse as a function of wavelength, verifying the resonant enhancement of the metamaterial. The recorded peak at 1580 nm and a full-width-at-half-maximum window of 200 nm spans the S-, C-, L-, and U communication bands with the potential to tune the center frequency simply by changing the length of the resonators.

Under 1550-nm illumination, the optical input power is swept from <1 μW to close to 100 mW. Figure 3C shows the linear behavior of the response over these five orders of magnitude. No saturation effect is observed. An external responsivity of R_{ext} = 0.75 mA/W is calculated. The internal responsivity corresponds to R_{int} = 1.57 mA/W at 1550 nm, which was found by using a lensed fiber that provides a better-matched illumination spot size (see materials and methods). The internal response is illustrated with the blue dotted line in Fig. 3C. At first sight, this responsivity looks low. Yet the signal-to-noise ratio (SNR) of a shot noise–limited detector linearly depends on the photocurrent, which is given by $I_{ph} = RP_{in}$.

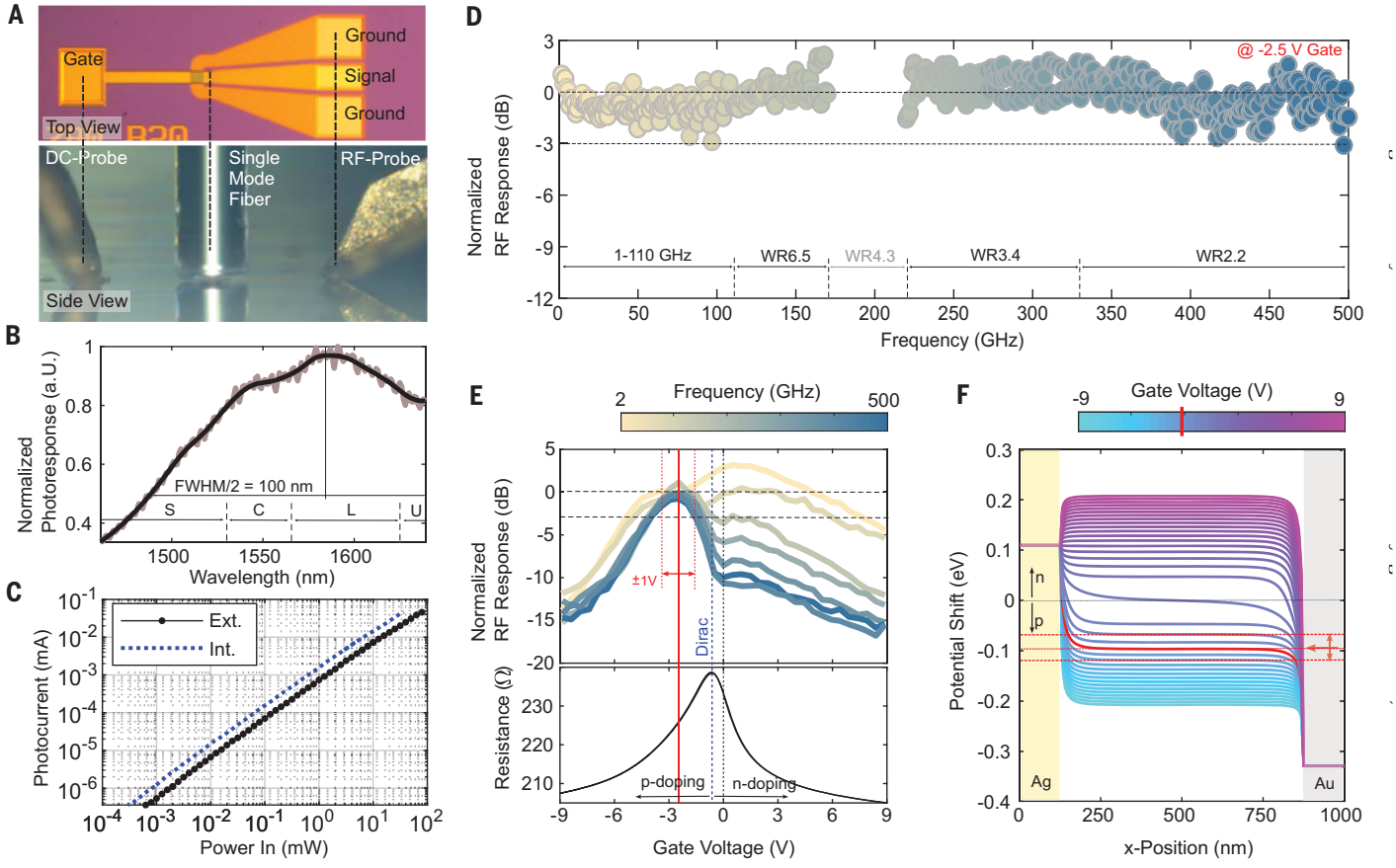


Fig. 3. Device performance for telecom wavelengths. (**A**) Top view (top) and side view (bottom) images of the device while it is in contact with the electrical probes, as taken with an optical microscope. The images show the direct optical coupling with a single-mode fiber. DC, direct current; RF, radio frequency. (**B**) Normalized photoresponse as a function of illumination wavelength showing the resonant enhancement and the wide-band operation. FWHM, full width at half maximum. (**C**) Extracted photocurrent for an optical input power sweep over five orders of magnitude (black line). The blue line corresponds to the optical power on the device (Int.), whereas the black line corresponds to the power out of the single-mode fiber (Ext.). The responsivities are R_{ext} = 0.75 mA/W and R_{int} = 1.57 mA/W. (**D**) Normalized frequency response of the graphene PD over a range of 2 to 500 GHz. A flat response with no roll-off behavior was measured. WR, waveguide rectangular. (**E**) The normalized RF response at various RF tones as a function of gate voltage. An ideal gate voltage at −2.5 ±1 V is found to enable the flat response, which corresponds to a slight p-doping, as can be seen from the resistance curve on the bottom. The resistance curve further reveals the Dirac points close to 0 V and a very small hysteretic behavior (further visualized in fig. S2). (**F**) The corresponding simulated potential profile for the measured gate-voltage range, which shows the ideal gate voltage (highlighted in red) that corresponds to a doping in the center of the two contact levels.

This shows that a good detection is either obtained by a high responsivity R or a high saturation power P_{in}. This detector provides linear operation for the highest input powers up to 100 mW, thereby reaching close to the largest reported photocurrents for zero-bias graphene PDs (see table S1). And, indeed, the detector has been tested for its capability in a high-speed data link in a 132 Gbit/s on-off keying experiment. These are the highest data rates yet demonstrated for a graphene PD (see fig. S1 for the data experiment results). Furthermore, destructive testing with a focused spot on similar devices resulted in damage threshold values of >1 W. So, the detector could be operated at even higher photocurrents (see materials and methods for more details). To conclude, the 100 distributed dipole antennas offer a resonant enhancement in the target wavelength range while the optical power is distributed to allow for high optical input powers. Further optimization could also lead to higher responsivities, as discussed in supplementary note 1.

To test the bandwidth of the device, we used a laser-beating scheme (see materials and methods for more details). The measured normalized radio frequency (RF) power of the device presented in Fig. 3D shows a response from 2 to 500 GHz in the measured setup range. The signal does not dip below the −3-dB line, and no clear roll-off behavior is visible. The device is still operated in a fully passive PV mode and is not biased, contrary to most other device technologies that work at high frequency. Yet a gate voltage of −2.5 V is applied because the frequency response has been found to be a function of gate voltage.

Figure 3E shows gate-voltage sweeps for different RF tones (2, 75, 150, 250, 350, 400, and 500 GHz). The RF response of the device is flat for a gate-voltage range of −2.5 ± 1 V. The −2.5-V operating point corresponds to slightly p-doped graphene, as is visible from the resistance-gate curve in the bottom panel of Fig. 3E, where we observe the Dirac point at −0.6 V. In the plot of the potential profile within the channel, −2.5 V corresponds to a center level between the calculated contact doping of the Au and Ag resonators; see the red solid line in Fig. 3F. Applying gate voltages above or beyond the range of −2.5 ± 1 V (red pointer and red dashed lines in Fig. 3F) results in diminished driving fields or even opposing currents at the contacts, leading to zero net photocurrent.

The bandwidth of a PD is limited by the resistor-capacitor (RC) time constant τ_{RC}, the relaxation time τ_{rel}, and the transit time τ_{tr}. The gate dependence of the bandwidth (see Fig. 3E) points away from any RC limit. High positive gate voltages lead to lower resistance, which results in a smaller τ_{RC} and therefore higher bandwidth. However, we observe a clear decrease in bandwidth for high positive gate voltages. We therefore exclude the influence of an RC limit in our device. This leaves the intrinsic response time of the graphene PD as $\tau_{tot} = (\tau_{rel}^{-1} + \tau_{tr}^{-1})^{-1}$. A recent publication (*41*) observed gate-dependent frequency responses and linked them to scattering and supercollision cooling effects of the carriers in graphene. In their work, higher doping levels resulted in smaller τ_{rel} values, which is opposite to our observations, where we do not observe a flat frequency for high gate voltages; see Fig. 3E. We therefore believe that our frequency response is enabled by the small τ_{tr}, which is achieved by the metamaterial structure. The localized absorption close to the resonators, which act as dopants and electrodes, leads to the short carrier-extraction pathways. The required driving force for the high-speed extraction is only achieved if the doping level is set by the appropriate gate voltage.

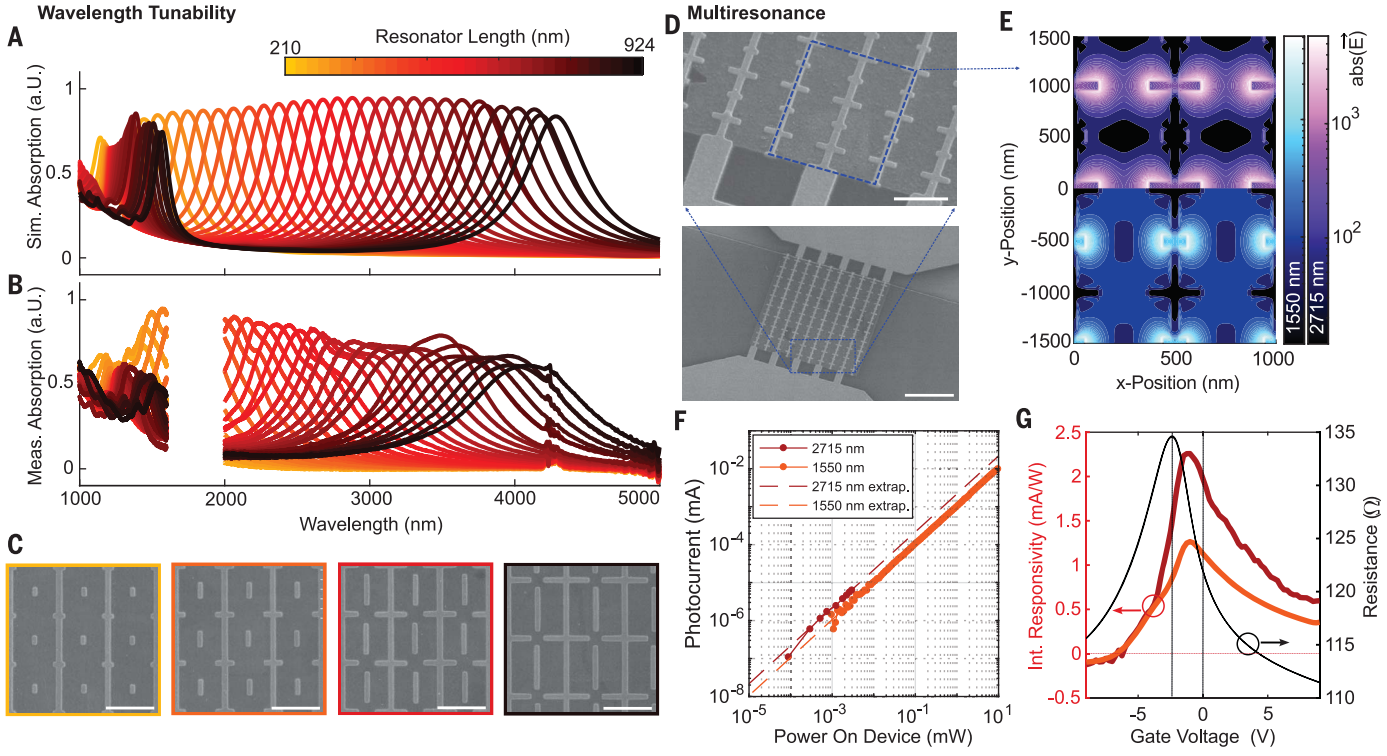


Fig. 4. Spectral tunability and multiresonance architecture. (**A** to **C**) Wavelength tunability of the metamaterial structure with the simulated (A) and measured (B) spectral absorption of different metamaterial dipole lengths, which shows the tunability of the structure. Four example scanning electron micrographs of the polarization-independent design are given in (C), where the colors correspond to the resonator-length scale bar shown in (A). Scale bar is 1 μm. (**D** to **G**) Concept of a multiresonance device. (D) Scanning electron micrographs of a dual-resonant device targeted for 1550 and 2715 nm. Scale bars are 1 μm (top) and 5 μm (bottom). (E) Corresponding electric-field simulations of the dual-resonant metamaterial of size 3 unit cells by 2 unit cells with excitation wavelengths of 1550 and 2715 nm, which show the clear dipole behavior for the two differently sized resonators. (F) Photocurrent versus optical power on device and (G) the measured responsivity for both wavelengths as a function of gate voltage. The arrows indicate the corresponding y axis for the encircled curves. extrap., extrapolated; int. responsivity, internal responsivity.

Because the experimental setup was limited to measure up to 500 GHz, the true bandwidth of the graphene PD is not yet resolved. Full RF simulations of the complete device with contacting pads predict a continuation of the flat response. However, this is under the assumption that the active area can be modeled as a lumped element and does not take into account any limitations that directly come from the intrinsic response time of the graphene's detection mechanisms. A clear prediction of the absolute bandwidth therefore requires further characterization.

Beyond telecommunication wavelengths

By making use of graphene's zero bandgap and linear dispersion that enables an almost spectrally independent absorption, the metamaterial can easily be changed to enhance the absorption at a different wavelength. Moreover, deploying a design with additional resonators that are oriented perpendicularly enables polarization-independent absorption (see figs. S3 to S5 for design logic and polarization-dependent behavior). By simply changing the dipole resonator length, the resonance and therefore the absorption can be tuned. Figure 4B shows the measured absorption spectrum for 30 different resonator lengths. Four examples of scanning electron micrographs with the respective resonators of a polarization-insensitive metamaterial are shown in Fig. 4C. The corresponding simulation results are depicted in Fig. 4A. Both the experimental and simulated spectral absorptions provided in Fig. 4, A and B, show a spectral tunability greater than 3000 nm. The experimental curves show that the shape of the peaks toward the mid-infrared range is broadened compared with that of the simulations. We attribute this to the nonideal material properties of the alumina passivation layer that is directly grown on the graphene. This presumption is supported by the comparison between simulated and measured spectra for designed nonpassivated devices as given in fig. S7, where we achieve an almost perfect match between the two.

Beyond the direct tunability of the resonator size, it is also possible to incorporate additional functionalities such as multiresonance, which enables, for example, near-infrared and mid-infrared photodetection with the same detector. Figure 4D shows two scanning electron micrographs of a device that incorporates resonators for both 1550 and 2715 nm. The 2715-nm resonators share the same connection lines and are spaced in between the same unit cell of the device used in Figs. 1 to 3. The corresponding simulated absolute electric-field intensities for both excitations are given in Fig. 4E, again showing the dipole-like behavior without any visible perturbation induced by the additional resonators. Figure 4, F and G, summarizes the performance under 1550- and 2715-nm illumination with an optical power sweep for both wavelengths (Fig. 4F) and the responsivity with an applied gate (Fig. 4G). The responsivity of the longer-wavelength resonance clearly outperforms the one at telecom wavelengths, which we attribute to the larger number of photons and the lower plasmonic losses in the resonators, which results in a higher absorption in graphene. In addition to the bandwidth and magnitude of the photoresponse, noise equivalent power (NEP) is a key figure of merit for mid-infrared–based PDs. The NEP benefits from the self-powered operation of the device and avoids the usual high dark currents of the semimetallic nature of graphene. At sufficiently high frequencies, that is, above the $1/f$ noise, we measured a noise floor on the device under zero bias that was close to both the theoretical limit (*42*) and the input noise of the measurement system. The resulting calculated NEP is then ~8 nW/√Hz at 2715 nm. This value is in the same order range as reported for other graphene-based devices that operate without any bias; see table S1 for a comparison and the materials and methods for details on the measurements.

Conclusions

Our demonstration of a 2- to >500-GHz electro-optic bandwidth PD rivals conventional PIN PD technologies (*43*) and uni-traveling carrier photodiodes (*44*). The vertical incidence metamaterial graphene PD offers many of the predicted advantages of graphene in a single device. Conceptually, the detector's performance is exploiting the advantages offered by metamaterial absorption enhancement, built-in field through graphene-metal contact doping, a well-controlled operation point by means of electrostatic gating, and effective passivation of CVD-grown graphene. The detector relies on a relatively simple layer stack of metal-insulator-graphene-metal-insulator, which potentially enables postprocessing on almost any substrate and supports a dense, monolithic integration with existing structures, similar to demonstrations with plasmonic modulators (*45*). Unlike in most prior works on graphene detectors, we show uncooled, air-stable operation with gate voltages in the complementary metal-oxide semiconductor (CMOS)–compatible, low-voltage range, which is an achievement that is due to the direct on-top–grown passivation in combination with the metamaterial bottom insulator design.

With these devices, we have demonstrated a data rate of 132 Gbit/s, which is the highest graphene data rate known to date. This high-speed detection has been enabled by the high saturation power of the detector. A high saturation power can offset a moderate responsivity, as in a shot noise–limited communication system where the SNR is proportional to the responsivity and input power. In addition, the moderate responsivities can be improved. Previous free space–illuminated graphene PD demonstrations that rely on carrier multiplication (*46*) or that are based on exfoliated multilayer graphene reached higher responsivities without any optical enhancement (*30*). Thus, there is ample room for concerted efforts to further refine the concept, improve the fabrication, and achieve higher-quality graphene (*47*, *48*). These efforts are likely to result in a new generation of graphene-based detectors with sufficient responsivities.

Finally, the high bandwidths of >500 GHz and the wavelength-independent absorption of graphene enable operation of the detector at any wavelength in a window extending from 1400 to 4200 nm and beyond. This is relevant for both sensing and communication. For example, in telecommunications, the ongoing capacity crunch has led to a strong push to unlock new communication bands (*49*–*51*). This new detector, with its compact footprint and ability to be integrated with CMOS, might thereby meet an ongoing dire need.

REFERENCES AND NOTES

1. K. S. Novoselov *et al*., *Science* **306**, 666–669 (2004).
2. A. K. Geim, K. S. Novoselov, *Nat. Mater.* **6**, 183–191 (2007).
3. W. Kong *et al*., *Nat. Nanotechnol.* **14**, 927–938 (2019).
4. D. Akinwande *et al*., *Nature* **573**, 507–518 (2019).
5. D. Neumaier, S. Pindl, M. C. Lemme, *Nat. Mater.* **18**, 525–529 (2019).
6. M. Liu *et al*., *Nature* **474**, 64–67 (2011).
7. M. Romagnoli *et al*., *Nat. Rev. Mater.* **3**, 392–414 (2018).
8. A. Montanaro *et al*., *Nat. Commun.* **12**, 2728 (2021).
9. F. Xia, T. Mueller, Y. M. Lin, A. Valdes-Garcia, P. Avouris, *Nat. Nanotechnol.* **4**, 839–843 (2009).
10. X. Gan *et al*., *Nat. Photonics* **7**, 883–887 (2013).
11. A. Pospischil *et al*., *Nat. Photonics* **7**, 892–896 (2013).
12. X. Wang, Z. Cheng, K. Xu, H. K. Tsang, J.-B. Xu, *Nat. Photonics* **7**, 888–891 (2013).
13. G. Konstantatos *et al*., *Nat. Nanotechnol.* **7**, 363–368 (2012).
14. Z. Ni *et al*., *ACS Nano* **11**, 9854–9862 (2017).
15. D. Schall, E. Pallecchi, G. Ducournau, V. Avramovic, M. Otto, D. Neumaier, paper M2I.4 presented at the Optical Fiber Communication Conference, San Diego, CA, 11 to 15 March 2018.
16. P. Ma *et al*., *ACS Photonics* **6**, 154–161 (2019).
17. Y. Ding *et al*., *Nanophotonics* **9**, 317–325 (2020).
18. K. F. Mak, L. Ju, F. Wang, T. F. Heinz, *Solid State Commun.* **152**, 1341–1349 (2012).
19. R. R. Nair *et al*., *Science* **320**, 1308–1308 (2008).
20. D. Schall *et al*., *ACS Photonics* **1**, 781–784 (2014).
21. Y. Gao, H. K. Tsang, C. Shu, *Nanoscale* **10**, 21851–21856 (2018).
22. Z. Ma *et al*., *ACS Photonics* **7**, 932–940 (2020).
23. J. E. Muench *et al*., *Nano Lett.* **19**, 7632–7644 (2019).
24. Y. Wang *et al*., *Nat. Commun.* **12**, 5076 (2021).
25. Y. Wang *et al*., *Adv. Opt. Mater.* **9**, 2001215 (2021).
26. J. Wei *et al*., *Nat. Commun.* **11**, 6404 (2020).
27. M. Freitag, T. Low, F. Xia, P. Avouris, *Nat. Photonics* **7**, 53–59 (2013).
28. F. H. L. Koppens *et al*., *Nat. Nanotechnol.* **9**, 780–793 (2014).
29. C. Liu *et al*., *Light Sci. Appl.* **10**, 123 (2021).
30. T. Mueller, F. Xia, P. Avouris, *Nat. Photonics* **4**, 297–301 (2010).
31. S. Schuler *et al*., *Nano Lett.* **16**, 7107–7112 (2016).
32. V. Mišeikis *et al*., *ACS Nano* **14**, 11190–11204 (2020).
33. S. Marconi *et al*., *Nat. Commun.* **12**, 806 (2021).
34. A. Urich, K. Unterrainer, T. Mueller, *Nano Lett.* **11**, 2804–2808 (2011).
35. J. C. Johannsen *et al*., *Phys. Rev. Lett.* **111**, 027403 (2013).
36. S. Winnerl *et al*., *J. Phys. Condens. Matter* **25**, 054202 (2013).
37. N. I. Landy, S. Sajuyigbe, J. J. Mock, D. R. Smith, W. J. Padilla, *Phys. Rev. Lett.* **100**, 207402 (2008).
38. A. A. Sagade *et al*., *Nanoscale* **7**, 3558–3564 (2015).

