THIS WEEK

EDITORIALS

PUBLISHING Dispute over open data signals need for more debate **p.140**

WORLD VIEW The robot telescope bringing practical science to schools **p.141**

EVOLUTION Separate origins for fish adipose fins reveal their value **p.142**

Don't hide the decline

US scientists should not be placated by the 'flat budget' myth. Funds are decreasing, and the situation will get worse.

For US researchers, the annual unveiling of the presidential budget request can be a time of both hope and trepidation. But after last year's fiscal battles with Congress, complete with an embarrassing government shutdown and painful across-the-board spending cuts, it was always clear that this year there would be little to celebrate.

In that atmosphere, the unveiling on 4 March of President Barack Obama's US$3.9-trillion budgetary vision for fiscal year 2015 brought both disappointment and a sigh of relief. In one sense, the proposal was optimistic: it exceeded congressional spending limits by $56 billion, and there were few deep cuts for science. But it leaves the budgets of major scientific funders, such as the US National Institutes of Health (NIH), the National Science Foundation (NSF) and the research efforts at the Department of Energy, essentially flat (see page 147).

Amid a sluggish economy and zealous calls to tighten federal purse strings, the prevailing wisdom is often to be grateful for a flat budget. Things could be worse. But those projects that stand to be gutted — such as the Stratospheric Observatory for Infrared Astronomy (SOFIA), an airborne observatory funded largely by NASA, which would have its budget slashed from $84 million to $12 million — stand as painful reminders that a flat budget is not something to celebrate. The proposed $200-million boost to bring the NIH's budget to $30.2 billion is paltry, but even worse is the $1.3-billion cut that could be in store for the Department of Health and Human Services, the NIH's parent agency.

What is more, inflation does not stand still for flat budgets. Overall spending on research and development would increase by 1.2% in 2015 if Obama has his way. But the rate of inflation that year is expected to be 1.7%. The outlook is worse for biomedical research — here, inflation is projected to rise by 2.2% in 2015, according to the Department of Health and Human Services' Biomedical Research and Development Price Index. The 0.7% budgetary bump that Obama has requested will not keep pace.

Indeed, 'flat' budgets such as those proposed last week have steadily eroded the NIH's coffers over the past decade. Controlling for inflation, the NIH's budget shrank by 10% between 2004 and 2014, according to the American Association for the Advancement of Science in Washington DC. The real decline is even steeper when the rate of biomedical inflation is taken into account.

A similar trend is emerging for research and development overall: federal spending on research and development in 2014 is 15.8% lower than in 2010 when inflation is considered.

Greener pastures are nowhere in sight. The president's request was sent to Congress, which will produce a plan of its own. Included in Obama's request is a proposed $56-billion Opportunity, Growth, and Security Initiative that would add $5.3 billion to the nation's research and development coffers. But there is little reason to hope that the initiative will make it through a US Congress determined to rein in spending, opposed to raising taxes and not generally known for a willingness to compromise. These are, after all, the same legislators who in October shut down the government for 16 days and allowed across-the-board spending cuts of 5% last year. Science suffered as a result: the NSF awarded 690 fewer grants in 2013 than the previous year, according to figures released last week by the Government Accountability Office. The NIH cut its grants by 750. The White House's budget proposal makes it clear: there will be no compensation for these lost opportunities.

"Rather than a relief, apparently flat budgets are a sure sign that competition for funds will grow still further."

Meanwhile, the economic strain on the country is immense. Mandatory spending obligations — on retirement and healthcare programmes, for example — are soaring, squeezing discretionary spending on other worthy areas, including research. As a result, discretionary programmes are battling over slices of a rapidly shrinking pie: in 2010, discretionary funds were 39% of the budget; in 2015, they will be 30%.

This means that the fight will only be more intense in years to come. Rather than a relief, apparently flat budgets are a sure sign that competition for funds will grow still further. And that things will get worse before they get better. ■

An elegant chaos

Universal theories are few and far between in ecology, but that is what makes it fascinating.

To some scientists in other fields, ecology must seem relatively straightforward. Many of the organisms live at a very human scale and are easy to access, especially in community ecology. Ecologists do not need special equipment to see and count elk. There are no electron microscopes, space telescopes or drilling rigs that can go wrong. Easy.

And yet, ecologists know that their subject can prove as troublesome as any other. Ecology would be easy, were it not for all the ecosystems — vastly complex and variable as they are. Even the most austere desert or apparently featureless moor is a dense, intricate network of thousands of species of photosynthesizers, predators, prey animals, parasites, detritovores and decomposers. As naturalist E. O. Wilson put it: "A lifetime can be spent in a Magellanic voyage around the trunk of a single tree." And not all of what one might learn from such a voyage would be transferable to the next tree. History, chance, climate, geology and — increasingly — human fiddling mean that no two ecosystems work in the same way.

Scientists like to impose structure and order on chaos, and ecologists

are no different. Ecology has its grand theories, but they are riddled with conditional clauses, caveats and exceptions. There are clear patterns at the global and single-species scales, but the middle ground is, as biologist John Lawton affectionately put it in 1999, "a mess". It is doubtful that the generalities that underlie the complex patterns of nature will ever be phrased succinctly enough to fit on a T-shirt.

This complexity is demonstrated by work that questions a famous and elegant 'trophic cascade' in Yellowstone National Park, Wyoming, discussed on page 158. The theory goes that wolves, restored to the park in the 1990s after decades of absence, scare elk away from certain areas. That has a knock-on effect for the rest of the food chain, allowing aspen and willows to flourish after decades of being browsed nearly to death. But studies in recent years suggest that wolves alone do not control the ecosystem. Other factors — the presence of beaver dams and grizzly bears, weather, hunting by humans and even climate change — also affect the elk population and the growth of trees and shrubs.

It would be useful to have broad patterns and commonalities in ecology. To know how ecosystems will respond to climate change, or to be able to predict the consequences of introducing or reintroducing a species, would make conservation more effective and efficient. But a unified theory of everything is not the only way to gain insight.

More ecologists should embrace the non-predictive side of their science. Teasing out what is going on in complex systems by looking at how ecosystems evolved, and by manipulating the environment in experiments, is just as much a science as creating formulae for how ecosystems work.