39. A. Dorodnyy, S. M. Koepfli, A. Lochbaum, J. Leuthold, *Sci. Rep.* **10**, 17941 (2020).
40. G. Giovannetti *et al*., *Phys. Rev. Lett.* **101**, 026803 (2008).
41. K. Yoshioka *et al*., *Nat. Photonics* **16**, 718–723 (2022).
42. Y. Fang, A. Armin, P. Meredith, J. Huang, *Nat. Photonics* **13**, 1–4 (2019).
43. S. Lischke *et al*., *Nat. Photonics* **15**, 925–931 (2021).
44. T. Kurokawa, T. Ishibashi, M. Shimizu, K. Kato, T. Nagatsuma, *Electron. Lett.* **54**, 705–706 (2018).
45. U. Koch *et al*., *Nat. Electron.* **3**, 338–345 (2020).
46. S. Cakmakyapan, P. K. Lu, A. Navabi, M. Jarrahi, *Light Sci. Appl.* **7**, 20 (2018).
47. M. Wang *et al*., *Nature* **596**, 519–524 (2021).
48. J. Li *et al*., *Nat. Mater.* **21**, 740–747 (2022).
49. P. J. Winzer, D. T. Neilson, *J. Lightwave Technol.* **35**, 1099–1115 (2017).
50. Y. Chen, Z. Xie, J. Huang, Z. Deng, B. Chen, *Optica* **6**, 884–889 (2019).
51. J. Guo *et al*., *Light Sci. Appl.* **9**, 29 (2020).

ACKNOWLEDGMENTS

We thank the Cleanroom and Operations team of the Binning and Rohrer Nanotechnology Center (BRNC) for their support. We thank Virginia Diodes, Inc. (VDI) for lending additional equipment. We thank M. S.-B. Hossain from the Christian-Albrechts-Universität zu Kiel for help with the digital signal processing in the data experiments. **Funding:** This work was funded by the European Union's Horizon 2020 research and innovation program through project aCryComm, FET Open Grant agreement no. 899558. Project SuperQuant (20FUN07) has received funding from the EMPIR program, which is cofinanced by the participating states and by the European Union's Horizon 2020 research and innovation program. **Author contributions:** Conceptualization: S.M.K., J.L.; Methodology: S.M.K., Y.K., R.G., Y.H., A.G., K.K., S.N., R.S., E.P., Y.F.; Investigation: S.M.K., M.B.; Visualization: S.M.K., M.D.; Funding acquisition: S.M.K., J.L.; Project administration: S.M.K., J.L.; Supervision: J.L., Y.F.; Writing – original draft: S.M.K.; Writing – review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg8017
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (*52*–*63*)

Submitted 23 January 2023; accepted 11 May 2023
10.1126/science.adg8017

CATALYSIS

Size of cerium dioxide support nanocrystals dictates reactivity of highly dispersed palladium catalysts

Valery Muravev[1], Alexander Parastaev[1], Yannis van den Bosch[1], Bianca Ligt[1], Nathalie Claes[2], Sara Bals[2], Nikolay Kosinov[1], Emiel J. M. Hensen[1]*

The catalytic performance of heterogeneous catalysts can be tuned by modulation of the size and structure of supported transition metals, which are typically regarded as the active sites. In single-atom metal catalysts, the support itself can strongly affect the catalytic properties. Here, we demonstrate that the size of cerium dioxide (CeO_2) support governs the reactivity of atomically dispersed palladium (Pd) in carbon monoxide (CO) oxidation. Catalysts with small CeO_2 nanocrystals (~4 nanometers) exhibit unusually high activity in a CO-rich reaction feed, whereas catalysts with medium-size CeO_2 (~8 nanometers) are preferred for lean conditions. Detailed spectroscopic investigations reveal support size–dependent redox properties of the Pd-CeO_2 interface.

The development of sustainable and efficient chemical processes relies heavily on catalysts. Most heterogeneous catalysts are based on transition metal nanoparticles (NPs) finely dispersed over an inexpensive solid metal oxide support. The high surface-to-volume ratio of supported NPs increases metal utilization, which is particularly important for structure-insensitive reactions that are catalyzed by precious noble metals (*1*). To maximize the use of these highly active yet costly metals, catalysts with atomically dispersed metals (single-atom catalysts, SACs) are being developed extensively (*2*–*5*).

Decreasing the size of noble metal species affects their reactivity (*1*, *6*–*8*). For example, size-induced perturbations in the electronic structure of few-nanometer Au clusters on TiO_2 render them very active in low-temperature CO oxidation, in contrast to larger gold NPs (*9*–*11*). However, further dispersing metal entities down to single atoms can also adversely affect catalytic activity. For example, clusters of metallic Pt supported on silica (SiO_2) and zirconia (ZrO_2) are more reactive toward CO oxidation than are isolated Pt species (*12*). In SACs, every (noble) metal center is coordinated by the atoms of the support. Therefore, the chemical properties of the support and metal-support interactions (MSIs) strongly modulate the activity of the catalyst (*13*–*16*).

CeO_2 (also known as ceria), a key component of automotive catalytic converters, is particularly known for strong MSIs with noble metals, rendering them stable in highly dispersed form (*17*–*19*). To limit agglomeration of the active phase in noble-metal SACs, low metal loadings, high–surface area (nanosized) supports, or both are commonly used (*20*–*24*). Nanostructuring can notably change the chemical properties of CeO_2 (*25*–*28*). Combining a surface science approach with theoretical calculations, Vayssilov *et al.* demonstrated that lattice O atoms can spontaneously migrate from nanosized CeO_2 to Pt clusters (~2 nm), a process that is energetically unfavorable on extended CeO_2 surfaces (*29*). Carrettin *et al.* used small (~4 nm) crystallites instead of bulk CeO_2 as a support for ~3-nm gold NPs to achieve a two orders of magnitude increase in low-temperature CO oxidation activity (*30*).

The effect of CeO_2 particle size on the catalytic properties of atomically dispersed noble metals is as yet unexplored. We used flame spray pyrolysis (FSP) (*31*) to prepare in a single step Pd-CeO_2 SACs with variable size of the support (4 to 18 nm). Using CO oxidation as a probe reaction, we demonstrated that the reactivity of the prepared nanocomposites strongly depended on the size of CeO_2 particles. Through a combination of advanced in situ spectroscopy tools and steady-state kinetic studies, the strong impact of the size-dependent redox properties of the Pd-CeO_2 interface on the catalytic performance is highlighted. This work demonstrates an unconventional approach to tailoring the properties of metal-based catalysts by changing the size of the support.

Pd catalysts with different sizes of CeO_2 support

We used FSP to prepare CeO_2 of variable sizes (*32*). The size and redox properties of the obtained CeO_2 NPs were controlled by the injection rate of the liquid feed containing a Ce precursor during FSP (figs. S1 to S4). We prepared 1 wt % Pd/CeO_2 catalysts (PdFSP) in a one-step FSP process using a solution containing both metal precursors, which resulted in partial doping of CeO_2 with the noble metal (*31*). High-resolution high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) images (Fig. 1A; for bright-field STEM images, see fig. demonstrated that octahedral NPs of Pd/ were synthesized. The size range from 4 to 18 nm was obtained by changing the injection rate from 1 to 15 ml/min (table S1). Throughout this work, we focus on three principal sizes of PdFSP catalysts: (i) small ~4-nm NPs, (ii) medium ~8-nm NPs, and (iii) large ~13-nm NPs (Fig. 1A).

Elemental mapping by energy-dispersive x-ray spectroscopy (EDX) revealed the highly dispersed nature of Pd species in small and medium PdFSP NPs, whereas additional Pd clusters of ~2 nm could be seen on large PdFSP after prolonged electron beam exposure (figs. S7 and S8). Synchrotron-based powder x-ray diffraction (XRD) exhibited exclusively CeO_2 reflections, indicating the absence of Pd/PdO phases (Fig. 1B). The sizes of PdFSP crystallites derived from XRD data (fig. S9) were in good agreement with microscopy and physisorption data (fig. S10 and table S1).

The Pd-CeO_2 interactions in PdFSP nanocomposites were strongly affected by the size of CeO_2, as determined by the temperature-programmed reduction in hydrogen (H_2-TPR) (Fig. 1C and fig. S11). The first peak of reduction (α) in small PdFSP NPs that appeared at ~200°C indicated a strong interaction between CeO_2 and Pd that is typical for a highly dispersed Pd-oxo species (*33*). As the size of the PdFSP NPs increased, this feature shifted to lower temperature and finally split into two peaks (α′ and α″), indicating weaker MSIs in these samples. We attributed an additional reduction peak for large PdFSP NPs to the presence of PdO_x clusters.

Quantification of H_2-TPR data (table S2) revealed that the reducibility of PdFSP nanocomposites strongly depended on their size, even when normalized to the total surface area. From analysis of x-ray photoelectron spectroscopy (XPS) data, we concluded that Pd in small and medium PdFSP samples (Fig. 2A and fig. S12) was present as highly dispersed Pd^{2+} species strongly interacting with CeO_2 (Pd $3d_{5/2}$ binding energy, ~338 eV) (*31*, *34*). For large PdFSP NPs (>10 nm), an additional weak component at ~336 eV was observed, which we attributed to PdO_x (where $x < 1$) clusters (*31*, *35*).

The fitting of Pd K-edge extended x-ray absorption fine structure (EXAFS) revealed the predominantly isolated nature of Pd-oxo species with about four O atoms in the first coordination shell and two to three atoms of Ce in the second shell (Fig. 2B, fig. S13, and table S3) for all samples. However, a weak contribution of Pd–Pd scattering at ~2.7 Å with a coordination number (CN) ~1 in the EXAFS of large PdFSP NPs pointed to the presence of small PdO_x clusters. To determine the speciation of Pd at the very surface of the NPs, we used diffuse reflectance infrared Fourier-transform

[1]Laboratory of Inorganic Materials and Catalysis, Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands. [2]Electron Microscopy for Materials Science (EMAT), University of Antwerp, 2020 Antwerp, Belgium.
*Corresponding author. Email: e.j.m.hensen@tue.nl

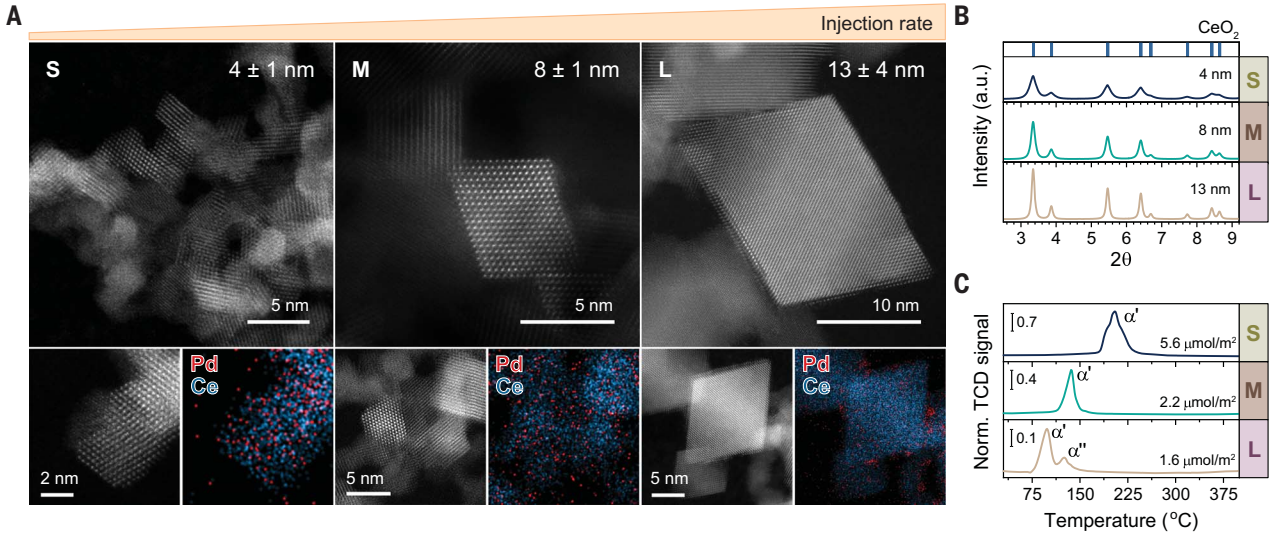


Fig. 1. Structure and reducibility of FSP-derived 1 wt % Pd-CeO_2 nanocomposites with varying size of the CeO_2 support. (**A**) HAADF STEM imaging and EDX elemental mapping of prepared materials. Labels "S", "M," and "L" refer to catalysts with small (~4 nm), medium (~8 nm), and large (~13 nm) CeO_2, respectively. (**B**) High-resolution synchrotron-based powder XRD data (λ = 0.18233 Å). (**C**) H_2-TPR profiles and hydrogen consumption quantification data.

spectroscopy (DRIFTS) of CO adsorbed at −20°C (Fig. 2C). The in situ DRIFTS spectra of small and medium PdFSP NPs exhibited one major infrared band at ~2140 cm^{-1} assigned to CO linearly adsorbed on oxidized single-atom Pd species (*31*, *33*, *36*, *37*). The additional CO bands observed at lower frequencies (2096, 2050, and 1900 cm^{-1}) for larger PdFSP NPs could be linked to the presence of semireduced and clustered Pd species (*31*, *33*, *36*, *37*).

Both the overall intensity of carbonyl bands in DRIFTS spectra and the Pd/Ce surface atomic ratios derived from XPS increased with CeO_2 size (table S1). These results indicated surface enrichment of Pd on large CeO_2 NPs, which could in turn cause lower Pd dispersion. To test this assumption, we prepared a series of reference Pd/CeO_2 samples using wet impregnation of FSP-made CeO_2 NPs of different sizes (denoted as PdFSPimp; figs. S14 to S18 and table S4). The CO DRIFTS spectra of the low-loaded (0.46 wt %) PdFSPimp sample with large CeO_2 NPs were similar to those of its 1 wt % counterparts prepared by impregnation and the one-step FSP method, displaying bands related to agglomerated Pd species (fig. S19).

We inferred that weak MSIs are the most likely reason for the observed clustering of Pd on large CeO_2. The size of CeO_2 NPs strongly influenced the Pd speciation in both one-step FSP-prepared (PdFSP) and wet-impregnated (PdFSPimp) samples. Strong MSIs in PdFSP NPs of 8 nm and below caused Pd to be exclusively present as isolated sites on and in the CeO_2 surface. Weaker MSIs in large PdFSP NPs (>10 nm) led to the coexistence of atomically dispersed and clustered Pd species in the as-prepared catalyst.

Pd speciation during catalytic CO oxidation

We used CO oxidation to determine the activity of FSP-made Pd/CeO_2 nanocomposites. Each catalyst was pretreated in O_2 and then three light-off measurements from −20° to 300°C were performed (fig. S20). All PdFSP catalysts could oxidize CO at room temperature (Fig. 3A and fig. S21). The low-temperature CO oxidation activity (<100°C) strongly depended on the size of the CeO_2 support. At moderate temperature (~150°C), small PdFSP NPs displayed a lower activity than medium and large PdFSP NPs, which already achieved full conversion at this temperature. According to light-off catalytic tests, CeO_2 NPs of ~8 nm were optimal for low-temperature CO oxidation on PdFSP NPs (Fig. 3B). Despite the difference in preparation method, the optimal support size for impregnated (PdFSPimp) catalysts was similar (9 nm) and followed the same overall activity trend (fig. S22).

We used in situ DRIFTS to shed light on the speciation of Pd under low-temperature CO oxidation conditions (75°C). Fig. 3C shows that Pd single atoms, represented by the ~2140 cm^{-1} carbonyl band, remained dominant in small PdFSP NPs, even after reaction at 300°C. These sites were particularly active in CO oxidation at low temperature (*31*, *33*, *36*, *37*). A higher concentration of atomically dispersed Pd in the medium-sized sample, manifested by a more intense ~2140 cm^{-1} band, could explain its superior activity at low temperatures (Fig. 3B).

In contrast to the as-prepared state, low-frequency bands (<2100 cm^{-1}) were also present and likely indicated the formation of metallic Pd clusters during CO oxidation. Metallic Pd species were less active than Pd single atoms at low temperature but could substantially contribute to overall activity at temperatures >125°C (*31*, *38*). On large PdFSP NPs, the fraction of metallic and clustered Pd species was considerably higher than that of atomically dispersed Pd, which explained its limited activity at low temperature. The plateauing of the CO oxidation activity at a temperature of ~125°C (Fig. 3A) on medium and small NPs predominantly populated by isolated Pd species stemmed from the weaker adsorption of CO on single-atom Pd sites compared with metallic sites (fig. S23).

To determine the stability of Pd-oxo species at elevated temperature (300°C), we used surface sensitive synchrotron-based near-ambient pressure XPS (NAP-XPS). During CO oxidation at elevated temperatures, Pd species were stable in the form of highly dispersed Pd–O_x–Ce moieties (~338 eV) (Fig. 3D). DRIFTS for medium and large PdFSP NPs indicated the formation of reduced Pd clusters at low temperature, and even more so after heating up to 300°C. We speculate that these clusters were of low nuclearity and may redisperse during CO oxidation at high temperature, akin to the dynamics observed for Pt clusters on CeO_2 (*39*). This hypothesis was supported by the minor contribution of Pd–Pd scattering (CN_{Pd-Pd} ~1) in EXAFS of used medium and large PdFSP samples (fig. S24). As we noted above, redispersion during CO oxidation at elevated temperature does not occur for larger Pd/PdO NPs on CeO_2 (*38*).

The resistance to sintering of single-atom Pd species in small PdFSP NPs originated from

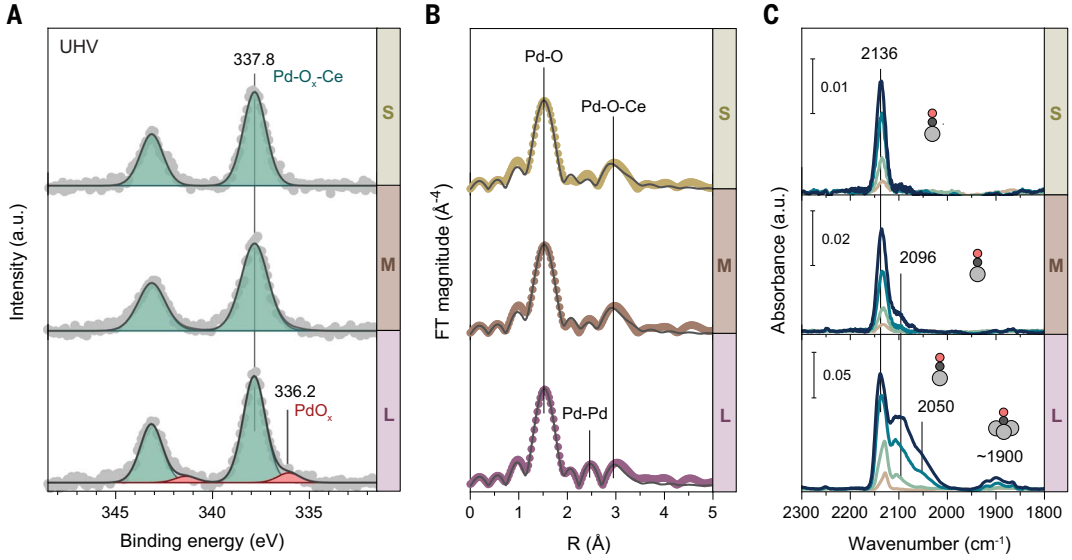


Fig. 2. Spectroscopic characterization of PdFSP samples with small (S), medium (M), and large (L) CeO_2 NPs. (**A**) XPS spectra of Pd 3d core line region acquired for fresh PdFSP samples. (**B**) R-space k^3-weighted EXAFS spectra of fresh PdFSP samples acquired at the Pd K-edge. (**C**) In situ DRIFTS during CO adsorption at −20°C on fresh PdFSP samples. The CO was set to 0.3 ml/min, resulting in a partial pressure of 2 mbar in the DRIFTS cell.

the strong MSI. With increasing size of CeO_2, these interactions became weaker, which resulted in lower stability of atomically dispersed Pd under reaction conditions. The facile transfer of O from the CeO_2 support to Pd sites strongly affected both the activity and stability of Pd-oxo species (*31*, *33*). Therefore, it is possible that the observed CeO_2 support size effects stem from the difference in redox properties and O mobility in the prepared nanocomposites.

Oxygen mobility at Pd-CeO_2 interface

We used in situ resonant photoelectron spectroscopy (RPES) to study the redox properties of the Pd-CeO_2 interface in the prepared nanocomposites. This technique allowed us to monitor the Ce^{4+}/Ce^{3+} ratio at the very surface (inelastic mean free path, ~5 Å), minimizing the contribution of bulk Ce atoms (*29*, *34*). Because the surface-to-volume ratio changes with the size of PdFSP NPs, RPES is a particularly well-suited spectroscopic probe. The valence band photoelectron spectra of small PdFSP NPs (Fig. 4A) exhibited the highest relative intensity of the Ce^{3+} resonant peak ($D_{Ce}{}^{3+}$) with respect to the Ce^{4+} peak ($D_{Ce}{}^{4+}$) and to off-resonance spectra. As shown in table S5, the fraction of Ce^{3+} in CO at 175°C decreased with the size of CeO_2 NPs, suggesting size-dependent reducibility of the surface. The reduction of Pd-oxo species was incomplete under these conditions, and medium PdFSP NPs displayed pronounced stabilization of Pd–O_x–Ce moieties (fig. S25). We infer that O removal by CO involved both Pd and Ce cationic species and was influenced by the size of the support, which is consistent with the temperature-programmed reduction in CO (fig. S26) and pulsing CO titration (table S6) data.

The size-induced changes in redox properties of PdFSP NPs strongly altered the kinetics of CO oxidation. For catalysts with ~4-nm CeO_2, the reaction order in CO was particularly high (+1.3), whereas it decreased substantially for larger NPs (>7 nm) to +0.2 for 13 nm PdFSP NPs (Fig. 4B). The opposite trend was observed for the reaction orders in O_2. For the small PdFSP NPs, it was unusually low (−0.2) and increased to +0.3 for large NPs. The trend in reaction orders was similar for the PdFSPimp catalysts (table S7). These data suggested that CO oxidation on large PdFSP NPs likely followed the Mars-van Krevelen mechanism, which involves O transfer from the support typical for conventional Pd-CeO_2 catalysts (*38*).

The unusual reaction orders observed for small PdFSP NPs indicate that O_2 poisoned Pd sites in catalysts with small CeO_2 NPs, necessitating high pressures of CO to achieve high catalytic activity. In a recent theoretical work, such reactivity was ascribed to Pd redox transitions between several formal oxidation states of metal during a catalytic cycle (*40*). In fact, as shown in Fig. 4C, under CO-rich reaction conditions (CO:O_2 = 4:1), small PdFSP NPs exhibited much higher activity than the larger PdFSP NPs. Catalysts highly active under these demanding conditions could contribute to solving the cold-start problem of current emission control systems, in which the exhaust feed, which is initially rich in CO, limits the activity of conventional noble-metal NP catalysts (*41*, *42*).

To determine directly the evolution of the CO coverage on Pd sites, we performed in situ DRIFTS experiments under conditions similar to those used in the reaction orders study (Fig. 4D). A higher partial pressure of CO for large PdFSP NPs caused only minor changes in Pd speciation and in the intensity of the $CO_{2(g)}$ infrared band, reflecting the CO oxidation activity. For medium PdFSP NPs, adding more CO increased the activity but also caused reduction and aggregation of Pd-oxo species, as manifested by an increase in the intensity of low-frequency bands (<2100 cm^{-1}). By contrast, no additional bands appeared in the spectra of small PdFSP NPs exposed to a CO-rich atmosphere. Consistent with the high CO reaction order, the increased CO_2 evolution and intensity of the ~2140 cm^{-1} band suggested that CO was competing with O_2 on single-atom Pd sites during the reaction. We argue that the high stability of Pd-oxo species in small PdFSP NPs and their poisoning by O_2 arose from facile O transport at the Pd-CeO_2 interface.

Further evidence for the high O mobility in PdFSP catalysts with small CeO_2 NPs was obtained from in situ Raman spectroscopy coupled with isotope labeling experiments. After the size of CeO_2 crystallites in PdFSP NPs decreased, the most intense Ce–O band located at 463 to 465 cm^{-1} (F_{2g}) broadened, whereas the relative intensity of bands assigned to intrinsic and extrinsic defects at 600 to 660 cm^{-1} (D_1 and D_2; see fig. S27) increased (Fig. 4, E to G) (*43*, *44*). When the O_2 feed was replaced by a CO + O_2 reaction mixture at 75°C, the band at 829 cm^{-1}, attributed to adsorbed peroxo $O_2{}^{2-}$ species, became more intense for all samples, indicating that there was a higher concentration of O vacancies (O_v) under reaction conditions (*33*, *43*).

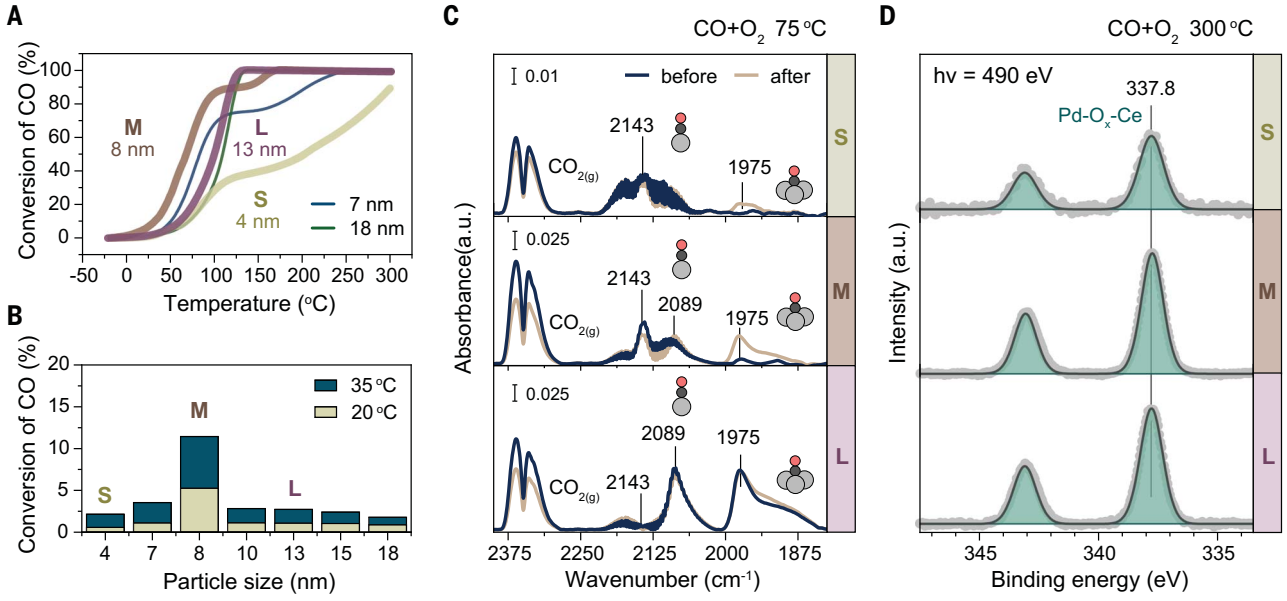


Fig. 3. Catalytic performance and Pd speciation under reaction conditions. (**A**) CO oxidation light-off curves from −20 to 300°C (third run). (**B**) Low-temperature activity as a function of CeO_2 size in PdFSP catalysts. Reaction conditions: 1% CO and 1% O_2 in He (GHSV ~180,000 ml g_{cat}^{-1} h^{-1}). (**C**) In situ DRIFTS spectra acquired during CO oxidation at 75°C before and after heating up to 300°C in reaction mixture. Reaction conditions: 1% CO and 1% O_2 in He at a total flow of 100 ml/min. (**D**) In situ NAP-XPS Pd 3d spectra of the catalysts exposed to an equimolar CO + O_2 mixture (1 mbar) at 300°C. Spectra were acquired at 490 eV, corresponding to an inelastic electron mean free path (IMFP) of ~5 Å.

In the absence of Pd, the increase in intensity of this peroxo peak was very small upon switching from O_2 to CO + O_2 (fig. S28). Accordingly, we conclude that noble metal species induce the formation of O_v and facilitate O_2 activation during low-temperature CO oxidation. The groups of Kwak and Corma arrived at a similar conclusion for Pd-CeO_2 and Au-CeO_2 catalysts, respectively (*33*, *45*, *46*). To test the possibility of CeO_2 lattice O exchange with gas-phase O_2 under actual reaction conditions, we replaced the $C^{16}O + {}^{16}O_2$ feed with $C^{16}O + {}^{18}O_2$ at 75° and 175°C (Fig. 4, E to G). As expected from surface adsorbates, the peroxo species exhibited a strong red shift (from 829 to 782 cm^{-1}) after complete replacement by ${}^{18}O_2{}^{2-}$ species (*47*). After prolonged exposure (8 hours), a slight red shift of the F_{2g} band was most pronounced for small PdFSP NPs with respect to larger NPs, as evidenced by the difference spectra (Fig. 4, E to G, bottom panel). Such a shift indicated a partial exchange of ${}^{18}O_2$ with Ce–${}^{16}O$ moieties (*48*), the extent of which increased at 175°C. The shift in the F_{2g} band was negligible in Pd-free sample regardless of temperature (fig. S28). In situ Raman data emphasize the distinctive redox properties of PdFSP catalysts with small CeO_2 NPs that give rise to facile O transfer at the Pd-CeO_2 interface.

Discussion

Aimed at maximum utilization of noble metals, the design of highly dispersed catalysts often involves using nanosized metal oxide support materials with a large surface area. We have demonstrated that the size of the CeO_2 support could strongly affect the reactivity of highly dispersed Pd-CeO_2 catalysts toward oxidation of CO. The observed size dependence was linked to different redox properties of the FSP-prepared Pd-CeO_2 nanocomposites. The strong MSIs between Pd and CeO_2 and the high O mobility in small PdFSP NPs (4 nm) led to O_2 poisoning of Pd single atoms and limited the activity of the catalyst in low-temperature CO oxidation. Moderate reducibility of CeO_2 NPs and O mobility at the metal-support interface in medium PdFSP NPs (8 nm) gave rise to a lower reaction order in CO and earlier CO oxidation light-off. Because of weaker MSI and muted O transfer, single Pd atoms are prone to reduction and sintering in large PdFSP NPs (13 nm), which leads to inferior activity of the catalyst at low temperature.

We have shown here that by changing the size of CeO_2 support, we could fine-tune the kinetics of low-temperature CO oxidation and achieve high activity, even in CO-rich reaction mixtures. Our results show that high O mobility leads to stabilization of atomically dispersed noble metals on small CeO_2 NPs, consistent with the recently reported unusual durability of few-atom Pd, Pt, and Rh species on ~5-nm CeO_2 supported on alumina (*37*). The concept of support size–dependent redox properties of the metal-CeO_2 interface could be extended to other important catalytic reactions involving Ce^{3+}/Ce^{4+} dynamics, such as CO_2 hydrogenation (*49*), the water-gas shift reaction (*48*), and propane dehydrogenation (*50*).

REFERENCES AND NOTES

1. C. Vogt, B. M. Weckhuysen, *Nat. Rev. Chem.* **6**, 89–111 (2022).
2. A. Wang, J. Li, T. Zhang, *Nat. Rev. Chem.* **2**, 65–81 (2018).
3. S. K. Kaiser, Z. Chen, D. Faust Akl, S. Mitchell, J. Pérez-Ramírez, *Chem. Rev.* **120**, 11703–11809 (2020).
4. X. F. Yang *et al.*, *Acc. Chem. Res.* **46**, 1740–1748 (2013).
5. L. Liu, A. Corma, *Chem. Rev.* **118**, 4981–5079 (2018).
6. W. E. Kaden, T. Wu, W. A. Kunkel, S. L. Anderson, *Science* **326**, 826–829 (2009).
7. J. Li, X. Li, H. J. Zhai, L. S. Wang, *Science* **299**, 864–867 (2003).
8. D. A. J. M. Ligthart, R. A. van Santen, E. J. M. Hensen, *Angew. Chem. Int. Ed.* **50**, 5306–5310 (2011).
9. M. Valden, X. Lai, D. W. Goodman, *Science* **281**, 1647–1650 (1998).
10. A. A. Herzing, C. J. Kiely, A. F. Carley, P. Landon, G. J. Hutchings, *Science* **321**, 1331–1335 (2008).
11. M. S. Chen, D. W. Goodman, *Science* **306**, 252–255 (2004).
12. K. Ding *et al.*, *Science* **350**, 189–192 (2015).
13. J. Li *et al.*, *J. Am. Chem. Soc.* **141**, 14515–14519 (2019).
14. P. Hu *et al.*, *Angew. Chem. Int. Ed.* **53**, 3418–3421 (2014).
15. T. W. van Deelen, C. Hernández Mejía, K. P. de Jong, *Nat. Catal.* **2**, 955–970 (2019).
16. R. Lang *et al.*, *Chem. Rev.* **120**, 11986–12043 (2020).
17. A. Beniya, S. Higashi, *Nat. Catal.* **2**, 590–602 (2019).
18. J. A. Farmer, C. T. Campbell, *Science* **329**, 933–936 (2010).
19. R. J. Farrauto, M. Deeba, S. Alerasool, *Nat. Catal.* **2**, 603–613 (2019).
20. L. DeRita *et al.*, *J. Am. Chem. Soc.* **139**, 14150–14165 (2017).
21. J. Resasco *et al.*, *J. Am. Chem. Soc.* **142**, 169–184 (2020).
22. Y. Lu *et al.*, *Nat. Catal.* **2**, 149–156 (2019).
23. J. Fu *et al.*, *Nat. Catal.* **3**, 446–453 (2020).
24. B. Qiao *et al.*, *Nat. Chem.* **3**, 634–641 (2011).
25. A. Corma, P. Atienzar, H. García, J.-Y. Chane-Ching, *Nat. Mater.* **3**, 394–397 (2004).
26. D. Prieur *et al.*, *Inorg. Chem.* **59**, 5760–5767 (2020).
27. C. Paun *et al.*, *J. Phys. Chem. C* **116**, 7312–7317 (2012).
28. J. Xu *et al.*, *Chem. Commun.* **46**, 1887–1889 (2010).
29. G. N. Vayssilov *et al.*, *Nat. Mater.* **10**, 310–315 (2011).
30. S. Carrettin, P. Concepción, A. Corma, J. M. López Nieto, V. F. Puntes, *Angew. Chem. Int. Ed.* **43**, 2538–2540 (2004).

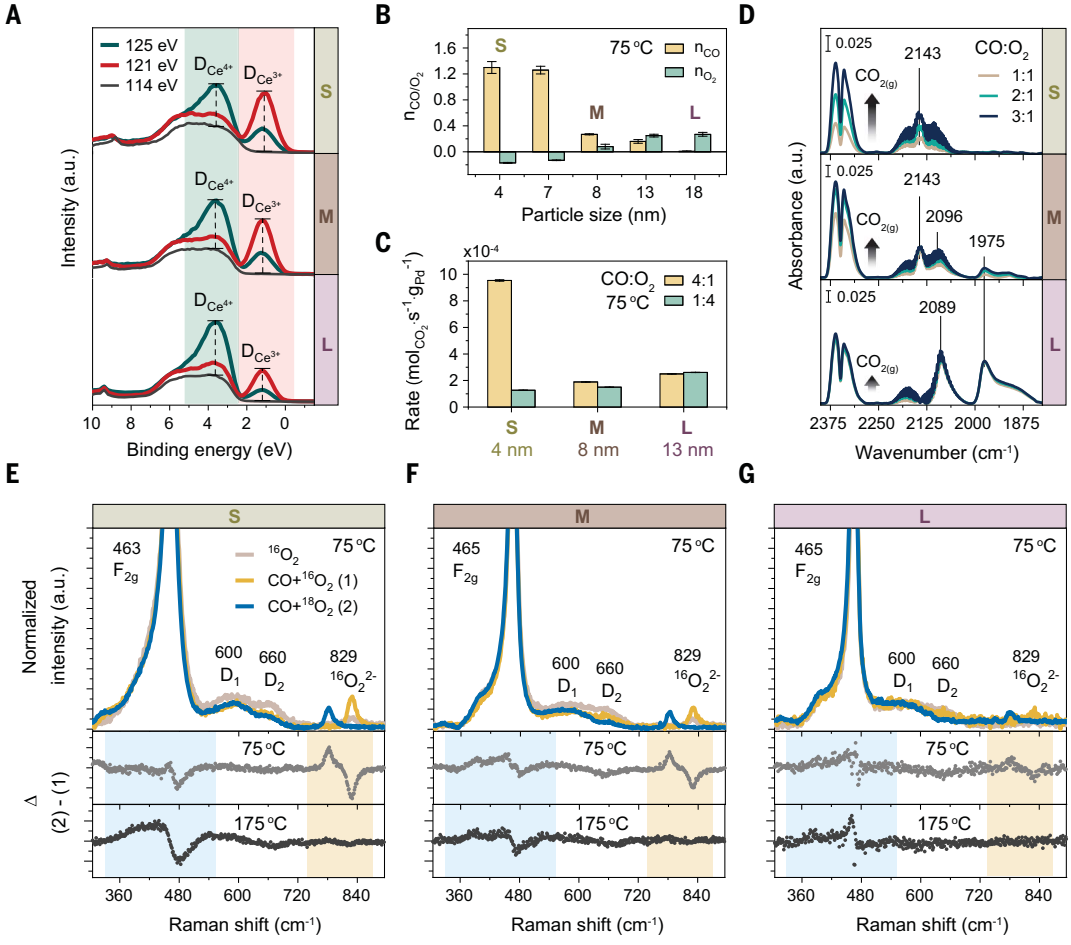


Fig. 4. CO oxidation kinetics and O transfer at Pd-CeO_2 interface. (**A**) In situ RPES valence band spectra of the catalysts exposed to 1 mbar of CO at 175°C. The intensities of the Ce^{4+} and Ce^{3+} spectral features, which appear at the respective resonant energies (125 eV for Ce^{4+}; 121 eV for Ce^{3+}; 114 eV for off-resonance), were used to estimate the Ce^{3+} atomic fraction. (**B**) Reaction orders measured at 75°C as a function of CeO_2 size in PdFSP catalysts. Before measurements, catalysts were stabilized in the reaction mixture at 75°C for ~14 hours. GHSV ~240,000 ml g_{cat}^{-1} h^{-1}, conversion kept below 8%. (**C**) Steady-state activity at 75°C in rich (CO:O_2 = 4:1) and lean (CO:O_2 = 1:4) conditions. GHSV ~240,000 ml g_{cat}^{-1} h^{-1}, conversion kept below 8%. (**D**) In situ DRIFTS study of CO partial pressure effect on CO oxidation at 75°C on PdFSP samples. Reaction conditions: 1 to 3% of CO and 1% of O_2 in He at a total flow of 100 ml/min. Before changing the partial pressures, the catalysts were stabilized on stream for ~30 min. (**E** to **G**) Normalized in situ Raman spectra for PdFSP samples in various gas environments. The spectra in isotope-labeled mixture were acquired after 8 hours of exposure. Reaction conditions: 1% CO and 1% $^{16}O_2$ ($^{18}O_2$) in He at a total flow of 100 ml/min.

31. V. Muravev *et al.*, *Nat. Catal.* **4**, 469–478 (2021).
32. L. Mädler, W. J. Stark, S. E. Pratsinis, *J. Mater. Res.* **17**, 1356–1362 (2002).
33. Y. Kim *et al.*, *Angew. Chem. Int. Ed.* **60**, 22769–22775 (2021).
34. A. Neitzel *et al.*, *J. Phys. Chem. C* **120**, 9852–9862 (2016).
35. O. A. Stonkus, T. Y. Kardash, E. M. Slavinskaya, V. I. Zaikovskii, A. I. Boronin, *ChemCatChem* **11**, 3505–3521 (2019).
36. D. Jiang *et al.*, *ACS Catal.* **10**, 11356–11364 (2020).
37. H. Jeong *et al.*, *Nat. Catal.* **3**, 368–375 (2020).
38. V. Muravev *et al.*, *Angew. Chem. Int. Ed.* **61**, e202200434 (2022).
39. F. Maurer *et al.*, *Nat. Catal.* **3**, 824–833 (2020).
40. L. Wang *et al.*, *ACS Catal.* **12**, 12927–12941 (2022).
41. D. S. Lafyatis *et al.*, *Appl. Catal. B* **18**, 123–135 (1998).
42. D. S. Lafyatis, T. H. Ballinger, G. S. Lammey, J. C. Frost, B. J. Cooper, *Stud. Surf. Sci. Catal.* **121**, 383–386 (1999).
43. S. Loridant, *Catal. Today* **373**, 98–111 (2020).
44. E. Sartoretti *et al.*, *Sci. Rep.* **9**, 3875 (2019).
45. J. Guzman *et al.*, *Angew. Chem. Int. Ed.* **44**, 4778–4781 (2005).
46. J. Guzman, S. Carrettin, A. Corma, *J. Am. Chem. Soc.* **127**, 3286–3287 (2005).
47. Z. Wu, M. Li, J. Howe, H. M. Meyer 3rd, S. H. Overbury, *Langmuir* **26**, 16595–16606 (2010).
48. C. Schilling, C. Hess, *ACS Catal.* **9**, 1159–1171 (2019).
49. A. Parastaev *et al.*, *Nat. Catal.* **3**, 526–533 (2020).
50. F. Xing, Y. Nakaya, S. Yasumura, K. Shimizu, S. Furukawa, *Nat. Catal.* **5**, 55–65 (2022).

ACKNOWLEDGMENTS

We thank the staff of the MAX IV Laboratory for time on beamline SPECIES under proposals 20200412 and 20190983; E. Kokkonen and A. Klyushin for assistance with NAP-XPS and RPES experiments conducted at SPECIES; staff of the MAX IV Laboratory for time on beamline BALDER under proposal 20200378; K. Klementiev for assistance with XAS measurements; J. Drnec at the ESRF for providing assistance in using beamline ID31; and V. Perez-Dieste and I. Villar Garcia at the CIRCE beamline at ALBA Synchrotron for help with acquiring preliminary RPES data obtained under proposal 2020024219. The synchrotron-based XRD measurements were performed on beamline ID31 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. **Funding:** This work was supported by the Netherlands Center for Multiscale Catalytic Energy Conversion (MCEC), a NWO Gravitation program funded by the Ministry of Education, Culture and Science of the Government of the Netherlands (V.M. and E.J.M.H.); the European Research Council (ERC consolidator grant 815128 REALNANO to S.B. and N.C.); and the European Union's Horizon 2020 Research and Innovation Program (grant 823717–ESTEEM to S.B. and N.C). Research conducted at MAX IV, a Swedish national user facility, is supported by the Swedish Research council under contract 2018-07152, the Swedish Governmental Agency for Innovation Systems under contract 2018-04969, and Formas under contract 2019-02496 (VM). **Author contributions:** Conceptualization: V.M., E.J.M.H.; Funding acquisition: E.J.M.H.; Investigation: V.M., A.P., Y.V.B., B.L., N.C., N.K.; Methodology: V.M., A.P., N.K., S.B., E.J.M.H.; Project administration: E.J.M.H.; Supervision: V.M., E.J.M.H.; Visualization: V.M.; Writing – original draft: V.M.; Writing – review and editing: V.M., N.K., E.J.M.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf9082
Materials and Methods
Supplementary Text
Figs. S1 to S31
Tables S1 to S8
References (*51*–*69*)

Submitted 19 November 2022; accepted 15 May 2023
10.1126/science.adf9082

MAGNETISM

Proximate deconfined quantum critical point in $SrCu_2(BO_3)_2$

Yi Cui[1]†, Lu Liu[2,3]†, Huihang Lin[1]†, Kai-Hsin Wu[4], Wenshan Hong[2], Xuefei Liu[1], Cong Li[1], Ze Hu[1], Ning Xi[1], Shiliang Li[2,5,6], Rong Yu[1,7]*, Anders W. Sandvik[4,2]*, Weiqiang Yu[1,7]*

The deconfined quantum critical point (DQCP) represents a paradigm shift in quantum matter studies, presenting a "beyond Landau" scenario for order-order transitions. Its experimental realization, however, has remained elusive. Using high-pressure ^{11}B nuclear magnetic resonance measurements on the quantum magnet $SrCu_2(BO_3)_2$, we here demonstrate a magnetic field–induced plaquette singlet to antiferromagnetic transition above 1.8 gigapascals at a notably low temperature, $T_c \simeq 0.07$ kelvin. First-order signatures of the transition weaken with increasing pressure, and we observe quantum critical scaling at the highest pressure, 2.4 gigapascals. Supported by model calculations, we suggest that these observations can be explained by a proximate DQCP inducing critical quantum fluctuations and emergent O(3) symmetry of the order parameters. Our findings offer a concrete experimental platform for investigation of the DQCP.

The theoretically proposed deconfined quantum critical point (DQCP) (*1*) connects two different ordered ground states of quantum matter by a continuous quantum phase transition (QPT). This type of criticality, which has been explored primarily in the context of two-dimensional (2D) quantum magnets (*2*), lies beyond the conventional paradigm of discontinuous (first-order) transitions between ordered phases with unrelated symmetries. The DQCP is associated with unconventional phenomena, including fractional spinon excitations and deconfined gauge fluctuations (*3*–*5*). Further investigations have extended the concept, introducing emergent symmetries (*6*–*11*) and exotic first-order transitions (*12*, *13*). In a very recent scenario, the DQCP is proposed to be a multicritical point (*14*, *15*) connected to a gapless quantum spin liquid (QSL) (*16*–*20*).