Paradigm shifts, after all, are rare in ecology. Debates are often resolved when competing concepts combine, rather than when one pushes the other completely off the table. Take the contrasting ideas of top-down regulation of ecosystems by carnivores and bottom-up regulation effected by the nutrition available from plants. The field is slowly working towards an integrated theory to predict when the top will rule and when the bottom will be in charge — and that theory will take the time to consider the middle players, the herbivores.

Other ecological debates have followed a similar path. Disagreement over whether complex ecosystems are more or less stable than simpler ones, for example, is also settling to a consensus: it depends.

Useful practical predictions need not stem from universal laws. They may come instead from a deep knowledge of the unique workings of each ecosystem — knowledge gained from observation and analysis. Proposing sweeping theories is exciting, but if ecologists want to produce work useful to conservation, they might do better to spend their days sitting quietly in ecosystems with waterproof notebooks and hand lenses, writing everything down.

"If ecosystems all worked in the same way, they would lose much of their mystery, their surprise and their beauty."

Ecological complexity, which may seem like an impenetrable thicket of nuance, is also the source of much of our pleasure in nature. If ecosystems were simple puzzles that all worked in the same way, they would lose much of their mystery, their surprise and their beauty. A lot of conservation work aims to protect the complexity and variability that makes ecosystems so hard to understand, and indeed to conserve.

Ecological rules are not the only reasons to promote conservation and fight extinctions. Sometimes we can argue for the conservation of particular species because ecology provides a scientific basis for it. At other times, we make the argument because there is a good chance that ecology will soon catch up and explain why the species are important.

But even if some predators do little but sit at the top of their food pyramids, creaming off a few herbivores, would we really want to live in a world without them? Answering that question really is easy. ■

Share alike

Research communities need to agree on standard etiquette for data-sharing.

Every mountaineer knows the sinking feeling of reaching a peak after a hard climb, only to see the true summit still above. Scientists who take on the tough terrain of open access may have a similar experience. After they reach the notable goal of sharing their research papers, they discover that a higher summit awaits: open data.

In many fields, making research data available online for all is a step beyond making research papers open-access. This might puzzle communities that have already agreed to share. Biologists routinely upload DNA sequences to the public repository GenBank, for example, creating a scientific commons for everyone's benefit. There are now more than 600 subject-specific repositories, with community-specific standards.

Yet even some of the most strident open-access supporters baulk at the concept of fully open data, judging by the reaction to a strengthened data-sharing policy instituted by the Public Library of Science (PLOS) this month. PLOS now requires researchers to make their papers' underlying data open online on publication, apart from data that they have a duty to keep private, such as that on human study participants (go.nature.com/rd27aa). Journals such as *Molecular Ecology* have mandated the same thing for years. But the PLOS move has provoked heated discussion and highlighted some important, yet unsettled, aspects of the practice and ethics of online data-sharing.

A few years ago, a survey found that scientists cited a lack of time and money, as well as technical barriers, to explain why they did not post data online (C. Tenopir *et al. PLoS ONE* **6,** e21101; 2011). It still takes time to prepare data, but increasingly, other excuses do not fly. General-purpose storage sites such as Dryad and figshare are cheap (or free) and suitable for all kinds of data sets; data journals provide publication venues appealing to the traditionally minded; and standards are emerging for citing other people's data sets (see *Nature* **500,** 243–245; 2013).

Harder to surmount are the feeling of data ownership and the fear of being 'scooped'. Years of toil to collect a data set that might support a decade of career-making publications could be rendered moot when another researcher jumps on the information online. This is a particular problem for early-career researchers, and for those working with unique data sets in small ecology or environmental-science laboratories.

Behind this fear is the worry that other scientists will not provide credit for the data they use. Research administrators place such importance on paper authorship that it is probably not enough for a study that leans significantly on another researcher's hard-won data set to merely cite that researcher, perhaps depriving them of a publication.

Communities need to debate the ethics of data-sharing and agree on etiquette. When a researcher relies on another's data, for example, it should be standard practice to invite the data-providers to be co-authors. Ecologists Clifford Duke and John Porter have suggested guidelines for deciding whether to extend such an invitation (C. S. Duke and J. H. Porter *BioScience* **63,** 483–489; 2013); these include noting whether the data are integral to the new analysis, whether the data are unique or particularly novel, and whether the data-provider can fully participate in manuscript-writing by approving draft and final versions. Another ecologist, Dominique Roche, has urged disclosure of data reuse, and better communication between data generators and reusers (D. G. Roche *et al. PLoS Biol.* **12,** e1001779; 2014).

It is not clear whether widespread online data-sharing will increase uncredited scooping. For now, *Nature* mandates uploading data when structured community data repositories exist, and encourages it otherwise. Before you can climb the highest mountains, you need proper safeguards and a decent map. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

Practical science has a global reach and appeal

As English schools consider downgrading practical science, **John Baruch** *points out that other nations are rushing to include more.*

England has a bizarre plan to downgrade the importance of practical science in schools. A consultation just closed by the Office of Qualifications and Examinations Regulation suggests that practical science and laboratory work should no longer contribute to the final mark for the A-level examination that students take at 18.

The move is especially odd given that other nations — Britain's competitors — are waking up to the need to include more practical science in their education systems. And British scientists are helping them to do it. I am one of them.

These countries — China, Poland and Ireland among them — realize that practical work is not just an integral part of science and essential to understand how science works, it is the best route to give students the skills they will need to support technological innovation. China, especially, has ambitious plans here: officials are working to change the culture of its school system so that it recognizes and rewards practical skills.

Practical science is more than hands-on science. It challenges the student to understand the real world, to create ways to test that understanding and to grasp the significance of statistics and errors in their arguments.

I am an astronomer, and my subject has a major advantage when it comes to practical and hands-on experience. We can automate and offer it remotely. At a stroke, this solves one of the obstacles to practical science in schools across the world: that lab work is expensive and requires skilled teachers and laboratory technicians, which are in short supply. Practical astronomy is easier — given the right equipment.

The Universe travels over our heads every night, and the only requirement for practical work is a telescope. In the late 1980s, the UK astronomical community, tired of the tedious need to guide these large instruments by eye, decided to investigate robotic telescopes. I was awarded a research contract to prove the concept of a telescope that could work autonomously.

The result was the Bradford Robotic Telescope (BRT). Initially perched high in the Yorkshire Pennines, it was the world's first fully automated instrument. Users submitted a list of objects they wished to observe and waited for the results to be returned to them by e-mail. Astronomers had priority, but the early years of the Internet allowed us to open up its use to thousands of others. We gave them free access to the instrument when the astronomers were not using it.

The telescope was then transferred to Mount Teide in the Canary Islands. It does everything for users: evaluates the weather, schedules itself to optimize observations, takes calibration data and returns the whole package along with analysis programmes. This is now an option on nearly all large telescopes, but for members of the public it remains unusual. (The norm is for users to have a fixed time slot of half an hour or so to drive the telescope, using web cameras at the observatory to see whether the weather is suitable and to move the telescope to point at the object they want to observe.)

Around 90,000 students and 2,500 teachers in Britain use the BRT. Secondary schools pay £195 (US$326) a year; primary schools £70. Every child has a username and can log on from home. More than one-third do so, and the results are stunning: children race back to school the next day to tell their teachers what they found.

In many schools, the telescope forms part of the GCSE astronomy programme (taken at around age 15), and astrophysics modules of the physics A level — one of the subjects for which officials are now trying to downgrade practical experience.

THE SKILLS OF **INNOVATION** AND CREATIVITY ARE THE **BEDROCK** OF A **KNOWLEDGE** ECONOMY.

Britain's loss could be China's gain. China has traditionally shown little interest in practical science in schools. Practical work does not feature in the school-leaving, or *Gaokao*, examination — the most important exam taken by Chinese young people (when they are 17) — so students, parents and teachers have never taken it very seriously.

That is now changing. A pilot programme run in the Beijing region and led by the Chen Jing Lun school will see practical-science projects contribute. The BRT will be the lead project offered to the students, starting this spring term. The British Council — which promotes international educational opportunities and cultural relations — has helped us to translate our website into Chinese.

Assuming that the pilot succeeds, there are already plans to expand it. We are talking to the Chinese Academy of Sciences about how it could build its own robotic telescopes. The Beijing Planetarium, Tsinghua University and the South China University of Technology in Guangzhou are already working on ways to give all Chinese students access to them.

There is interest elsewhere, too. Ireland ran a very successful pilot programme last year through University College Dublin and is now looking at how to roll it out across all secondary schools in Ireland. A pilot in Opole, Poland, organized through the University of Warsaw's physics department has the same objectives. All these international developments are driven by an aspiration to build technology-driven knowledge economies with high-paying employment.

The skills of innovation and creativity developed with practical science are the bedrock of a knowledge economy, and the Chinese and others using UK technology to boost their competitiveness must be looking at England's plans to drop practical science at A level as rather strange.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/csqq9b

John Baruch *is a senior lecturer at the University of Bradford and visiting professor at South China University of Technology in Guangzhou.*
e-mail: john@telescope.org

RESEARCH HIGHLIGHTS

Selections from the scientific literature

ORGANIC CHEMISTRY

Catalyst eases fuel production

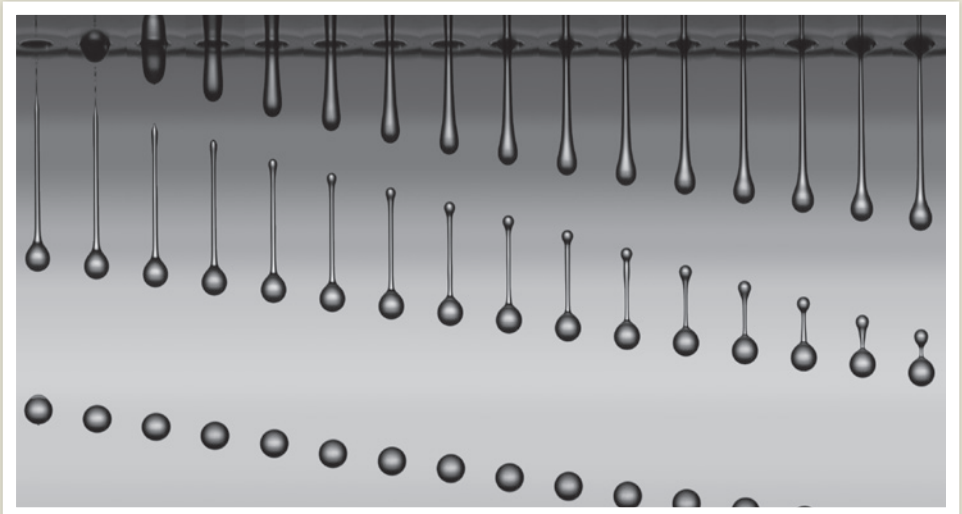
A catalyst could improve the manufacture of methanol, a promising fuel for renewable energy, from carbon dioxide. Current methods require high pressures or generate carbon monoxide, an undesirable by-product.

Jens Nørskov at Stanford University in California and his colleagues modelled the chemical reduction of CO_2 to methanol at ambient pressure and identified nickel–gallium-based compounds as promising catalysts.

The researchers synthesized and tested a series of these catalysts, and found that Ni_5Ga_3 produced the same or larger amounts of methanol compared with conventional catalysts, while also generating less CO, all at ambient pressure.

This catalyst could be used to make methanol as a fuel in, for instance, fuel cells, the authors say.

Nature Chem. **http://doi.org/rss (2014)**



MARK-JAN VAN DER MEULEN/ARJAN VAN DER BOS

PHYSICS

Fast imaging captures falling droplets

Researchers have obtained images of tiny oil droplets (pictured) forming in flight at high speeds. Such data on droplet formation could improve inkjet printing techniques.