Although DQCP phenomena are broadly relevant in quantum materials (*21*), there has been no supportive experimental identification in any system. Quantum magnets in which the interactions can be varied over a wide enough range to realize two phases bordering a DQCP are rare. An exception is the layered material $SrCu_2(BO_3)_2$ (*22*–*24*), in which antiferromagnetic (AFM) Heisenberg interactions between the $S = 1/2$ Cu^{2+} spins (Fig. 1A) provide a faithful realization of the 2D Shastry-Sutherland model (SSM) (*25*). In the SSM, three different $T = 0$ phases are well established to form as a function of the ratio $g = J/J'$ of the inter- to intradimer couplings (*26*, *27*): an exact dimer-singlet phase (DS, with singlets on the J' bonds), a twofold degenerate plaquette-singlet (PS) phase (Fig. 1B), and a Néel AFM phase (Fig. 1C). At ambient pressure, $SrCu_2(BO_3)_2$ is well described by the $g \simeq 0.63$ SSM with a DS ground state (*24*). An applied pressure increases g, driving the system into a PS phase at $P \simeq 1.8$ GPa (*28*, *29*), which persists with transition temperature $T_P \simeq 2$ K up to $P \simeq 2.6$ GPa (*30*, *31*). An AFM phase with Néel temperature T_N from 2.5 to 4 K has been detected between 3.2 and 4 GPa (*30*).

Here, we report a ^{11}B nuclear magnetic resonance (NMR) study of $SrCu_2(BO_3)_2$ in a magnetic field H up to 15 T and at pressures up to 2.4 GPa, aiming to characterize the field-driven PS-AFM transition. At 2.1 GPa, PS and AFM transitions are resolved using their NMR signatures and merge at a common point (H_C, T_C), with $H_C \simeq 6$ T and $T_C \simeq 0.07$ K (Fig. 1D). Such a low T_C in relation to T_P and T_N farther away from H_C indicates proximity to a $T_C = 0$ QPT. First-order discontinuities at (H_C, T_C) weaken with increasing pressure, and we observed quantum-critical scaling of the spin-lattice relaxation at 2.4 GPa for $T > T_C$.

Our results support the existence of a multicritical DQCP controlling the quantum fluctuations at 2.4 GPa, with T_C on the associated first-order line suppressed by an emergent O(3) symmetry of the combined scalar PS and O(2) AFM order parameters (*7*, *8*). By synthesizing past and present experiments on $SrCu_2(BO_3)_2$ and model calculations, we arrived at the global phase diagram depicted in Fig. 2. Before further discussing the DQCP scenario, we present our NMR detection of the various phases and transitions.

NMR identification of phases

We performed ^{11}B NMR measurements on $SrCu_2(BO_3)_2$ single crystals at pressures of up to 2.4 GPa in fields between 0.2 and 15 T and temperatures down to 0.07 K. Experimental details are provided in the supplementary materials (*33*). We first discuss NMR line shifts to detect the relevant quantum phases and transitions, followed by results of the spin-lattice relaxation rate $1/T_1$.

A typical ^{11}B NMR spectrum, shown in Fig. 3A, has a central peak with four satellite peaks on either side, from inequivalent sites B1 to B4 (Fig. 1A) caused by a small tilt angle between the field and the crystalline c axis (*33*). The satellites are sensitive to changes of the lattice structure because of the local coupling between the nuclear quadrupole moment and the electric field gradient (*33*). As shown at a low field and $P = 2.1$ GPa in Fig. 3B, the full-width at half maximum (FWHM) height of the satellites increases on cooling <10 K until a maximum at $T \simeq 3$ K, reflecting increasing lattice fluctuations when the spins form fluctuating plaquette singlets above the ordered PS phase (*31*). This PS liquid crosses over to the trivial paramagnetic (PM) state at higher temperature.

Below 1.8 K, the FWHM in Fig. 3B rises sharply and saturates around 1 K. As explained in section S2 of the supplementary materials (*33*), the rapid broadening follows from an orthogonal lattice distortion when a full-plaquette (FP) PS state (Fig. 1B) forms. The FWHM as a proxy for the PS order parameter is further corroborated by the consistency of $T_P \simeq 1.8$ K at the low field applied in Fig. 3B with the location of a sharp specific-heat peak (*30*, *31*), marked in Fig. 1D.

Figure 3C shows the evolution of the central peak with T at $P = 0.9$ GPa and $H = 4$ T. The negative Knight shift at the higher temperatures reflects the hyperfine coupling $A_{hf} \simeq -0.259\ T/\mu_B$ [see section S3 of the supplementary materials (*33*)] for $H \| c$ (*34*, *35*). The shift increases rapidly below $T^* \simeq 7$ K when dimer singlets form in the DS state. At 2.1 GPa (Fig. 3E), PS order forms < 2 K, but the Knight shift changes rapidly at $T \simeq 4$ K when the PS liquid forms.

The first-order transition between the DS phase and the PS or PS liquid phase terminates at an Ising-type critical point, which at $H = 0$ is located at $P \simeq 1.9$ GPa, $T \simeq 3.3$ K (*31*). At low T, the DS-PS transition takes place between 1.7 and 1.8 GPa (*30*). The first-order DS line must therefore bend slightly, as indicated in Fig. 2A, and can be crossed versus T at fixed P and H. Indeed, at 1.85 GPa (Fig. 3D),

[1]Department of Physics and Beijing Key Laboratory of Optoelectronic Functional Materials and Micro-nano Devices, Renmin University of China, Beijing 100872, China. [2]Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. [3]School of Physics, Beijing Institute of Technology, Beijing 100081, China. [4]Department of Physics, Boston University, Boston, MA 02215, USA. [5]School of Physical Sciences, Graduate University of the Chinese Academy of Sciences, Beijing 100190, China. [6]Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China. [7]Key Laboratory of Quantum State Construction and Manipulation (Ministry of Education), Renmin University of China, Beijing 100872, China.
*Corresponding author. Email: rong.yu@ruc.edu.cn (R.Y.); sandvik@bu.edu (A.W.S.); wqyu_phy@ruc.edu.cn (W.Y.)
†These authors contributed equally to this work.

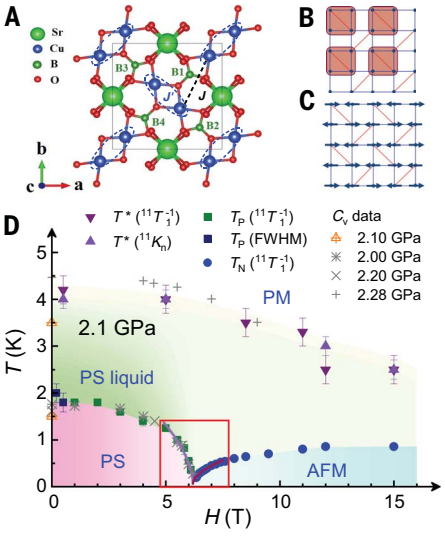


Fig. 1. Experimental overview. (**A**) Atomic structure of a $SrCu_2(BO_3)_2$ plane. Pairs of Cu^{2+} ions form spin dimers (ellipses) with Heisenberg intradimer (J') and interdimer (J) interactions (black dashed line). Each unit cell contains four B ions, for which we investigated the NMR response. (**B**) The PS phase in the equivalent square lattice of J (blue) and J' bonds (red). In $SrCu_2(BO_3)_2$, the singlets (shading) form on the full (J') plaquettes in one of two symmetry-equivalent patterns, whereas in the SSM, the singlets form on half of the empty plaquettes. (**C**) The AFM phase, which breaks O(3) symmetry when $H = 0$ and O(2) symmetry when $H \neq 0$. For $SrCu_2(BO_3)_2$ in a c-axis field, we found that the moments ordered along the a or b axis. (**D**) Field-temperature phase diagram at 2.1 GPa, showing the PM, PS liquid, ordered PS, and AFM phases resolved by our NMR measurements (Figs. 3 to 5). The transition temperatures T_P and T_N and the crossover temperature T^* are compared with specific-heat measurements (*30*, *31*). The data for 2.1 GPa come from (*30*) and the rest from (*31*). The red box marks the regime analyzed in Fig. 5F.

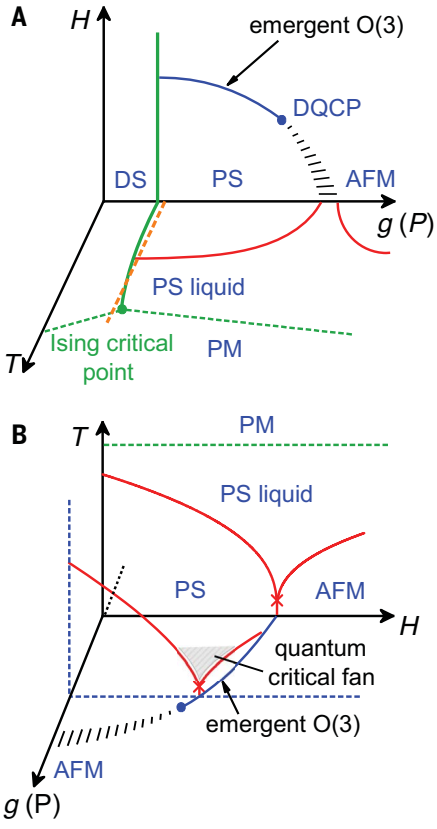


Fig. 2. Schematic phase diagram and DQCP scenario. (**A**) Phases in the space of coupling [$g = J/J'$ in the SSM, P in $SrCu_2(BO_3)_2$], temperature, and magnetic field. A multicritical DQCP separates a line of first-order QPTs and either a QSL phase (*17*) or a line of generic DQCPs (*38*); the region marked with dashed lines represents this undetermined feature. The first-order DS transition at fixed H (solid green line) terminates at an Ising critical point [green circle, based on previous experiments (*31*)]. The green dashed lines indicate crossovers at $T^*(g)$ into the PM phase. The dashed orange line shows how the slightly curved first-order DS transition line can be crossed versus T at fixed P (based on present experimental data). The ordered AFM phase at $T > 0$ requires interlayer couplings, as in $SrCu_2(BO_3)_2$. The magnetization plateau states at larger H (*28*, *32*) are not shown. (**B**) Phase diagram drawn to highlight (H,T) planes exemplified by Fig. 1D. Red crosses indicate $T_C > 0$ caused by weak 3D effects and violations of O(3) symmetry. The shading represents the "fan" in which quantum critical scaling is expected (supported by present experimental data). The blue dashed lines indicate the plane of highest-pressure (2.4 GPa) measurements.

the central peak is split at temperatures between 3 and 4 K, indicating phase coexistence. Previously, a different splitting was reported at 2.4 GPa (*35*, *36*), which was perhaps caused by pressure inhomogeneity, but we observed the double peak only at 1.85 and 1.95 GPa [see section S3 of the supplementary materials (*33*)]. Outside of this pressure range, T^* likely only marks a rapid crossover between the PM and PS liquid, with associated sharp specific heat peaks (*30*, *31*) which were observed also away from the critical point and reproduced (*31*) by SSM calculations. We have found no NMR signatures of a structural transition here or at higher temperatures [see section S3 of the supplementary materials (*33*)].

Above 1.95 GPa, AFM order emerges at high fields and leads to splitting of the NMR central peak by alternating positive and negative hyperfine fields, as shown at 2.1 and 2.4 GPa in Fig. 4, A and B, respectively, both at $T = 0.07$ K. The sudden rise with field of the peak-splitting $f_R - f_L$ (a proxy AFM order parameter), shown in Fig. 4C, signals a discontinuous onset of AFM order at $H_C(P)$, with the discontinuity much weaker at the higher pressure.

In the AFM state, the uniform magnetization does not exhibit any obvious discontinuity at H_C and remains < 2% of the saturated moment at our highest field of 15 T (fig. S5). A crossover temperature T^* persists also at high fields, where the PS liquid develops increasing spin fluctuations (discussed further below).

Spin-lattice relaxation rate

$1/T_1$ is a direct probe of low-energy spin fluctuations and can detect the PS and AFM transitions more precisely than the line shifts; the two probes give consistent results. Figure 5, A and B, show $1/T_1$ at $P = 2.1$ GPa for a wide range of applied fields that we grouped into those below and >6.2 T, corresponding, respectively, to the low-T PS and AFM phases; Fig. 5, C and D, show the same at 2.4 GPa with the separation at 5.8 T.

At 2.1 GPa (Fig. 5B), we found a sharp drop of $1/T_1$ at $T^* \simeq 3$ to 4 K and a broad peak or sharper kink <2 K. At low fields in Fig. 5A, the latter feature appeared up to 6.1 T and clearly marked the opening of a spin gap below T_P. At $P = 2.4$ GPa, we did not find a peak at T_P (Fig. 5C), but rather a sharp crossover from a low-T gapped regime to a window with power-law behavior that is analyzed in Fig. 5E and will be further discussed below. At the higher fields in Fig. 5, B and D, the low-T features (<0.8 K) are much sharper and coincide with the NMR peak splitting in Fig. 4, A and B. Thus, we can safely identify these peaks for $H \geq 6.33$ T as T_N (*37*). The minimum in $1/T_1$ around 1.5 K in Fig. 5B increases with the field, indicating increasing spin fluctuations in the PS liquid state.

Figure 5F shows very clear field-induced PS-AFM transitions revealed by these signals at both $P = 2.1$ and 2.4 GPa. The PS and AFM boundaries, $T_P(H)$ and $T_N(H)$, respectively, meet at a very low T_C. Given phase coexistence (Fig. 4, A and B), the QPT at H_C is clearly first order. The proxy AFM order parameter $f_R - f_L$ in Fig. 4C is consistent with H_C determined from $1/T_1$ at both pressures. The much smaller first-order discontinuity of $f_R - f_L$ at the higher pressure indicates the approach toward a continuous QPT.

We extracted the PS spin gap Δ by fitting $1/T_1$ below T_P to a semi-empirical form $T^{-a}e^{-\Delta/k_BT}$ with $a \approx 1$ [see section S5 of the supplementary materials (*33*)]. As expected, a linear decrease with H of Δ at both pressures is revealed in Fig. 6A and is caused by the field lowering of the $S = 1$ ($S^z = 1$) state above the singlet PS ground state. These results are compatible with previously determined $H = 0$ gap estimates (*29*, *30*) and the known g factor.

At a first-order transition into the AFM phase, the PS gap should jump discontinuously

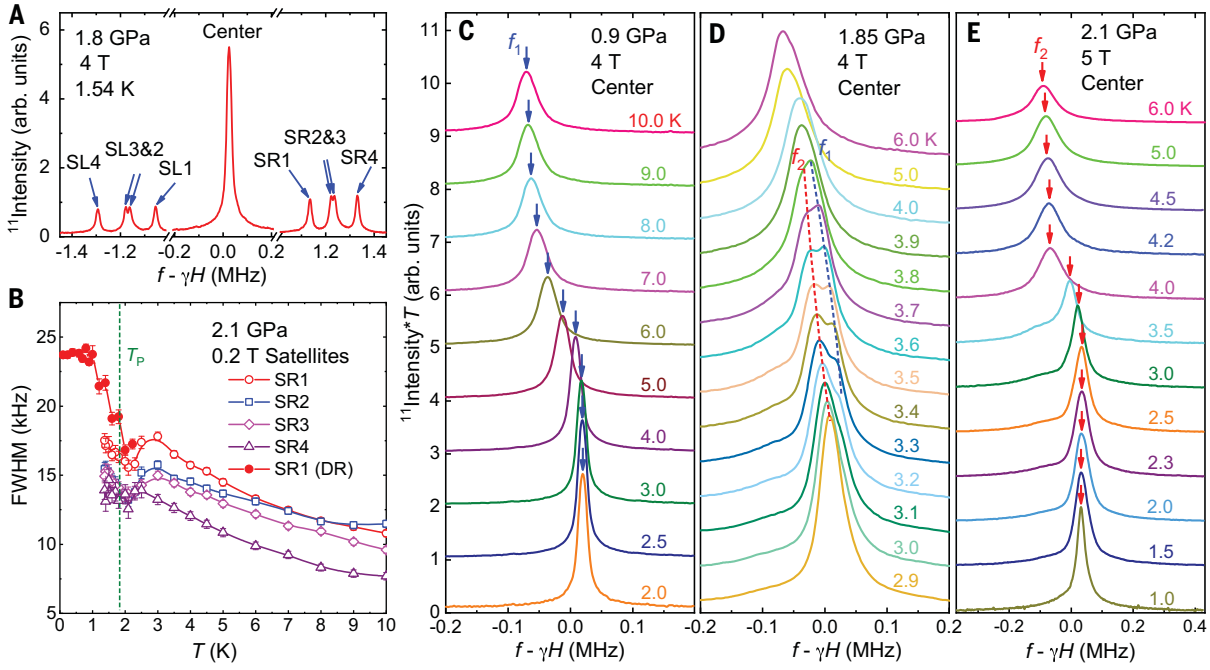


Fig. 3. NMR spectra and line shifts. (**A**) NMR ^{11}B spectrum at H = 4 T and P = 1.8 GPa, with the field applied at 8.6° from the crystalline c axis, showing the center line and two sets of satellites associated with the four B sites (Fig. 1A). (**B**) FWHM of satellites SR1 to SR4 shown as a function of temperature at P = 2.1 GPa and H = 0.2 T. The line at 1.8 K marks the onset of an upturn with further cooling. SR1 was measured in the dilution refrigerator in addition to the regular helium cryostat used for all cases [see section S2 of the supplementary materials (*33*)]. (**C** to **E**) NMR center line for a range of temperatures (curves shifted vertically for clarity) at P = 0.9 GPa and H = 4 T (C), 1.85 GPa and 4 T (D), and 2.1 GPa and 5 T (E). The peaks in the DS phase (C) and in and above the PS phase (E) are marked f_1 and f_2, respectively. The split peak in (D) reflects phase coexistence.

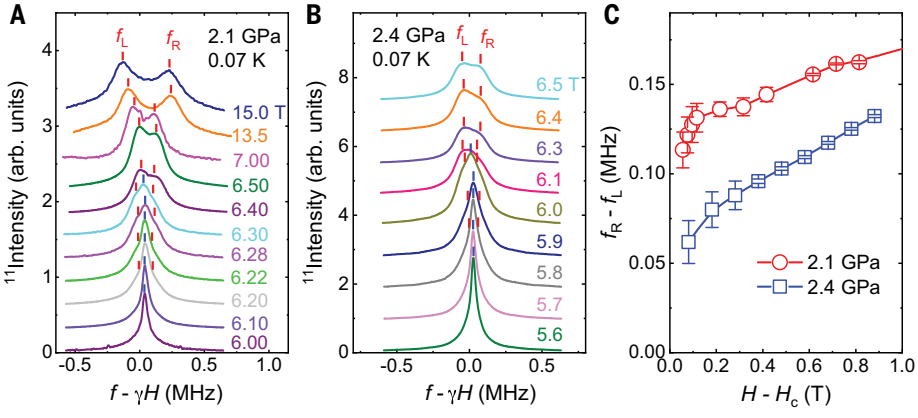


Fig. 4. AFM transition. Splitting of the NMR center line with increasing H at T = 0.07 K is shown in (**A**) and (**B**) for P = 2.1 and 2.4 GPa, respectively. The two peaks marked f_L and f_R (red bars) indicate AFM order developing above $H \simeq 6$ T. A center peak (blue bars) remaining at fields up to 6.1 T indicates phase coexistence. (**C**) Proxy AFM order parameter $f_R - f_L$ versus $H - H_C$, where H_C is determined using the spin-lattice relaxation rate (Fig. 5).

to zero at H_C (given that the AFM state is gapless), but despite the clear first-order signals described above (Fig. 4C), we found that $\Delta(H_C)$ values were indistinguishable from zero within statistical errors. We will discuss the anomalously small gap discontinuity in the context of the proximate DQCP scenario further below.

Deconfined quantum criticality

The SSM at H = 0 has been a candidate for a DQCP separating its coupling-induced PS and AFM ground states (*7*, *38*). The singlets in the PS phase of the model occupy the empty plaquettes, in contrast to the FP state in $SrCu_2(BO_3)_2$ (Fig. 1B). This aspect of the PS state depends sensitively on other possible weak interactions beyond the SSM (*11*, *39*), and the SSM description of the global phase diagram of $SrCu_2(BO_3)_2$ should remain valid.

There is mounting theoretical evidence that a gapless QSL phase can exist between a PS state (or closely related spontaneously dimerized state) and the AFM state in frustrated 2D quantum spin systems (*16*, *20*, *40*–*42*) and that these QSL phases generically end at multicritical DQCPs (*15*, *17*, *18*). Beyond such a point, the transition without intervening QSL is expected to be first order, with the coexistence state at H = 0 inheriting (and breaking) the emergent O(4) or SO(5) symmetry, depending on the type of singlet-ordered phase (*7*, *8*, *10*, *12*, *13*) of the DQCP.

In the H = 0 SSM, early calculations indicated a first-order PS-AFM transition (*27*), and a recent calculation suggested an O(4) [from the O(3) AFM and scalar PS order parameters] multicritical DQCP in an extended parameter space (*11*). A generic O(4) DQCP had previously been proposed (*38*). The intervening gapless QSL between the PS and AFM phases was identified very recently (*17*, *19*) and may be explained by an instability of the conventional DQCP (*15*). These theoretical insights, along with our NMR results for $SrCu_2(BO_3)_2$, suggest the scenario in Fig. 2. Because no experiment so far (including ours) has explicitly confirmed a QSL phase, the possibility remains that there is instead

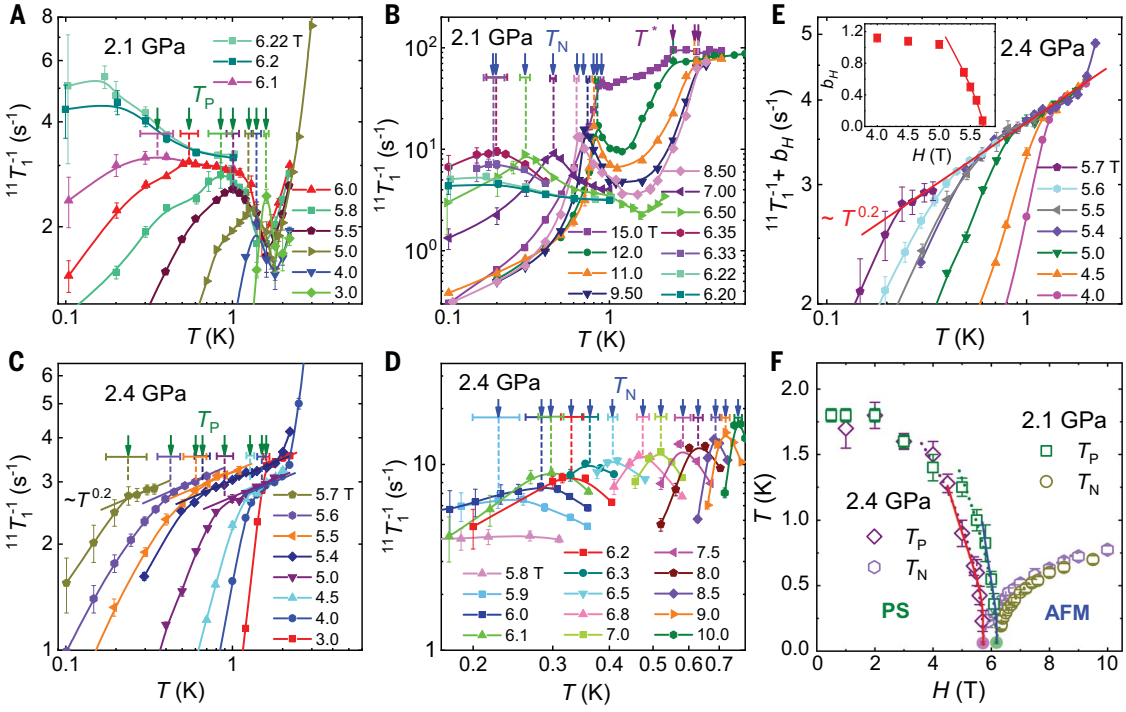


Fig. 5. Spin-lattice relaxation. $1/^{11}T_1(T)$ was measured at 2.1 GPa [(**A**) and (**B**)] and 2.4 GPa [(**C**) and (**D**)], which are separated to show the PS [(A) and (C)] and AFM [(B) and (D)] states. The drop in $1/T_1$ at a $T^* \simeq 4$ K [(A) and (B)] indicates the sharp crossover into the PS liquid. The peaks at lower T mark T_P and T_N, with uncertainties indicated by the horizontal bars. At 2.4 GPa, no low-T PS peak is observed (C), but T_P can be extracted from the sudden change from thermally activated to quantum critical behavior, $1/T_1 = aT^\eta - b_H$. (**E**) Inset: Power-law scaling of the offset, $b_H \propto (H_C - H)^d$ with $d \approx 0.8$, close to H_C. Main panel: The common scaling form with constant a and $\eta \approx 0.2$ is demonstrated, and b_H has been added. (**F**) Low-temperature phase diagrams at 2.1 and 2.4 GPa. The solid and dotted lines indicate the phase boundaries modeled by, respectively, a logarithmic form of T_P and near-critical forms of both T_P and T_N [see section S6 of the supplementary materials (*33*)]. The latter fits give the T_C values indicated with circles.

another line of PS–AFM transitions. Although the dashed regions in the phase diagrams in Fig. 2 can represent either possibility, specific heat measurements at $H = 0$ (*30*, *31*) found no phase transition between 2.6 and 3.2 GPa, consistent with a QSL ground state evolving into the $T > 0$ PS liquid.

A putative multicritical DQCP at $H > 0$ should evolve from a corresponding $H = 0$ DQCP with emergent O(4) symmetry (*7*, *38*). Although this O(4) point exists only in an extended parameter space outside of the (g, H, T) cube in Fig. 2, the fact that the field-induced magnetization is very small at H_C [see section S4 of the supplementary materials (*33*)] suggests that the putative $H > 0$ DQCP still hosts an approximate O(4) symmetry, with stronger O(3) character developing on the first-order line. Strictly speaking, at $H > 0$, the DQCP may evolve into a near-critical triple point with first-order signatures at the lowest energy scales.

Closer proximity of $SrCu_2(BO_3)_2$ to some continuous QPT with increasing pressure is certainly supported by our observation of a weaker discontinuity of the AFM order parameter at 2.4 GPa than at 2.1 GPa (Fig. 4C). Moreover, at a clearly first-order transition, correspondingly high T_C values would normally be expected. The low T_C at both pressures then point to a mechanism suppressing long-range order rather far away from the QPT. The DQCP scenario offers this possibility through its emergent continuous symmetry inherited (at least up to some large length scale) by the first-order line. An ideal 2D coexistence state with continuous order parameter symmetry must have $T_C = 0$, but weak violations of the symmetry [in combination with 3D effects (*43*)] would imply a low $T_C > 0$, as observed in $SrCu_2(BO_3)_2$.

In the scenario of a first-order transition with emergent O(3) symmetry, the Ising-type PS order can be understood as an uniaxial deformation of the O(3) order parameter. A logarithmic form of the PS transition temperature is then expected: $T_P \propto \ln^{-1}[a(H_C - H)]$ for some value of a (*7*, *44*). Fits of the experimental data to this form [see section S6 of the supplementary materials (*33*)] are shown with solid curves in Fig. 5F and indeed describe the behavior close to H_C.

To describe $T_N(H)$, we note again that interlayer interactions are required for $T_N > 0$ in a spin-isotropic system. These 3D couplings also change a continuous QPT ($T_C = 0$) into a first-order line extending to a bicritical or triple point at $T_C > 0$ (*38*, *43*) (red crosses in Fig. 2B). Given the extremely low T_C values in $SrCu_2(BO_3)_2$, a modified critical form with the same exponent ϕ governing both transitions above T_C may be expected from DQCP dualities (*13*, *45*): $T_{P,N} = T_C + a_{P,N}|H - H_C|^\phi$. Fits with independent exponents ϕ for the PS and AFM transitions [see section S6 of the supplementary materials (*33*)] indeed support a common value and motivate joint fitting with a single ϕ. Such fits are shown by the dashed curves in Fig. 5, where T_C is in the range of 0.05 to 0.07 K at both pressures. At 2.1 GPa, $H_C = 6.183 \pm 0.007$ and $\phi = 0.57 \pm 0.02$, and at 2.4 GPa, $H_C = 5.719 \pm 0.007$ and $\phi = 0.50 \pm 0.04$. These fits in which ϕ is close to estimates for both SO(5) (*12*, *14*) and O(4) (*45*) DQCPs [see section S6A of the supplementary materials (*33*)] do not rule out the alternative logarithmic form of T_P but do further validate the very low T_C values and common transition field H_C for both order parameters.

Quantum-critical relaxation

As shown in Fig. 5C, $1/T_1$ at 2.4 GPa exhibits T^η scaling with $\eta \approx 0.2$ within a window of temperatures for several fields close to H_C on the PS side. The ensemble of fits is further analyzed in Fig. 5E using the expected quantum-critical form $1/T_1 = aT^\eta - b_H$ (*46*), where a is a constant and $b_H > 0$ for $H < H_C$. The fact that scaling behavior is not observed at 2.1 GPa (Fig. 5A) suggests that only the system at 2.4 GPa is sufficiently close to a continuous QPT that it realizes the quantum critical fan (*46*). This is depicted in Fig. 2B, where T is the largest energy scale (but low enough so that the correlation length is well above the lattice constant). The value of η is compatible with an estimate for an O(4) DQCP (*45*) and slightly below the SO(5) value (*2*, *12*).

On the AFM side (Fig. 5D), $1/T_1$ is dominated by the 3D effects causing $T > 0$ AFM order, with

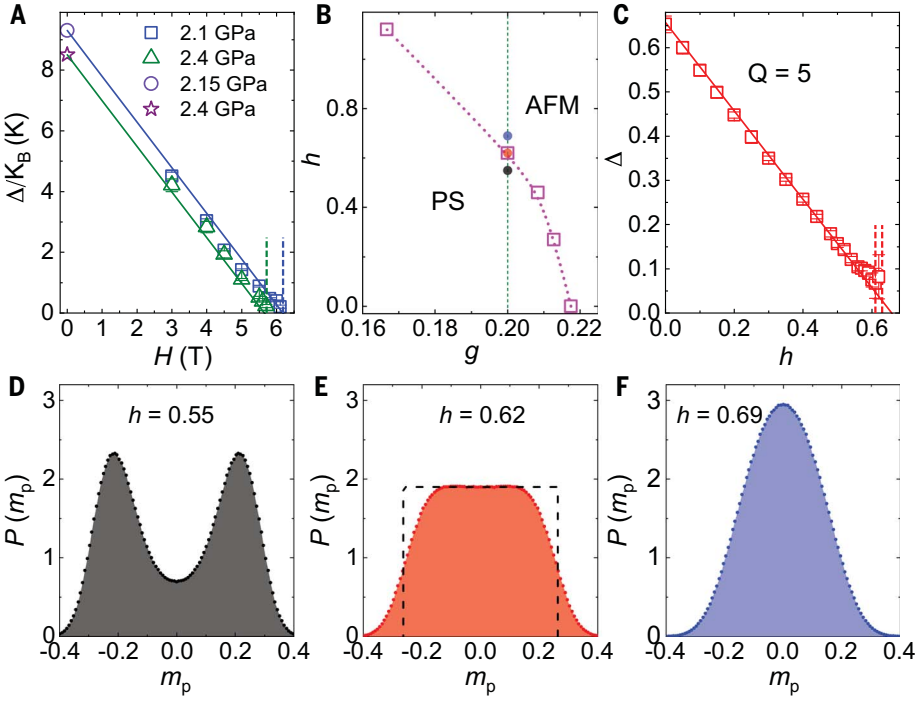


Fig. 6. Spin gap and emergent symmetry. (**A**) Field-dependent gap of $SrCu_2(BO_3)_2$ from experimental data. The lines show the expected form $\Delta(H) = \Delta(0) - \tilde{g}\mu_B H$, where $\Delta(0)$ represents the reported zero-field gaps with pressure at 2.15 GPa (*29*) and 2.4 GPa (*30*), and $\tilde{g} = 2.28$ is the known g factor [see section S5 of the supplementary materials (*33*)]. The vertical dashed lines represent H_C values from Fig. 5F. (**B**) Calculated ground-state phase diagram of the CBJQM versus $g = J/Q$ and field h. The PS and AFM phases are separated by a first-order transition. The vertical line and closely spaced points mark the parameters in (D) to (F). (**C**) Calculated spin gap of the CBJQM at $g = 0.2$. The dashed vertical line indicates h_C, and the solid line is a fit to $\Delta(h) = \Delta(0) - h$. (**D** to **F**) Calculated distribution of the plaquette order parameter. Double-peak (D), plateau (E), and single-peak (F) distributions are found, respectively, in the PS phase ($h = 0.55$), at the transition ($h = 0.62$), and in the AFM phase ($h = 0.69$).

the associated peak in $1/T_1$ masking any 2D quantum criticality, unlike the PS side, where the spin correlations and 3D effects are much weaker. We lack 2.4 GPa data at temperatures higher than those shown in Fig. 5C. At 2.1 GPa, no scaling is observed between T_N and T^* in Fig. 5B, where a sharp drop below T^* is immediately followed by strong precursors to AFM ordering.

Quantum spin model

We now turn to the checkerboard $J - Q$ model (CBJQM), in which four-spin interactions Q replace J' in the SSM. The CBJQM is amenable to quantum Monte Carlo simulations and hosts PS and AFM phases separated by a first-order transition with emergent O(4) symmetry at zero field (*7*). We here simulate [see section S1 of the supplementary materials (*33*)] the same model in a field, defining $g = J/Q$ and $h = H/J$ with $J = 1$.

In the phase diagram in Fig. 6B, the field-driven PS-AFM transition is first order. The PS gap $\Delta(h)$ obtained from the low-temperature susceptibility (fig. S22) is shown in Fig. 6C at $g = 0.2$, below the $h = 0$ transition at $g_C(0) \approx 0.217$. The expected linear form $\Delta(h) = \Delta(0) - h$ for an $S^z = 1$ excitation is observed for $h < h_C$, with h_C slightly less than $\Delta(0)$, implying a small gap discontinuity at h_C. We also observed a very small magnetization jump, ~0.002 per spin. These behaviors are reminiscent of the well-known "spin-flop" transitions from Ising to canted XY AFM phases, but with anomalously small magnetization discontinuity. In section S8 of the supplementary materials (*33*), we posit that the small magnetization and gap discontinuities, which decrease further upon moving closer to $g_C(0)$, reflect an approximate emergent O(3) symmetry in the CBJQM at $h > 0$.

The emergent symmetry can also be studied directly. At $h = 0$, the O(3) AFM order parameter (m_x,m_y,m_z) combines with the scalar PS order parameter m_p into an O(4) vector (m_x,m_y,m_z,m_p) at the $T = 0$ transition (*7*, *43*). To detect the putative O(3) symmetry of (m_x, m_y,m_p) at $h > 0$, we studied the distribution $P(m_p)$ along the vertical line in Fig. 6B. In the PS phase (Fig. 6D), $P(m_p)$ exhibits the expected double peak, reflecting the Z_2 symmetry that is broken in the thermodynamic limit. In the AFM phase (Fig. 6F), there is a single central peak, reflecting the lack of PS order.

At a conventional first-order transition, a three-peak distribution would follow from coexisting PS and AFM orders. By contrast, the distribution in the coexistence state in Fig. 6E is nearly uniform over a range of m_p values (with finite-size rounded edges). The distribution $P(m_p)$ obtained by integrating an O(3) symmetric $P(m_p,m_x,m_y)$ over m_x and m_y should indeed be uniform for $m_p \in [-R,R]$, where $R \equiv \max(|m_p|)$; therefore, the approximately flat distribution demonstrates emergent O(3) symmetry in the presence of finite-size fluctuations of R. Although this symmetry cannot be exact, i.e., it exists up to some finite length scale, it is responsible for suppressing T_C and the gap at H_C; see fig. S19, where we also show supporting results for cross-correlations between the PS and AFM order parameters.

We expect the same O(3) emergent symmetry at the PS-AFM transition in $SrCu_2(BO_3)_2$, in which the ordered coexistence state breaks the symmetry. The symmetry should be violated on long length scales because of the distance to the DQCP and also by 3D couplings. One of the Goldstone modes associated with the coexistence state then develops a small gap. Studies of the CBJQM with interlayer couplings suggest that the symmetry is unexpectedly robust (*43*).

Emergent O(3) symmetry on large length scales in $SrCu_2(BO_3)_2$ is supported, in particular, by our results at 2.1 GPa, where Fig. 4C shows a large discontinuity in the AFM order parameter, but T_C is low and the gap (Fig. 6A) is very small at H_C. Moreover, the uniform magnetization is extremely small and does not exhibit a discernible discontinuity (fig. S5). These behaviors are analogous to those in the CBJQM for g close to $g_C(0)$.

Discussion

Our high-pressure NMR experiments on $SrCu_2(BO_3)_2$ in a magnetic field establish an example of a quantum magnet realizing DQCP phenomenology, which thus far had existed only in the realm of field theories and model studies. We have demonstrated PS and AFM transitions, with $T_P(H)$ and $T_N(H)$ merging at $T_C \simeq 0.07$ K and $H_C \simeq 6$ T. The PS-AFM transition at H_C is first order, with discontinuity weakening with increasing pressure.

We have argued that the suppression of T_C and absence of statistically significant PS gap discontinuity are consequences of emergent O(3) symmetry generated by a nearby DQCP. At the highest pressure, 2.4 GPa, $1/T_1$ exhibits critical scaling for T between 0.2 and 2 K, indicating sufficient proximity to the DQCP [which is likely of the multicritical type (*14*, *15*, *17*, *18*)] for realizing the characteristic quantum-critical fan (*46*) on the gapped PS side

of the transition. Strong 3D AFM ordering effects on the gapless side of the transition mask putative quantum criticality in $1/T_1$ there, but the AFM ordering temperature T_N vanishes in a way very similar to the PS ordering temperature T_P, again in support of emergent symmetry of the order parameters.

The $H = 0$ AFM phase was previously detected in the specific heat between 3.2 and 4 GPa (*30*), with T_N from 2 to 3.5 K. Subsequently, results at $H > 0$ were also reported (*31*). However, whereas $T_P(H)$ from the specific heat agrees well with our PS transitions shown in Fig. 1D, the heat capacity peak assumed to signal the AFM transition did not drop below 1 K (*31*), extending above the PS phase at fields as low as 3 T. It may be difficult to detect the small specific-heat peak signaling the AFM transition (*30*) in high-field measurements at low temperatures.

Beyond the highest pressure reached here, a plausible scenario (*15*, *17*) is a QSL between the PS and AFM phases (Fig. 2). Our experiments do not directly address the putative QSL, and further investigations should elucidate the low-T, $H = 0$ state between 2.6 and 3 GPa [where no order has been detected (*30*, *31*)] and its evolution as H approaches 5.7 T, where our current experiments point to a DQCP slightly above 2.4 GPa.

REFERENCES AND NOTES

1. T. Senthil, A. Vishwanath, L. Balents, S. Sachdev, M. P. A. Fisher, *Science* **303**, 1490–1494 (2004).
2. A. W. Sandvik, *Phys. Rev. Lett.* **98**, 227202 (2007).
3. T. Senthil, L. Balents, S. Sachdev, A. Vishwanath, M. P. A. Fisher, *Phys. Rev. B* **70**, 144407 (2004).
4. H. Shao, W. Guo, A. W. Sandvik, *Science* **352**, 213–216 (2016).
5. N. Ma *et al*., *Phys. Rev. B* **98**, 174421 (2018).
6. A. Nahum, P. Serna, J. T. Chalker, M. Ortuño, A. M. Somoza, *Phys. Rev. Lett.* **115**, 267203 (2015).
7. B. Zhao, P. Weinberg, A. W. Sandvik, *Nat. Phys.* **15**, 678–682 (2019).
8. P. Serna, A. Nahum, *Phys. Rev. B* **99**, 195110 (2019).
9. G. J. Sreejith, S. Powell, A. Nahum, *Phys. Rev. Lett.* **122**, 080601 (2019).
10. J. Takahashi, A. W. Sandvik, *Phys. Rev. Res.* **2**, 033459 (2020).
11. N. Xi, H. Chen, Z. Y. Xie, R. Yu, arXiv:2111.07368 [cond-mat.str-el] (2021).
12. A. Nahum, J. T. Chalker, P. Serna, M. Ortuño, A. M. Somoza, *Phys. Rev. X* **5**, 041048 (2015).
13. C. Wang, A. Nahum, M. A. Metlitski, C. Xu, T. Senthil, *Phys. Rev. X* **7**, 031051 (2017).
14. B. Zhao, J. Takahashi, A. W. Sandvik, *Phys. Rev. Lett.* **125**, 257204 (2020).
15. D.-C. Lu, C. Xu, Y.-Z. You, *Phys. Rev. B* **104**, 205142 (2021).
16. W.-Y. Liu *et al*., *Sci. Bull.* **67**, 1034–1041 (2022).
17. J. Yang, A. W. Sandvik, L. Wang, *Phys. Rev. B* **105**, L060409 (2022).
18. W.-Y. Liu *et al*., *Phys. Rev. X* **12**, 031039 (2022).
19. A. Keleş, E. Zhao, *Phys. Rev. B* **105**, L041115 (2022).
20. H. Shackleton, A. Thomson, S. Sachdev, *Phys. Rev. B* **104**, 045110 (2021).
21. Y.-H. Zhang, S. Sachdev, *Phys. Rev. B* **102**, 155124 (2020).
22. H. Kageyama *et al*., *Phys. Rev. Lett.* **82**, 3168–3171 (1999).
23. S. Miyahara, K. Ueda, *Phys. Rev. Lett.* **82**, 3701–3704 (1999).
24. S. Miyahara, K. Ueda, *J. Phys.* **15**, R327–R366 (2003).
25. B. S. Shastry, B. Sutherland, *Physica B+C* **108**, 1069–1070 (1981).
26. A. Koga, N. Kawakami, *Phys. Rev. Lett.* **84**, 4461–4464 (2000).
27. P. Corboz, F. Mila, *Phys. Rev. B* **87**, 115144 (2013).
28. S. Haravifard *et al*., *Nat. Commun.* **7**, 11956 (2016).
29. M. E. Zayed *et al*., *Nat. Phys.* **13**, 962–966 (2017).
30. J. Guo *et al*., *Phys. Rev. Lett.* **124**, 206602 (2020).
31. J. L. Jiménez *et al*., *Nature* **592**, 370–375 (2021).
32. Y. H. Matsuda *et al*., *Phys. Rev. Lett.* **111**, 137204 (2013).
33. Materials and methods, additional NMR data, data analysis, and supporting theoretical results are available as supplementary materials.
34. K. Kodama *et al*., *J. Phys.* **14**, L319 (2002).
35. T. Waki *et al*., *J. Phys. Soc. Jpn.* **76**, 073710 (2007).
36. M. Takigawa, T. Waki, M. Horvatić, C. Berthier, *J. Phys. Soc. Jpn.* **79**, 011005 (2010).
37. T. Moriya, *J. Phys. Soc. Jpn.* **18**, 516–520 (1963).
38. J. Y. Lee, Y.-Z. You, S. Sachdev, A. Vishwanath, *Phys. Rev. X* **9**, 041037 (2019).
39. C. Boos *et al*., *Phys. Rev. B* **100**, 140413 (2019).
40. S.-S. Gong, W. Zhu, D. N. Sheng, O. I. Motrunich, M. P. A. Fisher, *Phys. Rev. Lett.* **113**, 027201 (2014).
41. L. Wang, A. W. Sandvik, *Phys. Rev. Lett.* **121**, 107202 (2018).
42. Y. Nomura, M. Imada, *Phys. Rev. X* **11**, 031034 (2021).
43. G. Sun, N. Ma, B. Zhao, A. W. Sandvik, Z. Y. Meng, *Chin. Phys. B* **30**, 067505 (2021).
44. V. Y. Irkhin, A. A. Katanin, *Phys. Rev. B* **57**, 379–391 (1998).
45. Y. Q. Qin *et al*., *Phys. Rev. X* **7**, 031052 (2017).
46. A. V. Chubukov, S. Sachdev, J. Ye, *Phys. Rev. B* **49**, 11919–11961 (1994).
47. Y. Cui *et al*., Data and simulation codes for: Proximate deconfined quantum critical point in $SrCu_2(BO_3)_2$, Zenodo (2023); https://doi.org/10.5281/zenodo.7600028.

ACKNOWLEDGMENTS

We thank B. Normand for extensive suggestions and constructive criticism and W. Guo, F. Mila, M. Takigawa, Y. Wang, Z.-Y. Xie, and Y.-Z. You for helpful discussions. **Funding:** This work was supported by the Ministry of Science and Technology of China (grants 2022YFA1402700, 2022YFA1403402, and 2021YFA1400401), the National Natural Science Foundation of China (grants 12134020, 12104503, 12174441, and 51872328), the Simons Foundation (investigator grant 511064), the Strategic Priority Research Program(B) of the Chinese Academy of Sciences (grant XDB33010100), the K. C. Wong Education Foundation (grant GJTD2020-01), the Beijing Institute of Technology Research Fund Program for Young Scholars, the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (grants 22XNH096 and 21XNLG18). Some of the numerical calculations were performed on the Shared Computing Cluster managed by Boston University's Research Computing Services. Part of the measurements were conducted at the Cubic Anvil Cell Station, the Synergetic Extreme Condition User Facility (SECUF). **Author contributions:** Y.C. performed NMR measurements and data analysis with assistance from C.L., Z.H., and W.Y.. W.H. and S.L. provided single crystals. X.L. and H.L. performed the Bayesian fitting analysis. L.L., H.L., N.X., and K.H.W. performed numerical simulations with guidance from R.Y. and A.W.S.. W.Y., A.W.S., and R.Y. guided the project and wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data and simulation codes needed to evaluate the conclusions in this study have been deposited at Zenodo (*47*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adc9487
Materials and Methods
Figs. S1 to S22
Table S1
References (*48*–*65*)

Submitted 10 May 2022; accepted 12 May 2023
10.1126/science.adc9487

BASALT FLOWS

A younger and protracted emplacement of the Ontong Java Plateau

Peter C. Davidson[1]*, Anthony A. P. Koppers[1], Takashi Sano[2], Takeshi Hanyu[3]

The submarine volcanic emplacement of the Ontong Java Plateau (OJP) is the suggested cause of Ocean Anoxic Event 1a (OAE 1a). However, no precise timing and duration exists for the formation of OJP, and its connection to OAE1a relies mainly on proxies in the sedimentary record. We provide high-precision $^{40}Ar/^{39}Ar$ data from OJP drill and dredge sites that considerably improve OJP's eruptive history. The ages determined from this work are as much as 10 million years younger than previous dates and indicate a protracted formation over at least 6 million years. OJP now appears too young to have caused OAE1a, but we suggest that it may have had a role in the later OAE1b. The protracted eruptive sequence has implications for the emplacement dynamics of OJP and other large igneous provinces.