Studying how droplets behave in inkjet printing is difficult because they move so rapidly. Detlef Lohse at the University of Twente in the Netherlands and his colleagues used 8-nanosecond-long flashes of a laser to light up picolitre-sized silicone oil droplets, and recorded these with a microscope and high-speed camera. By comparing two images of the same droplets taken 600 nanoseconds apart, the authors calculated the internal flow rate of the droplets as they formed, and found good agreement with mathematical simulations.

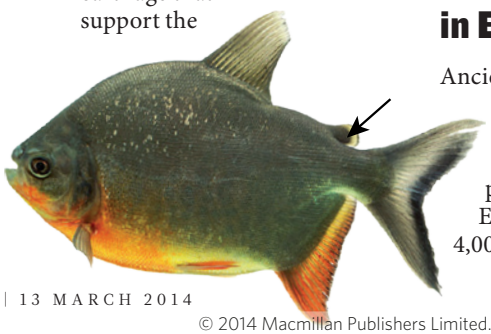
Phys. Rev. Applied **1, 014004 (2014)**

EVOLUTIONARY BIOLOGY

How the fish got its fins

The adipose fin, which sits between the dorsal fin and the tail on many fishes, might have evolved separately in different fish lineages rather than once from a single ancestor. This suggests that the fin (**pictured** with arrow) has an adaptive purpose and can evolve into various forms, contrary to previous thinking.

Thomas Stewart at the University of Chicago, Illinois, and his colleagues reconstructed the evolutionary relationships of 232 fishes, looking at the presence and absence of adipose fins. They also studied the skeletons of 620 fish species from 55 families. The team concludes that these fins have a wide variety of skeletal structures and have repeatedly evolved some of the same features, such as fin rays — rods of bone or cartilage that support the fin membrane.



Although the purpose of these fins is not clear, they could be a powerful tool for studying vertebrate limb evolution, the authors suggest.

Proc. R. Soc. B **281, 20133120 (2014)**

ARCHAEOLOGY

Cats tamed early in Egypt

Ancient Egyptians might have domesticated wild cats nearly 2,000 years earlier than previously thought. Egyptian artwork from 4,000 years ago depicts domesticated cats alongside humans. But in 2008, Wim Van Neer at the Royal Belgian Institute of Natural Sciences in Brussels and his colleagues discovered six cat skeletons buried in a cemetery for elite Egyptians that dates to the fourth millennium BC. The teeth and bones resemble those of modern domestic felines. The cats — two pairs of kittens, and an older female and male — seem to have been born outside the breeding season of wild cats, suggesting that humans had a role in rearing them, the researchers say.

J. Arch. Sci. **http://doi.org/rsg (2014)**

AMMIT/ALAMY

CONSERVATION BIOLOGY

Extinction looms for many mammals

Nearly one-quarter of the world's carnivores and hoofed mammals have moved closer to extinction since the 1970s.

Moreno Di Marco at Sapienza University of Rome and his colleagues looked at the conservation statuses of about 500 species of carnivores and ungulates over the past 40 years. The researchers found that for every species that saw improvements in status, eight deteriorated. Large animals are also sliding towards extinction faster than their smaller counterparts, and the sharpest declines in conservation status were seen in southeast Asian species.

The authors attribute these shifts to factors such as changes in international trade regulations, hunting, habitat loss and geopolitical events such as the collapse of the Soviet Union, which resulted in the loss of protected areas.

Conserv. Biol. **http://doi.org/rrs (2014)**

ECOLOGY

Warmer climate disturbs food web

A study of seabird feathers has revealed how climate change is shifting the food web in the Indian Ocean.

Alexander Bond at the University of Saskatchewan in Saskatoon, Canada, and Jennifer Lavers at the University of Tasmania in Hobart, Australia, inferred the diets of flesh-footed shearwaters (*Puffinus carneipes*) by looking at ratios of carbon and nitrogen isotopes in the birds' feathers that were collected between 1936 and 2011.

The duo found that levels of heavy isotopes — which are present at higher concentrations in species further up the food chain — fell in shearwater feathers over the years, hinting that the birds are eating animals that are lower on the food chain. This could be due to a lack of large fish caused by fishing. Furthermore, the length of the shearwaters' food chain could be shortening because of reduced nutrient flow to the Indian Ocean, owing to a warming climate that is weakening the Leeuwin Current near the western coast of Australia.

Glob. Change Biol. **http://doi.org/rrp (2014)**

CANCER

High cholesterol in prostate tumours

Prostate cancer could one day be treated by altering the cancer cells' abnormal cholesterol metabolism.

Ji-Xin Cheng at Purdue University in West Lafayette, Indiana, and his colleagues used Raman spectromicroscopy to analyse lipids inside single cells in tissue samples from people with prostate cancer. The team found that a cholesterol derivative, cholesteryl ester, accumulates inside the most aggressive of the cancer cells, but not in normal prostate cells. This build-up occurs because of the loss of PTEN, a tumour suppressor linked to many cancers.

Treating tumour-bearing mice with small molecules that block the accumulation of cholesteryl ester shrank the tumours and slowed their growth.

Cell Metabol. 19, **393–406 (2014)**

OPTICS

Hot air guides laser beams

A channel of hot air could enable high-power laser beams to travel through the atmosphere over long distances — which might be useful for applications such as communications.

High-power beams cannot be precisely focused over many kilometres, because the surrounding air absorbs and distorts the radiation. To solve this problem, Howard Milchberg and his colleagues at the University of Maryland in College Park created a conduit in air for the laser. They fired a square-shaped array of four intense, low-power light bursts. The dissipating shots left a channel of hot, dense air, through which the researchers then fired a powerful, focused laser beam. The team showed that the air channel guided the beam through 70 centimetres of air, and calculated that this waveguide could work over longer distances.

Phys. Rev. X 4, **011027 (2014)**

ENGINEERING

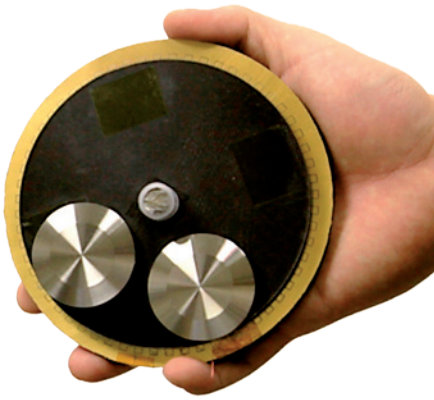
Shake to make power

A device that generates electricity through contact and friction might one day be used to harvest the energy from human motion to charge portable electronics.