The Ontong Java Plateau (OJP) is a submarine large igneous province (LIP) located in the equatorial western Pacific Ocean (Fig. 1). OJP is a fragment of the Ontong Java Nui (OJN) superplateau that comprises OJP, Manihiki Plateau (MP), and Hikurangi Plateau (HP)—all three hypothesized to have been emplaced together in the Cretaceous and rifted apart through plate spreading shortly after (*1*, *2*). OJP is among the best-sampled oceanic plateaus, encompassing seven Deep Sea Drilling Project (DSDP) and Ocean Drilling Program (ODP) drill sites along with several kilometers of obducted volcanic stratigraphy in the Solomon Islands (*3*). Despite the extensive sampling of OJP, its emplacement history remains poorly constrained. Previous $^{40}Ar/^{39}Ar$ groundmass ages reported in the literature show a large range, from 128.2 million to 119.6 million years ago (Ma) with additional volcanism around 90 Ma and 2σ uncertainties ranging from >1.5 to 8.5 Ma (*4*–*8*). More recent attempts to date groundmass samples from ODP Leg 192 revealed a significant recoil issue resulting from the loss of ^{39}Ar (and ^{37}Ar) during the irradiation of samples, and no reliable ages came from these studies (*9*). The ambiguity and inaccuracy in knowing the emplacement timing of OJP limits our understanding of the potential environmental impact that this largest-ever submarine eruption may have caused in the world oceans (*10*) and in the mantle-plume processes underlying its formation (*11*, *12*).

Many have suggested that the emplacement of OJP caused Ocean Anoxic Event 1a (OAE 1a) at ~120 Ma (*13*–*18*), and a better constraining of the eruptive history and tempo of OJP has substantial implications for its responsibility in causing this event (*13*–*18*). It is proposed that massive, prolonged submarine eruptions can supply excess CO_2 and nutrients to the oceans, which increases biologic productivity and leads to widespread ocean anoxia in certain cases (*19*). OAE1a was a period of worldwide black-shale deposition in the ocean basins and a major perturbation to the carbon cycle (*20*) and thus to global climate. To date, the main indicators tying OJP to OAE1a are indirect evidence, such as various isotopic excursions and changes in trace element concentrations in the sedimentary records, all of which may suggest a volcanic source. However, a direct age correlation remains rather tenuous given the large uncertainties in dating the OJP lava flows (Fig. 2) (*13*–*18*). Other OAEs seem to correlate with submarine LIP emplacement, such as the Caribbean and Kerguelen plateaus' associations with OAE2 and OAE1a and -1b, respectively (*20*–*23*).

In addition, the cause of melting that could produce such a large magmatic event as the one that formed OJP is still debated (*1*, *24*–*28*). The currently favored hypothesis is that OJP is the result of a mantle-plume head impinging on the lithosphere, resulting in high-volume decompression partial melting (*6*, *28*, *29*). However, several lines of evidence do not support this, which include the lack of a plume tail producing an obvious linear track of seamounts as well as geophysical evidence of an unusually deep and mostly submarine emplacement with little subsequent lithospheric subsidence (*24*, *25*). These latter observations suggest a bolide impact (*24*) or dense fertile mantle entrained by rapid seafloor spreading (*25*) as

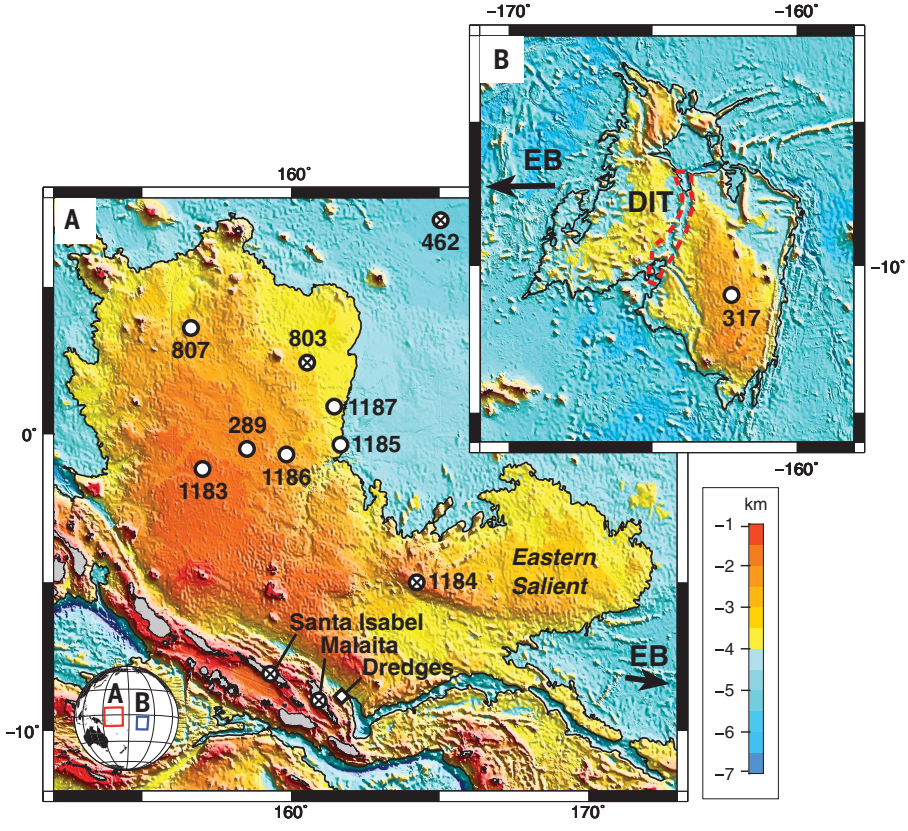


Fig. 1. Bathymetric maps of the OJP and MP. (**A**) OJP. (**B**) MP. DSDP and ODP drill sites with ages from this study are shown as solid white circles. The white diamond in (A) indicates the locations of the five dredge samples from this study (detailed map in fig. S1). Santa Isabel and Malaita (Solomon Islands) are terrestrial exposures of the Ontong Java basement with geochronological and geochemical data referenced in this study. DIT, Danger Islands Troughs (dotted red outline); EB, Ellice Basin.

[1]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA. [2]National Museum of Nature and Science, Tsukuba, Japan. [3]Japan Agency for Marine-Earth Science and Technology, Yokosuka, Japan.
*Corresponding author. Email: davidspe@oregonstate.edu

potential emplacement mechanisms. Evaluating these geodynamical hypotheses requires robust geochronological constraints.

This study updates the geochronology with high-precision $^{40}Ar/^{39}Ar$ incremental heating ages from plagioclase mineral separates from seven drill sites at OJP and MP and five dredge locations at OJP (Fig. 1). Although the drill sites reach only from <10 to ~216 m into the lavas at any individual site, their dispersed locations throughout OJP indicate that they may have sampled up to ~1500 m or >25% of the total volcanic stratigraphy (supplementary materials). Plagioclase analyses from 37 of 44 experiments yielded reliable age plateaus interpreted as eruption ages (table S1) while meeting current community quality standards (*30*). Individual ages range from 116.85 to 107.96 Ma, up to 10 million years (Myr) younger than previously published dates (Fig. 2). These preferred ages (Fig. 2 and table S3) include combined ages for each location but also include some individual analyses that are considered reliable and statistically different from other dates at that location (supplementary materials). These ages are all within the Cretaceous Normal Superchron (CNS), which is consistent with all OJP lavas being normally magnetized, and are generally consistent with biostratigraphy (supplementary materials).

Plagioclase analyses from the dredged samples agree with drilled samples but include a younger group of alkalic lavas not sampled at the drill locations. The three samples from DSDP Hole 317A located on MP are the oldest, with a combined age of 116.50 ± 0.30 Ma (Fig. 2). This age agrees with previously published ages from this site but reduces the reported 2σ standard error by an order of magnitude (*31*).

Our results show that the upper lava-flow units of the OJP are much younger than literature data suggest. Most previously published age plateaus contain just three to six incremental heating steps with uncertainties on individual steps in the ±10 Myr range. This prevents the resolution of certain problematic degassing patterns (Fig. 3A) when dating submarine basaltic groundmass, such as the recoil of ^{39}Ar. Improvements in mass spectrometry and sample preparation techniques allow for analyses with much higher resolution and with detailed age spectra that can resolve issues such as ^{39}Ar recoil. In a prior analytical campaign (supplementary materials), we dated 13 groundmass samples from OJP from Holes 807C, 1183A, and 1185B, including samples from a drill-hole interval that was previously dated (Fig. 3A). We found recoil in nearly all groundmass analyses (Fig. 3 and fig. S4) that manifests as a stepwise decrease in age between consecutive steps (Fig. 3, red steps) because recoil tends to affect low-temperature (low-T) sites (in the finest-grained phases) more than high-T sites within a sample. The consistent recoil patterns in the groundmass data render most or all previous groundmass ages from the literature apparently older and inaccurate. These results led us to reattempt analyses with plagioclase, which proved to have minimal recoil and produced flat age spectra (Fig. 3B).

Our plagioclase age results have important implications for the eruptive history of OJP. First, we show that the identified geochemical groups at OJP are now also distinguishable by age. The interpreted OJP main eruptive phase that produced the Kwaimbaita lavas (*26*) is indeed the oldest in the sequence and erupted across the entire plateau over a period of at least 3 Myr, starting at 115.51 Ma and ending around 111.42 Ma. This is much longer than the main eruptive phase of many terrestrial LIPs,

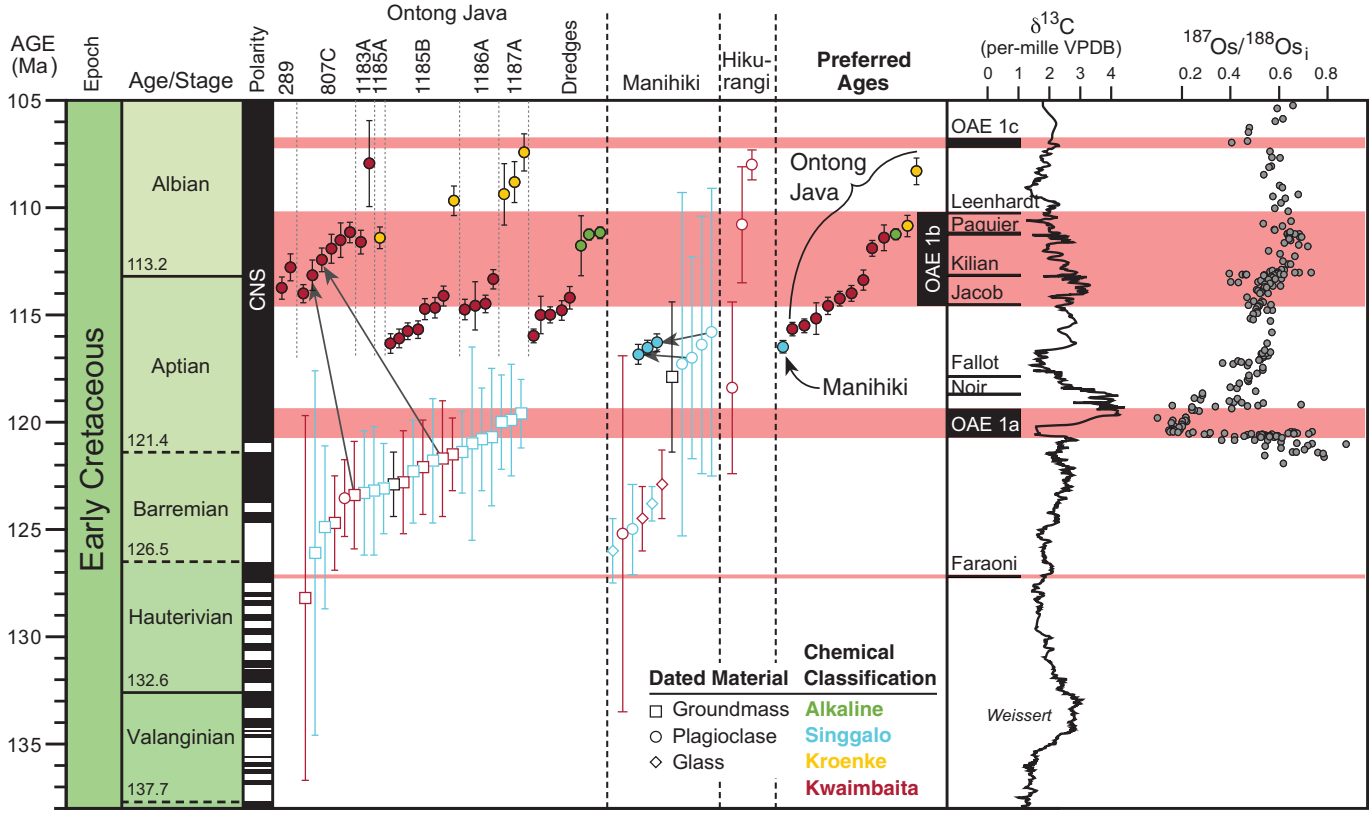


Fig. 2. Individual analyses from each location compared with literature data. Analyses are compared with data from OJP (*4*–*8*), MP (*29*, *31*), and HP (*31*). Preferred plagioclase ages from our analyses are shown in a separate column. The sample colors (solid, ages from this study; open, from literature) correspond to identified geochemical groups in the legend. The arrows connect our analyses with older analyses from similar drill-core intervals. Previous groundmass samples are older, whereas the MP plagioclase samples remain the same age. The Geologic Time Scale 2020 (GTS2020) and $\delta^{13}C$ curve (*38*) are shown for comparison, with labeled carbon perturbations and OAEs (also shown as pink horizontal bars). Compiled $^{187}Os/^{188}Os_i$ data (gray circles) from sedimentary land sections and drill cores are shown at far right (*23*).

such as the Deccan and Siberian Traps, which appear to have erupted in <1 Myr (*32*–*35*). The prolonged eruptive history of OJP compared with that of terrestrial LIPs could indicate a fundamental difference in the emplacement dynamics of LIPs that erupted in these two different environments. Second, our results support previous work suggesting that the Kroenke units are later-stage lavas erupting after the main-phase Kwaimbaita lavas (*35*), from 110.86 to 108.27 Ma, which is consistent with the stratigraphy in Hole 1185B despite the two groups having nearly identical mantle isotopic signatures (*26*). Third, the younger 111.25-Ma alkalic lavas from the dredge samples likely indicate a waning stage of volcanism erupting coincidently with the Kroenke units, but these lavas are similar in formation to the postshield alkalic volcanism seen at Hawai'i (*36*).

The geochemistry-age relations are more complex when considering the Singgalo samples from MP (referred to as the high-Ti group at MP), which are about 1 Myr older than the oldest-dated Kwaimbaita samples on OJP. On the basis of drilled sequences and field relations, the Singgalo type is recognized as a stratigraphically younger geochemical group at both OJP and MP (*4*, *29*), and the older age we obtained for the Singgalo lavas at MP relative to those at OJP implies that MP was emplaced before OJP. It is unclear how the Singgalo lavas at MP compare in age with Singgalo lavas at OJP because none were dated in this study, but they must be more than 5 Myr younger at OJP because they overlie the younger Kwaimbaita lavas there. This introduces a significant discrepancy in the emplacement history of the greater OJN system as outlined in the Taylor hypothesis (*1*). The older ages of MP samples relative to OJP samples—and the fact that these samples come from the stratigraphically younger Singgalo magmas—demonstrate that MP is almost certainly older than most of OJP. This is the first evidence for an age-progressive nature within OJN as a whole, indicating that the source of volcanism may have started centered below MP and then migrated westward across the Pacific over time. It is also unclear whether the Singgalo lavas are younger, older, or the same age as Kroenke lavas at OJP.

Our age results also have profound implications for OJP as the causal eruptive event that resulted in global OAE1a (*13*–*16*, *18*). Our results suggest that the main-stage Kwaimbaita lavas erupted about 3 Myr after OAE1a, with the oldest OJP lavas dated at ~116 Ma, whereas OAE1a ended at ~119 Ma (Fig. 2). However, sedimentary Os isotope records show strong negative excursions during OAE1a, suggesting a large volcanic contribution of a low mantle $^{187}Os/^{188}Os$ source to the oceans that is captured in these sedimentary records (*13*, *14*, *16*). Other sedimentary proxies such as Mn, Eu/Eu*, and Hg mimic the Os isotope record and also suggest a volcanic source for OAE1a (*15*, *18*). We posit that on the basis of the younger OJP ages we determined, we must seek another, older source for OAE1a.

Two revised scenarios for the emplacement of OJP relative to OAE1a should be considered. The first scenario entails a prolonged emplacement of OJP over several million years, of which the OJP drill sites characterize only the past 4 Myr of the main-stage Kwaimbaita eruptive history. The drill sites sample up to perhaps 1.5 km or 25% of the total volcanic stratigraphy (supplementary materials). Upward of 4 km of OJP's stratigraphy is exposed on the Solomon Islands (*37*), and lavas located deeper in the stratigraphy there could extend OJP's eruptive history back at least 3 to 5 Myr further in time. In this scenario, the bulk of the eruptive material could therefore have been erupted before the 115.5-Ma lava flows dated in this study, thus OJP could still be the cause of OAE1a. Rebounds in the Os isotope record between OAE1a and OAE1b suggest perhaps three volcanic pulses, one occurring between ~120.5 and 119.5 Ma, a second at 117 Ma, and the third between ~115 and 113 Ma (Fig. 2). This third pulse at the beginning of OAE1b is captured with our OJP age results, whereas the second pulse could be the result of volcanism at MP. In this scenario, main-stage OJP volcanism in this submarine LIP would have had a highly protracted nature, lasting ~7 to 9 Myr.

The second scenario assumes that our age results for the Kwaimbaita samples at the drill sites encompass most of the eruptive history of the main phase of OJP construction and that samples from the Solomons would yield similar ages and only extend the eruptive history of OJP back by less than 1 to 2 Myr. In this scenario, OJP would not be older than ~117 to 118 Ma and could not be the cause of OAE1a, leaving this global anoxic event without an identified volcanic source.

Although our findings suggest that OJP may not be the cause of OAE1a, they do point toward a connection with OAE1b. OAE1b occurred between about 114.5 and 110.5 Ma during the well-constrained Aptian-Albian boundary (*38*), and the ages we obtained show that Kwaimbaita volcanism was well underway by about 2 Myr before the onset of OAE1b and continued through the end of OAE1b. Several negative Os excursions are present during OAE1b and are

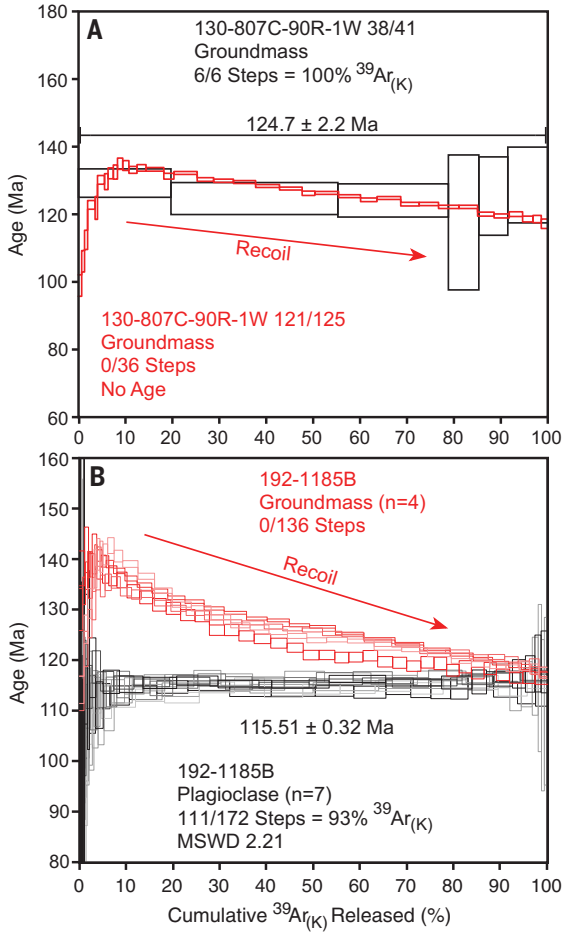


Fig. 3. Examples of the effects of recoil in Ontong Java samples. (**A**) Groundmass analysis (black) from sample 130-807C-90R-1W (interval 38 to 41 cm) published in 1993 (*6*), compared with the nearby sample (red) (interval 121 to 125 cm) analyzed in this study. The previous analysis has only six steps with large uncertainties that could not resolve the recoil apparent in our analysis. (**B**) Groundmass analyses (red) relative to plagioclase analyses (black and gray) from this study from ODP Hole 1185B. No groundmass analyses yielded age plateaus, whereas plagioclase analyses all have flat age spectra and are concordant.

suggested to be caused by volcanism. Although the Kerguelen Plateau has been invoked as the source (*22*, *23*) for OAE1b, we suggest that OJP could be a second contemporary source of these negative Os excursions.

These significantly younger ages for OJP and MP illuminate the need for modern, high-precision geochronology of difficult-to-date volcanic features. The previously used and too-old ~124 to 120 Ma emplacement age of OJP has been cited in numerous studies since radiometric ages were first acquired (*1*, *2*, *13*, *17*) and was incorrectly used to draw conclusions that went far beyond understanding how old this volcanic feature is. Despite the recognition of problematic ^{39}Ar recoil being present in OJP samples, these submarine volcanic plateau ages were the "best" available and are still used throughout the literature. The need for high-precision ages is especially important when trying to evaluate causal relationships between igneous and sedimentary records, as has been done with OAEs and other major biosphere perturbations (such as extinction events) and LIPs. Our data demonstrate the need to reevaluate other LIPs and their geochronological records to better understand their histories, the differences between submarine and terrestrial flood basalt systems, and how they relate to broader Earth-systems changes.

REFERENCES AND NOTES

1. B. Taylor, *Earth Planet. Sci. Lett.* **241**, 372–380 (2006).
2. K. Hochmuth, K. Gohl, G. Uenzelmann-Neben, *Geochem. Geophys. Geosyst.* **16**, 3789–3807 (2015).
3. M. G. Petterson *et al.*, *Tectonophysics* **283**, 1–33 (1997).
4. M. L. G. Tejada, J. J. Mahoney, C. R. Neal, R. A. Duncan, M. G. Petterson, *J. Petrol.* **43**, 449–484 (2002).
5. M. L. G. Tejada, J. J. Mahoney, R. A. Duncan, M. P. Hawkins, *J. Petrol.* **37**, 361–394 (1996).
6. J. J. Mahoney, M. Storey, R. A. Duncan, K. J. Spencer, M. Pringle, *Geophys. Monogr.* **77**, 233–261 (1993).
7. L. M. Chambers, M. S. Pringle, J. G. Fitton, *Spec. Publ. Geol. Soc. Lond.* **229**, 325–331 (2004).
8. I. J. Parkinson, B. F. Schaefer, ODP Leg 192 Shipboard Scientific Party, American Geophysical Union Fall Meeting, San Francisco, CA, 10 to 14 December 2001, Abstract V51C–1030 (2001).
9. L. M. Chambers *et al.*, American Geophysical Union Fall Meeting, San Francisco, CA, 6 to 10 December 2002, Abstract V71B–1271 (2002).
10. M. E. Clapham, P. R. Renne, *Annu. Rev. Earth Planet. Sci.* **47**, 275–303 (2019).
11. A. A. P. Koppers *et al.*, *Nat. Rev. Earth Environ.* **2**, 382–401 (2021).
12. M. F. Coffin, O. Eldholm, *Rev. Geophys.* **32**, 1–36 (1994).
13. M. L. G. Tejada *et al.*, *Geology* **37**, 855–858 (2009).
14. S. Méhay *et al.*, *Geology* **37**, 819–822 (2009).
15. G. Charbonnier, K. B. Föllmi, *Geology* **45**, 63–66 (2017).
16. C. Bottini, A. S. Cohen, E. Erba, H. C. Jenkyns, A. L. Coe, *Geology* **40**, 583–586 (2012).
17. E. Erba *et al.*, *Spec. Pap. Geol. Soc. Am.* **511**, 271–303 (2015).
18. K. W. Bauer *et al.*, *Geology* **49**, 1452–1456 (2021).
19. H. C. Jenkyns, *Geochem. Geophys. Geosyst.* **11**, Q03004 (2010).
20. E. Erba, *Mar. Micropaleontol.* **52**, 85–106 (2004).
21. Q. Jiang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202039119 (2022).
22. H. Matsumoto *et al.*, *Sci. Rep.* **10**, 12601 (2020).
23. H. Matsumoto *et al.*, *Nat. Commun.* **13**, 239 (2022).
24. S. Ingle, M. F. Coffin, *Earth Planet. Sci. Lett.* **218**, 123–134 (2004).
25. J. Korenaga, *Earth Planet. Sci. Lett.* **234**, 385–399 (2005).
26. J. G. Fitton, M. Godard, *Spec. Publ. Geol. Soc. Lond.* **229**, 151–178 (2004).
27. J. A. Tarduno *et al.*, *Science* **254**, 399–403 (1991).
28. J. J. Mahoney, K. J. Spencer, *Earth Planet. Sci. Lett.* **104**, 196–210 (1991).
29. C. Timm *et al.*, *Earth Planet. Sci. Lett.* **304**, 135–146 (2011).
30. A. J. Schaen *et al.*, *Geol. Soc. Am. Bull.* **133**, 461–487 (2021).
31. K. Hoernle *et al.*, *Geochim. Cosmochim. Acta* **74**, 7196–7219 (2010).
32. C. J. Sprain *et al.*, *Science* **363**, 866–870 (2019).
33. B. Schoene *et al.*, *Science* **363**, 862–866 (2019).
34. S. D. Burgess, S. A. Bowring, *Sci. Adv.* **1**, e1500470 (2015).
35. T. Sano, S. Yamashita, *Spec. Publ. Geol. Soc. Lond.* **229**, 185–218 (2004).
36. F. A. Frey *et al.*, *J. Geophys. Res.* **95**, 1271–1300 (1990).
37. M. G. Petterson, *Spec. Publ. Geol. Soc. Lond.* **229**, 63–81 (2004).
38. A. S. Gale *et al.*, in *Geologic Time Scale 2020*, vol. 2, F. M. Gradstein, J. G. Ogg, M. D. Schmitz, Gabi M. Ogg, Eds. (Elsevier, 2020), 1023–1086.
39. P. C. Davidson, A. A. P. Koppers, T. Sano, T. Hanyu, Geochemistry of dredged samples from near the Solomon Islands, Version 1.0, Interdisciplinary Earth Data Alliance (IEDA) (2023); https://doi.org/10.26022/IEDA/112922.

ACKNOWLEDGMENTS

We thank D. Heaton and D. Miggins for laboratory assistance. We also thank B. Duncan for his edits and comments. **Funding:** This project was funded by an IODP US Science Support Program Schlanger Fellowship award to P.C.D. and JSPS KAKENHI grant 18H03746 to T.S. and T.H. **Author contributions:** Conceptualization: P.C.D., A.A.P.K., and T.S. Methodology: P.C.D. and A.A.P.K. Investigation: P.C.D. and T.S. Visualization: P.C.D., T.S., and T.H. Funding acquisition: P.C.D., A.A.P.K., and T.S. Project administration: A.A.P.K. and T.S. Supervision: A.A.P.K. and T.S. Writing – original draft: P.C.D. Writing – review and editing: P.C.D., A.A.P.K., T.S., and T.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All ^{40}Ar/^{39}Ar data are available at EarthRef.org Digital Archive, and geochemistry data (discussed in the supplementary materials) are available at EarthChem.org (*39*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade8666
Materials and Methods
Supplementary Text
Figs. S1 to S8
References (*40*–*60*)
Tables S1 to S4
Data S1 to S3

Submitted 3 January 2023; accepted 11 May 2023
10.1126/science.ade8666

PLANT EVOLUTION

Leaves and sporangia developed in rare non-Fibonacci spirals in early leafy plants

Holly-Anne Turner[1]†, Matthew Humpage[2], Hans Kerp[3], Alexander J. Hetherington[1]*

Lateral plant organs, including leaves and reproductive structures, are arranged on stems in distinct patterns termed phyllotaxis. Most extant plants exhibit phyllotactic patterns that are mathematically described by the Fibonacci series. However, it remains unclear what lateral organ arrangements were present in early leafy plants. To investigate this, we quantified phyllotaxis in fossils of the Early Devonian lycopod *Asteroxylon mackiei*. We report diverse phyllotaxis in leaves, including whorls and spirals. Spirals were all n:(n+1) non-Fibonacci types. We also show that leaves and reproductive structures occurred in the same phyllotactic series, indicating developmental similarities between the organs. Our findings shed light on the long-standing debate about leaf origins and demonstrate the antiquity of non-Fibonacci spirals in plants.

Lateral plant organs such as leaves and reproductive structures are positioned on stems in a regular arrangement termed phyllotaxis. On theoretical grounds there could be a large number of possible phyllotactic patterns. However, this is not the case. A comparatively small number of discrete patterns are found in plants, and spirals are the most common (*1*–*3*). When a leafy shoot with spiral phyllotaxis is viewed from above, adjacent leaves are arranged in clockwise and anticlockwise spirals that radiate out from the center, and when viewed from the side these spirals form helices running around the circumference of the stem. Quantification of spiral phyllotaxis requires counting the number of clockwise and anticlockwise spirals, termed contact parastichies (*1*, *4*). Most commonly the number of clockwise and anticlockwise contact parastichies are integers in the Fibonacci sequence (1, 1, 2, 3, 5, 8, 13, 21…), and we will refer to these here as Fibonacci spirals. In a survey of angiosperms and gymnosperms, including 12,000 observations from 650 species, Fibonacci spirals occurred in >91% of observations (*1*, *3*). This number increased to >96% (*1*, *3*) when the duplicated Fibonacci series (2, 4, 6, 10, 16…) known as bijugate spirals (*1*, *5*) were included. Although there is some variation in the prevalence of Fibonacci spirals between different plant groups and between different organs (*3*–*6*), it remains the most common type of spiral phyllotaxis present in angiosperms, gymnosperms, ferns, and two of the three groups of lycopods (*1*, *3*–*5*). Why Fibonacci spirals are so common in plants has perplexed scientists for centuries (*1*, *7*–*9*), but their evolutionary origin has been largely overlooked (*7*). Given the widespread occurrence of Fibonacci spirals in living species, it was predicted that they were likely an ancestral and highly conserved characteristic of all seed plants (*7*) and likely all vascular plants (*4*, *10*, *11*). However, this hypothesis has not been vigorously tested (*5*, *6*, *12*), in part because of the difficulty of quantifying phyllotaxis in fossils. Thus, it remains unclear what phyllotactic types were present in early land plants and whether Fibonacci spirals were as common in the geological past as they are today.

Here, we investigated phyllotaxis in fossils of *Asteroxylon mackiei*, an Early Devonian lycopod and member of the earliest clade of leafy plants, the Drepanophycales (*13*–*16*). Lycopod leaves, termed microphylls, evolved entirely independently from the leaves of euphyllophytes (*17*–*19*), suggesting that the evolution of leaf phyllotaxis was divergent in the two clades. Despite this independent evolution, Fibonacci spirals are present in all three living clades of lycopods, just as they are in the major groups of euphyllophytes (*5*). However, in members of the Lycopodiales, non-Fibonacci spirals are present at a much higher frequency than Fibonacci spirals (*4*–*6*, *20*–*24*). In some species, such as *Lycopodium clavatum* and *Lycopodium annotinum*, Fibonacci spirals occur at a frequency of <1% (*5*, *20*), whereas in others Fibonacci spirals occur at a higher frequency but are only associated with a small number of developmental stages (*5*, *22*, *23*). Why non-Fibonacci spirals predominate in the Lycopodiales compared with the other two clades of living lycopods is a subject of ongoing debate (*6*, *20*, *23*), and it is currently unknown whether non-Fibonacci spirals in the Lycopodiales represent a derived or ancestral characteristic (*4*, *11*). Using a combination of classic fossil preparation techniques and three-dimensional (3D) digital reconstruction methods, we quantified phyllotaxis in multiple *A. mackiei* shoots to shed light on the origin of phyllotaxis in lyco leaves.

[1]Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, Max Born Crescent, Edinburgh EH9 3BF, UK. [2]Northern Rogue Studios, 18 Hunsdon Road, Iffley, Oxford OX4 4JE, UK. [3]Research Group for Palaeobotany, Institute for Geology and Palaeontology, University Münster, Heisenbergstrasse 2, 48149 Münster, Germany.
*Corresponding author. Email: sandy.hetherington@ed.ac.uk
†Present address: School of Biological, Earth and Environmental Sciences, University College Cork, Distillery Fields, North Mall, Cork T23 N73K, Ireland.

Quantification of *A. mackiei* phyllotaxis

To characterize phyllotaxis in *A. mackiei*, we first took a centric approach (*1*), in which leaf arrangement is characterized based on transverse sections close to the shoot apex. This approach was made possible in *A. mackiei* because of the exceptional level of preservation of leafy shoots in the Rhynie chert (*14*, *16*, *25*–*27*). We identified *A. mackiei* apices in two previously published thin sections, Pb 4123 (*25*) and GLAHM Kid 2554 (*28*), and in two new preparations generated during this study (Fig. 1). To quantify phyllotaxis, we represented the center point of each leaf by a point and connected neighboring leaves with both clockwise (red) and anticlockwise (blue) spirals, termed contact parastichies (*4*). Phyllotaxis of each apex was then presented as a ratio of the number of clockwise and anticlockwise contact parastichies, such as x:y where y>x. Based on our analysis, we identified the phyllotactic patterns 5:5 (Fig. 1A), 6:6 (Fig. 1B), 7:8 (Fig. 1C), and 8:9 (Fig. 1D). Phyllotaxis differed between each specimen and included both spiral and whorled types. Whorled types contain equal numbers of clockwise and anticlockwise contact parastichies (Fig. 1, A and B), whereas in spiral types they differ (Fig. 1, C and D) (*1*, *5*, *9*). The two specimens with spiral phyllotaxis 7:8 (Fig. 1C) and 8:9 (Fig. 1D) can both be described by the equation n:(n+1). n:(n+1) spiral phyllotaxis, in which $n \geq 3$, is rare in living plants and represents non-Fibonacci spirals (*1*, *3*–*5*). Based on our centric approach, we demonstrated variable phyllotaxis in *A. mackiei* with both non-Fibonacci spirals and whorled types.

To investigate phyllotaxis further, we took a cylindrical approach (*1*) by investigating leaf arrangement around the circumference of plant stems. We digitally reconstructed leafy shoots in 3D using serial peel preparations from the A. Bhutta Collection at the University of Cardiff (*29*). We digitally reconstructed three leafy shoots (Fig. 2). Figure 2A shows a preserved oblique section through a small leafy shoot, 4.1 mm in diameter, with only the front surface preserved. Figure 2B shows a preserved leafy shoot apex; the specimen was 5.8 mm in diameter, 28 mm long, and included 62 well-preserved leaves. Figure 2C shows the largest specimen at 46 mm in length and 7.9 mm in diameter, with 60 fully preserved leaves. Figure 2C also shows a preserved lateral leafy shoot bud (Fig. 2C, arrowhead, and fig. S1). This lateral bud, along with small apices identified on the other two reconstructions (fig. S1), demonstrates that lateral buds were common in *A. mackiei*. To quantify phyllotaxis in these specimens, we used the software Blender to digitally "unwrap" the

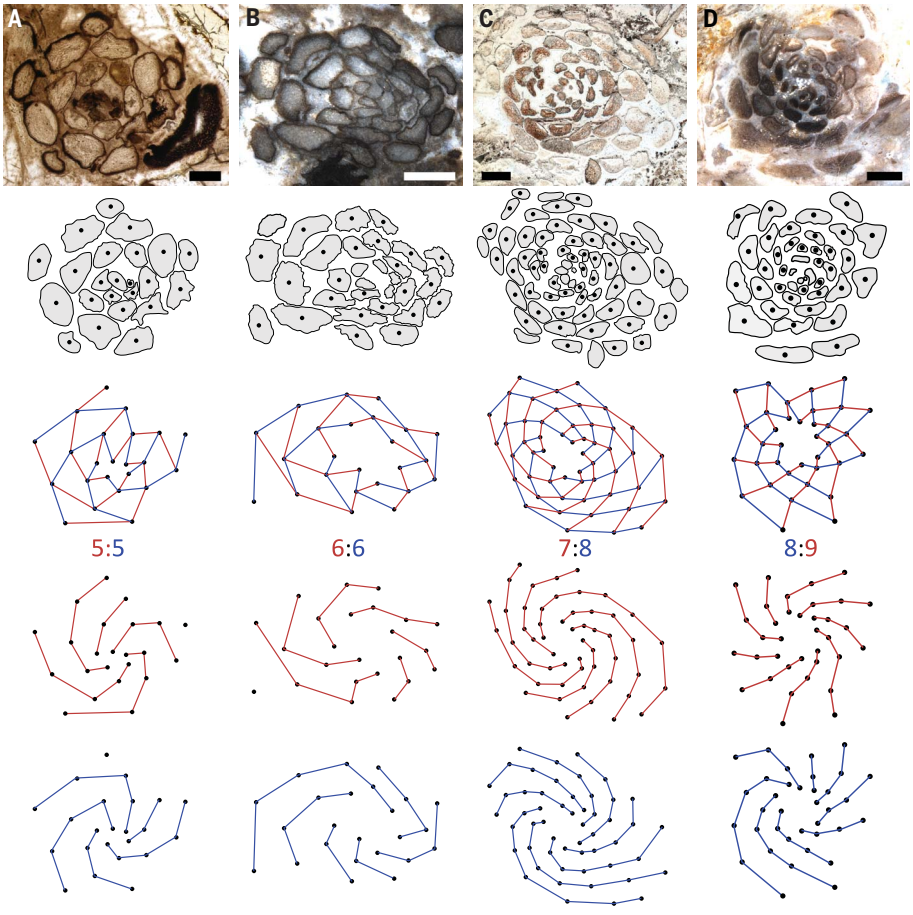


Fig. 1. Phyllotactic diversity in *A. mackiei* based on a centric investigation of shoot apices. (**A** to **D**) Row 1 shows transverse sections of four *A. mackiei* apices. Row 2 shows a line drawing made from (A) to (D), with leaves in gray and the center of all leaves included in the analysis marked with a black point. Rows 3 to 5 show clockwise (red) and anticlockwise (blue) contact parastichies. Specimen accession codes: Pb 4123 (A), Pb 5022 (B), GLAHM Kid 2554 (C), NMS G.2022.15.1 (D). Scale bars, 1 mm.

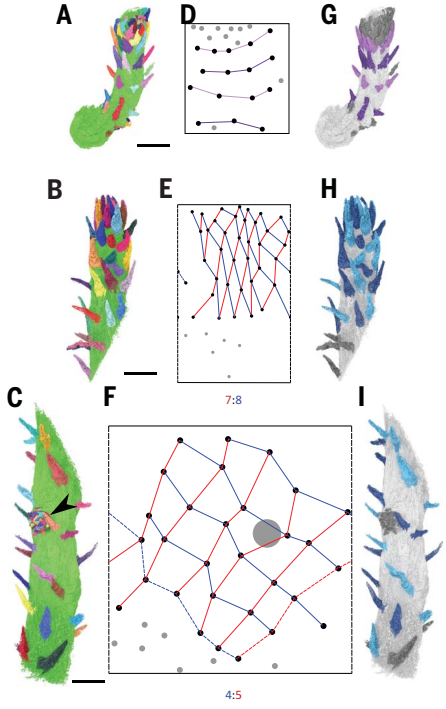


Fig. 2. Cylindrical quantification of phyllotaxis based on 3D digital reconstructions of leafy shoots. (**A** to **C**) 3D digital reconstruction of *A. mackiei* leafy shoots showing leaves randomly colored. Black arrow in (C) marks lateral bud. (**D** to **F**) 3D leaf arrangement represented on a 2D lattice with points representing leaf positions and lines representing contact parastichies: clockwise (red), anticlockwise (blue). Dashed lines represent continuations of parastichies from one side of the lattice to the other. Black points indicate leaves included in lattice, and gray points indicate leaves or putative leaves excluded from the analysis. Large gray circle in (F) marks position of the bud. (**G** to **I**) 3D digital models with leaves colored to highlight phyllotaxis. (G) Alternating light and dark purple leaves indicate successive whorls. (H) and (I) Alternating light and dark blue leaves indicate successive anticlockwise contact parastichies. Scale bars, 5 mm.

stem to visualize 3D leaf arrangement as a 2D lattice (Fig. 2, D to F). We represented the positions of leaves used for quantification by black points, whereas gray points were used to denote leaves not included in the analysis. In Fig. 2A, only the front surface of the axis is preserved, so we could not quantify parastichies around the complete diameter of the specimen. However, based on the front portion, we observed four successive horizontal whorls of leaves separated by internodes lacking leaves, which is consistent with a whorled arrangement (Fig. 2G). For the other two specimens, for which leaves extended around the full circumference of the shoots, we annotated the lattice diagrams with both clockwise (red) and anticlockwise (blue) contact parastichies. Based on this analysis, we recognized phyllotaxis as 7:8 (Fig. 2B) and 4:5 (Fig. 2C), both non-Fibonacci n:(n+1) spirals. As with our centric analysis, we again recognized both spiral and whorled phyllotaxis in *A. mackiei*.

To further build on our investigation, we next quantified phyllotaxis using both a centric and cylindrical approach within a single specimen. To achieve this we leveraged our 3D reconstruction of a shoot apex (Fig. 2B). Although the original peel preparations used to reconstruct Fig. 2B were longitudinal sections (Fig. 3A), we predicted that it would be possible to create a virtual transverse section necessary for a centric investigation of phyllotaxis. Using the software Blender, we first filled holes within our original skeletal model (Fig. 3B) and smoothed pixilated regions of the apex. This step allowed us to clearly visualize the dense clustering of leaves at the apex (Fig. 3C). Next we cut a digital transverse section through the apex, allowing us to quantify phyllotaxis (Fig. 3D). At the center of the apex a 7:7 phyllotaxis was apparent. However, when the parastichies are followed outward from the center, one of the anticlockwise parastichies splits, resulting in the 7:8 phyllotaxis observed when using a cylindrical approach (Fig. 3, E and F; Fig. 2E; and fig. S2). This shoot apex therefore preserves evidence of a transition between whorled and spiral phyllotaxis. Taken together our analysis of phyllotaxis in *A. mackiei* indicated that phyllotaxis was highly diverse, consisting of spirals and whorls; that spirals were non-Fibonacci n:(n+1) types; and that transitions in phyllotaxis were possible in a single leafy shoot.

Comparison of *A. mackiei* phyllotaxis with extant species

Based on studies of extant species alone, it was predicted that Fibonacci spirals were likely an ancestral and highly conserved characteristic of all major groups of vascular plants (*4*, *10*, *11*). Our findings from *A. mackiei* do not support this hypothesis. Instead they indicate that n:(n+1) non-Fibonacci types were present early

in land plant evolution despite being a rare phyllotactic type in living species (*1*, *3*–*5*). There are at least two groups of living plants where non-Fibonacci n:(n+1) spirals occur more commonly, some species of cacti (*4*, *30*, *31*) and members of the Lycopodiales (*4*–*6*, *20*–*24*). Given that both *A. mackiei* and the Lycopodiales are members of the lycopods (*13*), we sought to interpret our new findings specifically in the context of lycopod evolution. Extant lycopods consist of three clades: the Lycopodiales, Selaginellales, and Isoetales (*32*). In members of the Isoetales, all reported spiral phyllotaxis is Fibonacci (*6*, *33*, *34*). In the Selaginellales, spiral phyllotaxis occurs in the Homoeophyllae and *Selaginella selaginoides* clades (*35*). Spirals are Fibonacci in the larger Homoeophyllae clade, with only the single species *S. selaginoides* reported as developing more complex spirals (*6*). However, detailed characterization of 10 members of the Lycopodiales demonstrates that they share all the characteristics of *A. mackiei* phyllotaxis described here, including high variability, regular transitions in phyllotaxis, and the almost exclusive development of non-Fibonacci spirals including n:(n+1) types (*4*–*6*, *20*–*24*). Based on investigations of living species and fossils from the Carboniferous, Church (*4*, *11*) predicted that the non-Fibonacci n:(n+1) spirals found in living Lycopodiales were a derived condition from an ancestor with Fibonacci spirals. However, our findings from *A. mackiei*, a member of the extinct Drepanophycales, challenge this hypothesis and instead suggest that non-Fibonacci n:(n+1) spirals and whorls were ancestral in lycopods. Evidence for this hypothesis includes data presented here for *A. mackiei*, the low frequency of Fibonacci spirals in living Lycopodiales, and the absence of Fibonacci spirals from other fossil lycopods such as Early Carboniferous *Oxroadia* species (*36*, *37*). To our knowledge, the earliest reports of Fibonacci spirals in lycopods are in arborescent taxa in the Carboniferous (*11*, *38*), suggesting a late origin of Fibonacci spirals in lycopods. Lycopods were the earliest group of leafy plants, therefore suggesting that non-Fibonacci spirals in leaves predated the widespread occurrence of Fibonacci spirals in land plants. Finally, because leaves evolved independently in lycophytes and euphyllophytes (*17*–*19*) and non-Fibonacci types appear ancestral in lycophytes, this suggests that Fibonacci spirals present in living members of both groups may have evolved independently. Collectively our analysis of *A. mackiei* leaves changes our interpretation of the evolution of phyllotaxis in lycopods and demonstrates the antiquity of non-Fibonacci spirals.