Zhong Lin Wang and his colleagues at the Georgia Institute of Technology in Atlanta designed a compact, light-weight generator (**pictured**), consisting of a copper-plated disk that spins and rubs against a static base containing a layer of electrodes and a conducting surface. The device can harvest mechanical energy from gentle wind, tap-water flow and normal body movements.

The technology could be developed for large-scale power generation, the authors say.

Nature Commun. 5, **3426 (2014)**



GUANG ZHU/ZHONG LIN WANG/BEIJING INST. NANOENERGY NANOSYSTEMS/GEORGIA TECH

COMMUNITY CHOICE

The most viewed papers in science

NEUROSCIENCE

Why warm caresses feel so good

HIGHLY READ on www.jneurosci.org in February

Nerve fibres in human skin that are sensitive to gentle touch are specially tuned to respond to slow, skin-temperature strokes.

Rochelle Ackerley at the University of Gothenburg, Sweden, and her colleagues used a robotic probe to stroke the forearms of volunteers at different speeds and temperatures. In one experiment, the researchers recorded the electrical responses of the nerves, called C-tactile fibres, in the skin of 18 participants. In another, they assessed how pleasurable 30 different participants considered each stroke.

The C-tactile fibres fired more frequently, and participants reported more pleasure, when strokes were applied slowly and the probe was close to typical skin temperature. The findings suggest that the fibres have a role in evolutionarily important social interactions that rely on touch, such as in romantic relationships or when nurturing a baby.

J. Neuro. 34, **2879–2883 (2014)**

NATURE.COM

For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

RESEARCH

Infant HIV cure

A second HIV-infected child seems to have been successfully treated, according to the doctor who last year reported the first child to be cured of the infection (see *Nature* **504,** 357–365; 2013). Deborah Persaud, who studies paediatric infectious diseases at Johns Hopkins Children's Center in Baltimore, Maryland, presented the results on 5 March at the annual Conference on Retroviruses and Opportunistic Infections in Boston, Massachusetts. She said that a girl born last April to an HIV-infected mother was treated with retroviral drugs within four hours of birth, and that tests suggest she has been cured. The child continues to receive anti-HIV medication.

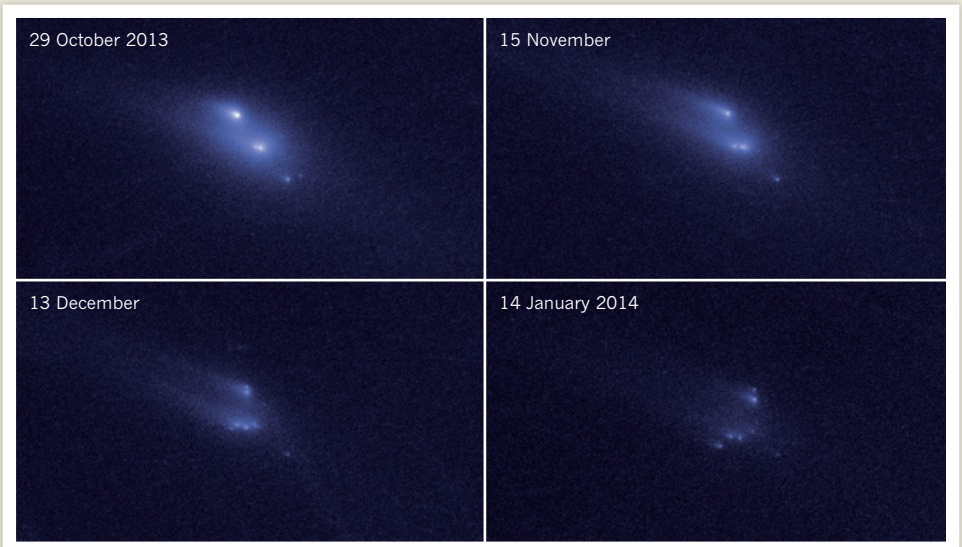
Retraction call

Two prominent research papers that describe a method for reprogramming mature cells into an embryonic state should be retracted, a co-author of the papers said in media reports on 10 March. The studies, published in *Nature* (H. Obokata *et al. Nature* **505,** 641–647, 676–680; 2014), apparently contain images duplicated from the doctoral dissertation of the lead author, Haruko Obokata of the RIKEN Center for Developmental Biology in Kobe, Japan. "I have lost faith in the paper," said Obokata's co-author, Teruhiko Wakayama, a mouse-cloning expert at the University of Yamanashi. See go.nature.com/5gpzog for more.

NUMBER CRUNCH

€182.5 m

The total amount (US$253 million) that Italian regulators fined the Switzerland-based drug companies Novartis and Roche for colluding to block the use of an eye treatment in order to promote a more-expensive alternative drug that they jointly market.



D. JEWITT (UCLA)/ESA/NASA

Asteroid caught in the act of falling apart

For the first time, the Hubble Space Telescope has captured an asteroid in the process of breaking into pieces. A series of close-up images taken over several months (pictured) revealed that a fuzzy object first spotted in September 2013 is actually a set of ten rocky fragments that are slowly drifting apart. A team led by astronomer David Jewitt at the University of California, Los Angeles, describes the findings this month in *Astrophysical Journal Letters* (D. Jewitt *et al. Astrophys. J.* **784,** L8; 2014).

BUSINESS

Live longer

Genomics pioneer Craig Venter hopes to discover how to keep ageing adults fit and healthy for longer with the launch of a new company, Human Longevity, announced on 4 March. Based in San Diego, California, the business is a joint venture with Robert Hariri, chief executive of the stem-cell company Celgene Cellular Therapeutics in Summit, New Jersey, and the X Prize Foundation founder Peter Diamandis. It will sequence the genomes of cancer patients and their tumours, and the genomes of 40,000 people per year from all age groups, with the aim of increasing this to 100,000 per year, to build the largest database yet of human genetic information.

POLICY

Budget cuts bite

The US National Institutes of Health awarded 750 fewer new research grants in 2013 compared with 2012, an 8.3% drop, as a result of the sweeping budget cuts known as sequestration that hit government agencies last year. The finding is part of a report published on 6 March by the US Government Accountability Office, which notes that it will be years before the full effects of the cuts are known. The 2013 sequestration also hit the US National Science Foundation, which awarded 690 fewer grants. See pages 139 and 147 for more.

China congress

China's Premier Li Keqiang declared a war on pollution in his opening speech to the annual National People's Congress on 5 March. At the meeting, attended by the country's leading lawmakers and politicians, Li said that green measures will target outdated energy production

and industrial processes. He also gave the nation's researchers a boost, pledging cash for basic science. See page 148 for more.

PEOPLE

ACS head retires

The world's largest scientific society, the American Chemical Society (ACS), will lose its chief, Madeleine Jacobs, when she retires at the end of this year. Jacobs has spent more than 24 years with the society, including 11 years as its chief executive. She came to the ACS after spending 21 years in science journalism and public affairs, and during her tenure oversaw a boom in fund-raising activity that netted the society more than US$500 million per year.

Smithsonian head

David Skorton, the president of Cornell University in Ithaca, New York, will be the new head of the US Smithsonian Institution from July 2015. Skorton, a cardiologist who was formerly president of the University of Iowa in Iowa City, will replace Wayne Clough, the Smithsonian's Board of Regents announced on 10 March. The institution, a collection of museums and research complexes headquartered in Washington DC, has an annual budget of US$1.3 billion.

NOAA

New NOAA chief

The US Senate approved Kathryn Sullivan to lead the National Oceanic and Atmospheric Administration (NOAA) on 6 March. Sullivan (**pictured**) is a former NASA astronaut and the first US woman to walk in space. She replaces marine ecologist Jane Lubchenco, who resigned in February 2013. Sullivan returned to the agency in 2011 as its deputy administrator after serving as its chief scientist in the 1990s.

Forensic fraud

Disgraced forensic chemist Annie Dookhan, who worked at the Hinton State Laboratory Institute drug lab in Jamaica Plain, Massachusetts, until 2012, acted alone when she falsified data and tampered with drug samples. So concludes Glenn Cunha, Massachusetts' Inspector General, in a report dated 4 March. However, the report adds that poor lab management and training, and weak security enabled the crimes. Dookhan admitted to the fraud in 2012 and is currently serving a prison sentence. See go.nature.com/oohole for more.

EVENTS

El Niño cometh?

The eastern equatorial Pacific Ocean might shift into a warm phase, known as El Niño, in the next few months, the US National Oceanic and Atmospheric Administration (NOAA) said on 6 March. The phenomenon disrupts weather patterns around the globe and could become more frequent as a result of global warming. NOAA says that there is a 50% chance of an El Niño developing during the Northern Hemisphere's summer or autumn, but that the accuracy of forecasts will improve over the next two months.

Coal-mine fine

One of the largest coal companies in the United States has agreed to spend around US$200 million to reduce water pollution and monitor the environment at mines in the Appalachian Mountain region, the US Environmental Protection Agency announced on 5 March. Alpha Natural Resources, based in Bristol, Virginia, and its subsidiaries will also pay fines totalling $27.5 million for thousands of violations of the Clean Water Act. The settlement covers 79 mines and 25 processing plants in 5 states, including West Virginia.

Radiation contained

No radioactive contamination or air-quality problems were detected by the first probes lowered into the Waste Isolation Pilot Plant near Carlsbad, New Mexico, the US Department of Energy announced on 9 March. One section of the nuclear-waste facility suffered a leak last month (see *Nature* http://doi.org/rtw; 2014), but energy-department officials say that the plant's air-filtration system stopped the radiation reaching other parts of the facility. The results must be confirmed before employees can re-enter the plant and begin investigating the cause of the leak.

COMING UP

16–18 MARCH
A symposium entitled 'The Evolution of Modern Humans — From Bones to Genomes' takes place in Sitges, Spain. Hosted by Cell Press, the meeting will discuss multidisciplinary approaches to studying the evolution of *Homo sapiens*.
go.nature.com/63txud

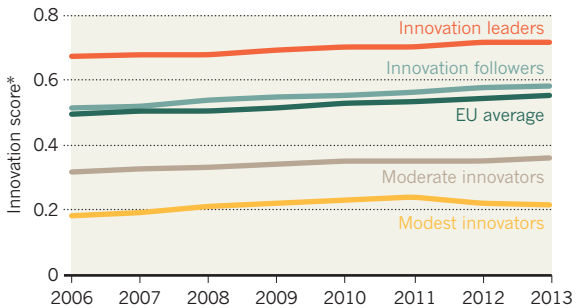
24–25 MARCH
Representatives of more than 50 nations meet at the third Nuclear Safety Summit in The Hague, the Netherlands. They will discuss ways to reduce the amount of nuclear material in the world and tackle smuggling.
go.nature.com/pdiaam

TREND WATCH

The European Union (EU) is getting more innovative, but its member states remain divided into leaders and laggards, suggests the European Commission's *Innovation Union Scoreboard 2014* (see chart). The worst countries barely improved, as measured by a set of metrics covering research systems, papers, patents, entrepreneurship and innovative firms. At the sub-national level, performance got worse in one-fifth of EU regions. Europe closed the gap between itself and the United States and Japan, however.

SPLIT STREAMS IN EUROPE'S INNOVATION

There are stark differences in innovation performance in EU member states (clustered into four groups by the European Commission).



*Based on a set of 25 indicators.
Innovation leaders: Sweden, Denmark, Germany, Finland. **Innovation followers:** Austria, Belgium, Cyprus, Estonia, France, Ireland, Luxembourg, the Netherlands, Slovenia, UK. **Moderate innovators:** Croatia, Czech Republic, Greece, Hungary, Italy, Lithuania, Malta, Poland, Portugal, Slovakia, Spain. **Modest innovators:** Bulgaria, Latvia, Romania.