The identification of the antiquity of non-Fibonacci spirals in lycopods begs the question of why non-Fibonacci spirals are present in members of the Lycopodiales and *A. mackiei* but are found less frequently in other groups of vascular plants (*1*, *3*–*5*). Based on studies of extant members of the Lycopodiales, a number of hypotheses have been put forward to explain the higher occurrence of non-Fibonacci spirals. These include predictions related to the shape or organization of the meristem, the relative size of leaf primordia compared with the apex, the presence of dichotomous branching, or different regulation of phyllotaxis in lycopods compared with euphyllophytes (*5*, *6*, *20*, *23*). Of these, dichotomous branching is considered the most likely cause (*6*, *20*, *23*) as a result of the major changes to meristem size and shape involved in branching and associated restriction of the sites for new primordia development (*6*, *23*, *39*, *40*). In support of this hypothesis is the observation that non-Fibonacci spirals occur at higher frequencies in euphyllophytes with dichotomous branching (*31*, *39*, *40*), and dichotomous branching was common in *A. mackiei* (*15*, *16*). The link between dichotomous branching and non-Fibonacci spirals does not hold true for all species, including members of the Selaginellales and palms (*5*, *6*, *40*, *41*), suggesting that it is not the only contributing factor to the development of non-Fibonacci spirals. However, if dichotomous branching is correlated with the development of non-Fibonacci spirals in land plants, it suggests that non-Fibonacci spirals may have been more frequent during the early diversification of vascular plants from the Silurian to the Carboniferous, when dichotomous branching was more common than today (*42*). The analysis of phyllotaxis in the fossil record therefore represents a key line of evidence for uncovering the evolutionary history of Fibonacci spirals in land plants.

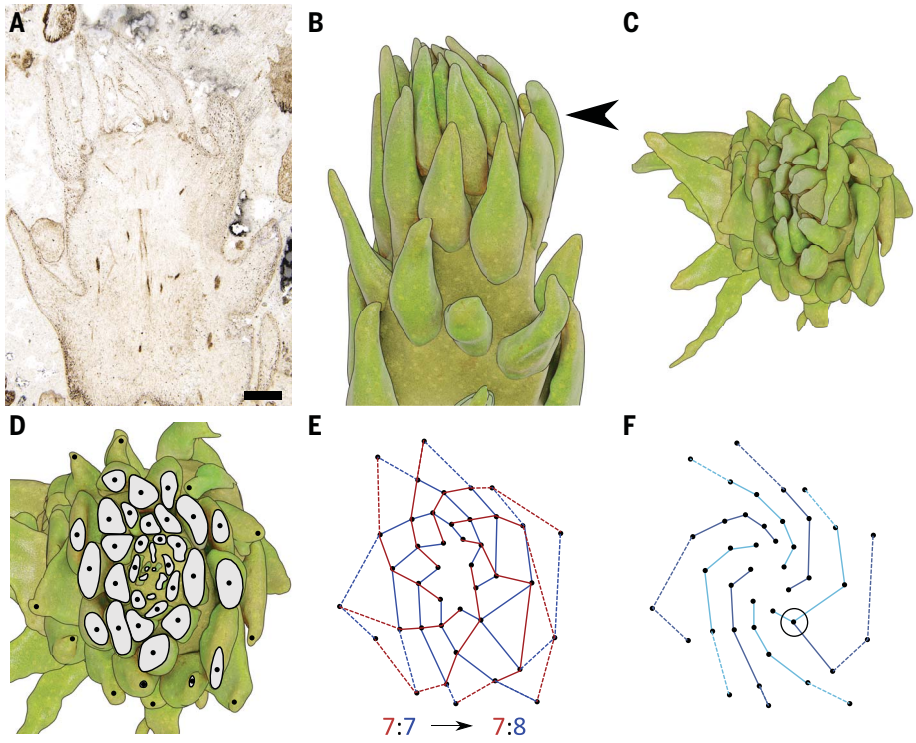


Fig. 3. Evidence of a phyllotactic transition within *A. mackiei*. (**A**) Peel used to create 3D model showing a longitudinal section through the shoot apex. (**B** and **C**) 3D digital rendering of a side-on (B) and top-down (C) view of the apex shown in Fig. 2B. (**D**) Virtual transverse section made through the apex at position marked by the black arrow in (B). Leaves cut in transverse section are shown in gray with black points indicating the center of sectioned leaves and unsectioned leaf tips very close to the apex. (**E**) Lattice diagram of leaf arrangement based on (D) showing clockwise (red) and anticlockwise (blue) contact parastichies. Solid lines connect leaves cut in transverse section in (D), and dotted lines connect leaves just below the plane of the section. (**F**) Dark and light blue lines show successive anticlockwise parastichies, with the circled leaf marking the transition from 7 to 8 anticlockwise contact parastichies. Specimen accession code: A. Bhutta BL29A/40 (A). Scale bar, 1 mm.

Phyllotaxis of leaves and sporangia in a fertile *A. mackiei* shoot

Having characterized phyllotaxis in *A. mackiei* leaves and drawn conclusions concerning the evolution of phyllotaxis, we finally wanted to use our analysis to investigate a related question, the origin of leaves. The origin of leaves in lycophytes has been debated for over a century, with two major competing hypotheses in the literature. The enation hypothesis predicts that leaves evolved as de novo structures on the shoot, whereas the sterilization hypothesis predicts that leaves evolved by

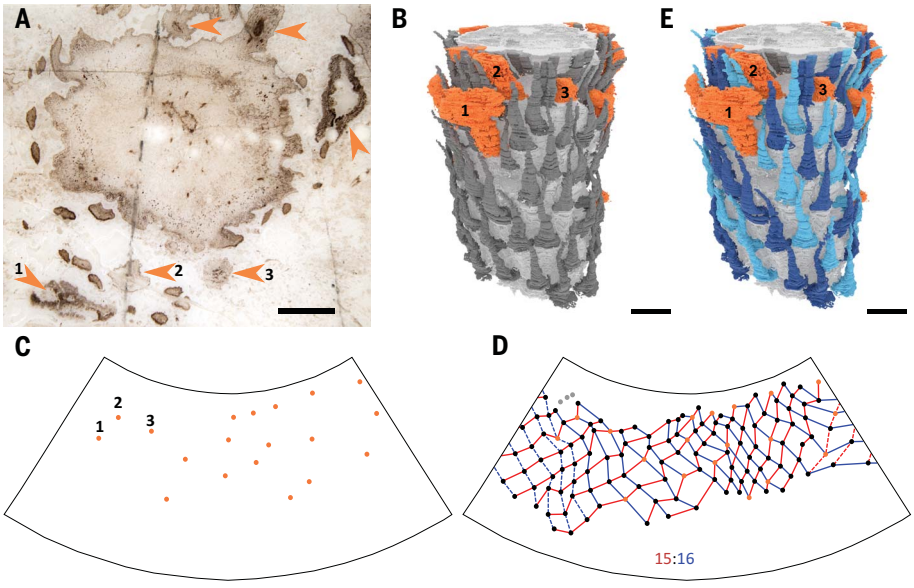


Fig. 4. *A. mackiei* sporangia are arranged in the same phyllotactic series as leaves. (A) Peel used to create the 3D model showing a transverse section of the stem with six sporangia surrounding the axis (orange arrowheads). Numbered arrowheads 1 to 3 highlight the position of three sporangia in (**A** to **C**) and (**E**). (B) 3D digital reconstruction of the fertile leafy shoot, with leaves in dark gray and sporangia in orange. (C) 3D arrangement of sporangia represented on a 2D lattice. (**D**) 2D lattice depiction of all leaves and sporangia. Orange points are sporangia, black points are leaves included in lattice, and gray points indicate leaves, or putative leaves, excluded from the analysis. Contact parastichies are marked with red (clockwise) and blue (anticlockwise) lines. (E) 3D digital model with leaves colored in alternating light and dark blue to highlight successive anticlockwise parastichies. Specimen accession code: A. Bhutta XYA/29 (A). Scale bars, 2 mm.

the sterilization of sporangia (*13*, *17*, *43*–*47*). *A. mackiei* leaves were traditionally interpreted as supporting the enation hypothesis for leaf origins (*44*). This is because the leaves of *A. mackiei* only developed a vascular trace that extended to the leaf base rather than along the full length of the leaf as in other lycopods (*13*, *16*, *25*, *27*, *29*). The leaves of *A. mackiei* were therefore interpreted as a transitional stage between unvascularized enations and fully vascularized leaves, (*44*). However, subsequent detailed cladistic analysis of early lycophytes, including *A. mackiei*, instead suggests that the leaves of *A. mackiei* are homologous to leaves of all lycopods and the absence of the leaf trace in *A. mackiei* is due to a secondary loss (*13*, *17*, *45*). We investigated the phyllotactic relationships between leaves and sporangia to help evaluate the enation and sterilization hypotheses. To investigate phyllotaxis in both sporangia and leaves, we reconstructed a fertile shoot of *A. mackiei* (Fig. 4). Unlike in living lycopods, where sporangia are always positioned on the adaxial sides of specialized leaflike structures termed sporophylls (*48*), sporangia of *A. mackiei* were borne laterally on stems connected by short stalks (*25*). Sporangia of *A. mackiei* could be readily differentiated from leaves as a result of their larger size, prominent vascular trace and valvate tips, or abscission marks (*25*) (Fig. 4A, orange arrowheads).

The fertile region consisted of 19 sporangia and 118 small and densely packed leaves. We quantified phyllotaxis by taking a cylindrical approach employing a conical lattice because of different diameters at the top and base of the axis (Fig. 4, C and D). On the lattice we first plotted the position of the sporangia (Fig. 4C). Based solely on sporangia, we were unable to identify any regular phyllotactic arrangement. We next investigated leaves and sporangia together (Fig. 4D). Based primarily on the positions of leaves, we identified a 15:16 spiral phyllotaxis; notably, the positions of the sporangia were included in this phyllotactic pattern. This suggests that leaves and sporangia developed in a single phyllotactic series and can be considered as topographically homologous (*49*). Taken together, the lack of phyllotaxis in sporangia when examined in isolation but their incorporation with leaves into a recognizable n:(n+1) spiral highlights the developmental similarity between the two organs.

The occurrence of leaves and sporangia in the same phyllotactic series alone does not allow us to conclusively reject either the enation or sterilization hypothesis. Shared development could result from either shared origin of the two organs (sterilization hypothesis) or convergence on the same developmental patterning mechanism for two distinct organs (enation hypothesis). However, our findings do add to a growing body of evidence in support of the sterilization hypothesis. Previous support for the sterilization hypothesis suggests that this is a developmentally parsimonious interpretation for leaf evolution because it does not require the evolution of a new organ system and leaves and sporangia share developmental similarities, including similarities in gene expression (*13*, *17*, *45*, *46*). Development in the same phyllotactic series of both sporangia and leaves in *A. mackiei* adds a further point of similarity between the two organs, consistent with the sterilization hypothesis (*13*, *17*, *45*, *46*). Our quantitative assessment of phyllotaxis in a fertile apex therefore adds evidence to the long-standing debate about the origin of leaves in lycophytes.

REFERENCES AND NOTES

1. R. V. Jean, *Phyllotaxis* (Cambridge Univ. Press, 1994).
2. R. D. Meicenheimer, in *Symmetry in Plants*, R. V. Jean, D. Barabé, Eds. (World Scientific, 1998), pp. 125–143.
3. R. V. Jean, *J. Theor. Biol.* **156**, 41–62 (1992).
4. A. H. Church, *On the Relation of Phyllotaxis to Mechanical Laws* (Williams & Norgate, 1904).
5. E. M. Gola, A. Banasiak, *Acta Soc. Bot. Pol.* **85**, 1–21 (2016).
6. J. C. Schoute, in *Manual of Pteridology*, F. Verdoorn Ed. (Martinus Nijhoff, 1938), pp. 1–64.
7. D. Reinhardt, E. M. Gola, *Trends Plant Sci.* **27**, 1017–1032 (2022).
8. I. Adler, D. Barabé, V. Jean, *Ann. Bot.* **80**, 231–244 (1997).
9. X. Yin, *J. Plant Res.* **134**, 373–401 (2021).
10. R. V. Jean, *Ann. Bot.* **61**, 293–303 (1988).
11. A. H. Church, *On the Interpretation of Phenomena of Phyllotaxis* (Oxford Univ. Press, 1920).
12. E. Véron, T. Vernoux, Y. Coudert, *Trends Plant Sci.* **26**, 124–131 (2021).
13. P. Kenrick, P. R. Crane, *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institute Press, 1997).
14. D. Edwards, *Trans. R. Soc. Edinb. Earth Sci.* **94**, 397–410 (2004).
15. A. J. Hetherington *et al.*, *eLife* **10**, e69447 (2021).
16. R. Kidston, W. H. Lang, *Trans. R. Soc. Edinb.* **52**, 643–680 (1920).
17. P. R. Crane, P. Kenrick, *Plant Syst. Evol.* **206**, 161–174 (1997).
18. C. K. Boyce, *Curr. Opin. Plant Biol.* **13**, 102–107 (2010).
19. A. M. F. Tomescu, *Trends Plant Sci.* **14**, 5–12 (2009).
20. E. Gola, *Acta Soc. Bot. Pol.* **65**, 235–247 (1996).
21. D. W. Stevenson, *Bot. J. Linn. Soc.* **72**, 81–100 (1976).
22. E. Vindt-Balguerie, *Bull. Soc. Bot. Fr.* **138**, 135–146 (1991).
23. X. Yin, R. D. Meicenheimer, *Am. J. Bot.* **104**, 8–23 (2017).
24. R. Rutishauser, in *Symmetry in Plants*, R. V. Jean, D. Barabé, Eds. (World Scientific, 1998), pp. 171–212.
25. H. Kerp, C. H. Wellman, M. Krings, P. Kearney, H. Hass, *Int. J. Plant Sci.* **174**, 293–308 (2013).
26. H. Kerp, *Philos. Trans. R. Soc. B.* **373**, 20160495 (2018).
27. F. M. Hueber, *Ann. Mo. Bot. Gard.* **79**, 474–499 (1992).
28. R. Kidston, W. H. Lang, *Trans. R. Soc. Edinb.* **52**, 831–854 (1921).
29. A. A. Bhutta, Studies on the Flora of the Rhynie Chert. University of Wales. (1969).
30. J. D. Mauseth, *Int. J. Plant Sci.* **181**, 518–528 (2020).
31. E. Gola, *Acta Soc. Bot. Pol.* **66**, 237–257 (1997).
32. PPG I, *J. Syst. Evol.* **54**, 563–603 (2016).
33. W. F. B. Hofmeister, *On the Germination, Development, and Fructification of the Higher Cryptogamia*, and on the Fructification of the Coniferae. Translated by F. Currey (Hardwicke, 1862).
34. C. West, H. Takeda, *Trans. Linn. Soc. Bot* **8**, 333–376 (1915).
35. X.-M. Zhou *et al.*, *Cladistics* **32**, 360–389 (2016).
36. R. M. Bateman, *Palaeontogr. Abteilung B.* **228**, 29–103 (1992).
37. K. L. Alvin, *Palaeontology* **8**, 281–293 (1965).
38. A. Dickson, *Trans. Bot. Soc. Edinb.* **11**, 145–147 (1873).
39. B. Zagórska-Marek, *Can. J. Bot.* **63**, 1844–1854 (1985).
40. E. M. Gola, *Front. Plant Sci.* **5**, 263 (2014).
41. P. B. Tomlinson, *Ann. Bot* **35**, 865–879 (1971).
42. G. Chomicki, M. Coiro, S. S. Renner, *Ann. Bot.* **120**, 855–891 (2017).
43. F. O. Bower, *The Origin of a Land Flora* (Macmillan, 1908).

44. F. O. Bower, *Primitive Land Plants also known as the Archegoniatae* (Macmillan, 1935).
45. P. Kenrick, in *Developmental genetics and plant evolution*, Q. C. B. Cronk, R. M. Bateman, J. Hawkins, Eds. (Taylor & Francis, 2002), pp. 365–387.
46. A. Vasco *et al*., *New Phytol.* **212**, 745–758 (2016).
47. C. J. Harrison, M. Rezvani, J. A. Langdale, *Development* **134**, 881–889 (2007).
48. D. W. Bierhorst, *Morphology of Vascular Plants* (Macmillan, 1971).
49. E. G. Cutter, in *Trends in Plant Morphogenesis: Essays presented to C. W. Wardlaw on His Sixty-fifth Birthday*, E. G. Cutter, Ed. (Longmanns, 1966), pp. 220–234.
50. A. J. Hetherington, H-A. Turner, *Asteroxylon mackiei* 3D reconstructions for investigation of phyllotaxis. (University of Edinburgh, Edinburgh DataShare, 2023); https://doi.org/10.7488/ds/3808.

ACKNOWLEDGMENTS

We thank D. Edwards and C. Berry for access to the A. Bhutta collection at the University of Cardiff. This invaluable assistance allowed the project to be carried out during the pandemic. We thank N. Clark for access to the Kidston Collection. We thank N. Fraser, A. Ross, and Y. Candela for assistance accessioning material into National Museums Scotland. We thank F. Buckley and North Sea Core for their assistance with Rhynie chert specimens. We thank the five reviewers for constructive comments on the manuscript. **Funding:** This work was funded by the following: UK Research and Innovation Future Leaders Fellowship MR/T018585/1 (to A.J.H.); Royal Society research grant RGS\R2\212063 (to A.J.H.); Deutsche Forschungsgemeinschaft KE 584/13- 1 (to H.K.); Deutsche Forschungsgemeinschaft KE 584/13- 2 (to H.K.). **Author contributions:** Conceptualization: A.J.H. Methodology: A.J.H., H.-A.T., and M.H. Investigation: H.-A.T., A.J.H., M.H., and H.K. Visualization: H.-A.T., A.J.H., M.H., and H.K. Funding acquisition: A.J.H. and H.K. Supervision: A.J.H. Writing – original draft: A.J.H. and H.-A.T. Writing – review and editing: H.-A.T., A.J.H., M.H., and H.K. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Fossil specimens described in this study are deposited in four collections. Collection abbreviations: Pb: Forschungsstelle für Paläobotanik, Institut für Geologie und Paläontologie, University Münster, Münster, Germany. GLAHM: The Hunterian, University of Glasgow. NMS: National Museums Scotland. A. Bhutta: Akhlaq Ahmed Bhutta peel collection, University of Cardiff. Accession numbers for all figured specimens are: Pb 4123 (Fig. 1A), Pb 5022 (Fig. 1B), GLAHM Kid 2554 (Fig. 1C), NMS G.2022.15.1 (Fig. 1D), A. Bhutta BL29A/40 (Fig. 3A), and A. Bhutta XYA/29 (Fig. 4A). 3D reconstructions are based on peel numbers A. Bhutta 139A/2-072 (Fig. 2A), A. Bhutta BL29A/15-090 (Fig. 2B, Fig. 3), A. Bhutta BL29A/ 50-182 (Fig. 2C), and A. Bhutta XYA/4-181 (Fig. 4B). Cropped and aligned images of peels used to generate 3D reconstructions, are deposited on Edinburgh University DataShare (*50*), as are the 3D reconstructions. An interactive 3D model of the specimen Fig. 3B can be viewed on Sketchfab: https://sketchfab.com/3d-models/asteroxylon-mackiei-f090e445a53a47cd8b394b1ee7d9f517. All other data are available in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

Submitted 22 December 2022; accepted 15 May 2023
10.1126/science.adg4014

WORKING LIFE

By **Azmi Ahmad**

First aid for the psyche

On the second day of a conference, I awoke to a 7 a.m. phone call from my dad. "Azmi, your grandfather had a brain hemorrhage and was admitted to the hospital." I told my dad I'd fly out to meet him that day, but he encouraged me to stay to present my research. By the time I arrived the following night—poster and conference clothes in hand—my grandfather had passed away. The rush of grief was overwhelming. Over the next few days, I focused on my family and helped plan the funeral. Then, when the time came to get back to work, I developed a self-care plan and communicated my needs to others. I knew what to do thanks to a mental health first aid course I had taken, which taught me how to care for myself as well as others.

After finishing my Ph.D. in the fateful spring of 2020, I had been unemployed for 4 months, so I was excited to start my postdoc, moving 3000 kilometers across the country in the middle of the pandemic. Yet something felt off. Workplace hours were staggered and much of the in-person training for my projects was virtual. I was lucky to have a supportive supervisor, but I did not know anyone outside my direct colleagues in the lab. I coped with my feelings of anxiety and isolation by burying myself in work, putting in long hours at the bench.

After getting involved with my postdoctoral association's advocacy committee, I quickly saw that many incoming postdocs were facing similar troubles. Wanting to do what I could to help, I took on the role of disseminating information about work-life balance and mental health, announcing seminars, meditation sessions, and mental well-being programs.

One day, the organizers of a mental health first aid course asked me to send out an announcement. I was intrigued. I had never dealt with a mental health emergency. But it struck me that if I wanted to be the best advocate for my community, I should lead by example and take the course myself.

When training day came, I expected to mostly learn about how to handle emergencies. But the 6-hour session was much more broad ranging, covering how to deal with chronic situations and what to do to take care of yourself. A key lesson was that no one should struggle alone, and that it's important to develop a self-care plan by thinking about the people, programs, and activities in your support network that can help you deal with mental health challenges.

"A key lesson was that no one should struggle alone."

After taking the course, I felt more confident in my ability to defuse difficult situations and direct people toward professional help should I ever come across an emergency. The training helped me become a better listener. For instance, I've learned to be mindful of my body language—to make eye contact, maintain a soft tone, and focus on being present and actively listening. When a friend reached out to say they felt dismissed at work and wanted to quit, I listened and gave reassurance as best I could. I've also found I can listen more attentively to my own thoughts and feelings and identify steps to help me feel better.

When my grandfather died, I didn't bury myself in work as I had done in the past. Instead, I put my well-being first and reached out to others who could help me. I told my adviser that I'd need to take a break and miss a lab meeting, and he understood. I created a physical health plan—adding more exercise to my daily routine—and took time each morning to practice guided meditation. I also opened up about my feelings with close friends, who were supportive and would then regularly check in on how I was doing. These steps didn't magically erase my grief. But they helped me work through it and connect with those around me.

You don't need a formal course. But develop your own mental health first aid kid with self-care strategies, and don't be afraid to connect with others. Academia—and life in general—can be challenging. We're better off when we can rely on others and support ourselves. ■

Azmi Ahmad is a postdoctoral researcher at Yale University. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.