SOURCE: *INNOVATION UNION SCOREBOARD 2014*

NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

JONATHAN ERNST/REUTERS/CORBIS

President Barack Obama's 2015 budget has disappointed.

POLICY

Obama's budget request falls flat

Hopes dim for a science-funding increase in 2015.

BY LAUREN MORELLO, JESSICA MORRISON, SARA REARDON, JEFF TOLLEFSON AND ALEXANDRA WITZE

What a difference a year makes. In April 2013, US President Barack Obama unveiled a budget proposal that set out plans to map the human brain and to capture an asteroid that could be towed near to the Moon for astronauts to study. But that sense of ambition is absent from Obama's plan for fiscal year 2015, which was released on 4 March, frustrating scientists and policy analysts who hoped for a show of support from the White House after years of pressure to reduce government spending.

The US$3.9-trillion proposal would keep budgets flat or nearly so at major science agencies, including the National Institutes of Health (NIH), the National Science Foundation (NSF) and NASA. Overall, it includes just $65.9 billion for non-defence research and development, a paltry 0.7% above the current level.

The proposal seeks an extra $5.3-billion boost for science from a new initiative that is separate from normal agency budgets, to be paid for in large part by new taxes on the wealthy. But few observers expect Congress to accept that plan, dimming hopes for a significant boost for science funding in 2015. "We have to convince the legislators that they have to make some tough decisions and give more money to science," says Samuel Rankin, director of the American Mathematical Society's Washington DC office. "That's easier said than done."

At the NIH, Obama is seeking $30.2 billion in 2015, compared with this year's $30 billion (see 'Budget highlights') — a rise of less than 1%. That is not enough to outpace the expected rise in research costs in 2015, which according to the agency's own Biomedical Research and Development Price Index will be 2.2%, higher than the US inflation rate. One of the few bright spots in the NIH plan is continued support for the Brain Research through Advancing Innovative Neurotechnologies Initiative. It would receive $100 million in 2015, up from $40 million this year.

"Scientists are constitutionally not comfortable with the status quo, and that's what this is — not only the money but the idea," says Howard Garrison, deputy executive director for policy at the Federation of American Societies for Experimental Biology in Washington DC. Stefano Bertuzzi, executive director of the American Society of Cell Biology in Bethesda, Maryland, is blunter. "A 1% bump is not going to cut it," he says.

The NSF would see its budget creep up to $7.3 billion, roughly 1% above the current level. Of the agency's seven research directorates, only Social, Behavioural and Economic Sciences (SBE) would receive a noticeable boost, rising by 6% to $272 million. Most of that is earmarked for the National Center for Science and Engineering Statistics (NCSES), which tracks trends in scientific research, education and workforce development. Rick Wilson, a political scientist at Rice University in Houston, Texas, would have liked to see more SBE money made available to individual researchers, but says that the NCSES data are "really useful" for social scientists.

At NASA, the proposed $17.5-billion total budget pushes forward the administration's plans to capture and study an asteroid, despite lawmakers' scepticism about the mission's value. Asteroid experts at NASA's Jet Propulsion Laboratory in Pasadena, California, have been hunting for candidate targets, while astronauts at the Johnson Space Center in Houston, Texas, are designing equipment for a close-up encounter with a space rock. NASA has asked for $133 million — an increase of $55 million over this year — for the programme, which NASA chief Charles Bolden has described as a way to test new exploration capabilities aimed ▶

BUDGET HIGHLIGHTS
How science agencies fared in the budget (US$ millions).

Agency	2013 actual*	2014 actual	2015 request
National Institutes of Health	29,001	30,003	30,203
Centers for Disease Control and Prevention	5,503	5,882	5,474
Food and Drug Administration	2,386	2,640†	2,584
National Science Foundation	7,105	7,142	7,255
NASA (science)	4,782	5,151	4,972
Department of Energy Office of Science	4,681	5,066	5,111
Environmental Protection Agency	7,918	8,200	7,890
National Oceanic and Atmospheric Administration	4,906	5,322	5,497
US Geological Survey	1,012	1,032	1,073

*2013 figures include the roughly 5% across-the-board cut of the budget sequester. †Includes one-time transfer of $79 million in user fees.

SOURCE: WHITE HOUSE OFFICE OF MANAGEMENT AND BUDGET

at eventually sending astronauts to Mars. If Congress approves the budget proposal, the agency will select the basic outlines of the mission plan early next year.

Obama's budget request also includes $15 million for a spacecraft to Europa, an icy moon of Jupiter. The move is an apparent nod to Congress, which last year inserted $80 million into the NASA budget explicitly for Europa, despite the administration's objections. At times, NASA has sounded less than enthusiastic about this moon; Beth Robinson, the agency's chief financial officer, told reporters last week that Europa was a challenging environment and that she did not foresee a mission to the moon launching before the mid-2020s.

Overall, NASA's science budget would drop nearly 3.5% under the president's proposal, to $4.97 billion. The biggest casualty would be the Stratospheric Observatory for Infrared Astronomy, a joint US–German project involving a 2.5-metre airborne telescope. NASA says that if other partners cannot supply most of the plane's operating budget, it will ground the observatory from 1 October (see *Nature* http://doi.org/rtk; 2014).

Also facing a cut is high-energy physics research at the Department of Energy (DOE), which would fall by 6.8% to $744 million. That has sparked concerns that the US physics community has been too slow to unite behind a viable research agenda. With unfortunate timing, a DOE advisory panel charged with producing such an agenda is scheduled to release its results two months from now. Under consideration are proposals for a new domestic neutrino experiment and US participation in a linear collider, likely to be destined for Japan, that would build on advances at the Large Hadron Collider in Geneva, Switzerland.

"The White House is certainly sending a signal to the high-energy physics community that it needs to get its act together," says Michael Lubell, director of public affairs for the American Physical Society in College Park, Maryland.

To Andrew Lankford, a physicist at the University of California, Irvine, who leads the DOE's High Energy Physics Advisory Panel, the move is not surprising, given Obama's emphasis on climate and clean-energy research and development at the department — these saw a significant boost in the White House budget proposal. Lankford says that the proposal would make it "a challenge to maintain the vitality of our research community" — but he is confident that his panel's report will be completed in time to influence budget negotiations in Congress. ■ SEE EDITORIAL P.139

BUDGETS

China goes back to basics on research funding

Core science gets budget boost in a bid to change research culture and increase innovation.

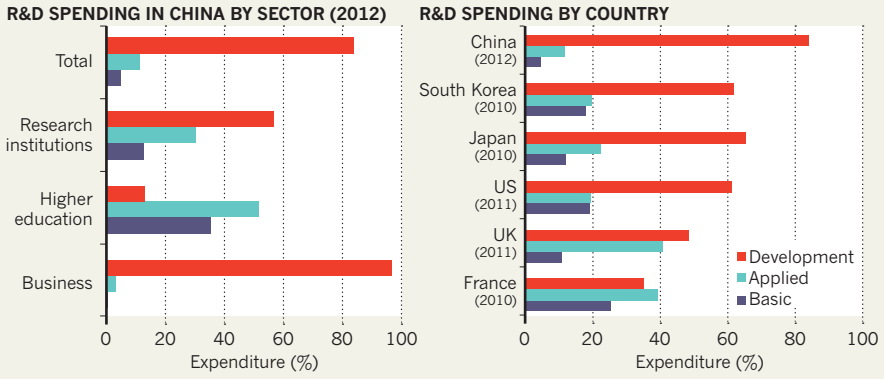
BY JANE QIU

Last week, Chinese science saw some big wins as Premier Li Keqiang delivered his first budget since taking office a year ago. Yet observers have warned that to translate that support into innovation, the country must invest more in basic research and move away from its desire for quick successes.

China's total expenditure on research and development (R&D) has increased by 23% a year on average over the past decade. But with uncertainties arising from a new government and the effects of the economic slowdown, scientists had feared cutbacks this year.

At the opening session of the annual National People's Congress in Beijing, however, Li reassured the research community by stressing

CASH DRAW
In most sectors, China spends more on developing existing technologies than on basic or applied science. The difference is more pronounced than in some other countries.



SOURCES: PRC STATISTICAL YEARBOOK 2013/ US NSF *SCI. AND ENG. INDICATORS 2014*

the importance of scientific innovation to economic growth — and by pledging hard cash.

The central government's expenditure on science and technology this year was set at US$43.6 billion (267.4 billion yuan renminbi), an 8.9% rise on last year, which slightly trails the overall projected budget increase of 9.3%. The biggest winners are 16 'megaprojects' with an emphasis on engineering and applied research in areas such as transgenic crops, nuclear power plants and lunar exploration, which together will receive a whopping $8.1 billion.

China's basic-research spending has historically been extremely low — about 4.8% in 2012 and 2013, compared with 10–25% in developed nations (see 'Cash draw'). But this year, the appropriation for basic research will increase by 12.5% to $6.6 billion — of which the National Natural Science Foundation of China is slated to get $3.1 billion, says its president, Yang Wei. The major areas that the foundation will fund include studies of biodiversity, air pollution, supercomputers, neurodegenerative diseases and scientific equipment.

Two of the 16 megaprojects have a substantial basic-research component: these are in the areas of drug discovery and major infectious diseases, including HIV/AIDS and influenza. And with a combined budget of $488 million, the two initiatives "will continue to strengthen the capacity for drug screening, rapid detection of pathogens and vaccine development", says Liu Qian, deputy director of the National Health and Family Planning Commission, who oversees the projects.

The Ministry of Science and Technology will spend about 8% of its total budget of $8.1 billion on basic research — including $211 million on six major science programmes in areas such as nanotechnology, quantum physics, stem cells and protein science — and $1.1 billion on developing key technologies.

The Chinese Academy of Sciences (CAS), which relies mostly on extramural grants to fuel its research, will receive $423 million for 20 'strategic priority projects' in areas ranging from neuroscience to studies of the Tibetan Plateau.

Although such strong support is welcomed, observers such as Richard Suttmeier, a policy researcher at the University of Oregon in Eugene who advises China's science ministry, fear that more money may not bring more innovation without a sea change in how it is spent, and without a shift in China's research culture and institutions. This is because of serious concerns about the country's quality of research.

"While China's output of publications and patents is impressive, there are very few genuine innovations," says CAS president Bai Chunli. A key reason, says Yang, is the modest government support for basic research. Moreover, he adds, "a large chunk of China's R&D expenditure comes from industry". In 2012, industry contributed 76% of all R&D funding but spent almost nothing on basic science and only 3% on applied research, he says (see 'Cash draw'). Consequently, 84% of China's total R&D spending goes on product development, such as the commercialization of technologies, compared with 35–65% in developed economies.

Even research institutes spend only 13% of their funding on basic science and less than one-third on applied research, and are under increasing pressure to engage in development projects that focus on real-world problems. Without a firm footing in solid research, such projects "have limited value in developing an innovative economy", says Su Jun, a policy researcher at Tsinghua University in Beijing.

Chinese scientists are also concerned by the serious misuse of research grants. A report released by China's National Audit Office last October shows that the problem is widespread — and that up to half of all research funding has been misused. In a crackdown, one vice-president of Zhejiang University was arrested last December on "suspicion of occupational crime", according to the Xinhua news agency, and 50 officials at the Guangdong Provincial Department of Science and Technology are under investigation for embezzling R&D funds.

When research grants are used as intended, Chinese science "suffers from excessive bureaucratic interference and a culture of *jigong jinli* — seeking quick success and short-term gain", says Muming Poo, director of the CAS Institute of Neuroscience in Shanghai. Officials often demand the demonstration of productivity on an almost yearly basis, and grants can be slashed by 50% if researchers fail in this task.

"The Chinese government is aware of the problems and institutional reforms are firmly on the agenda of the congress to boost R&D efficiency," says Bai. Topics under discussion at the congress included increasing the proportion of basic research in total R&D to 10% by 2020; instigating policies to encourage investment in basic and applied research; reforming funding and evaluation systems; raising the budget for research overheads; and giving scientists more freedom. Such changes "will be crucial for Chinese science to reach the next level", says Bai. ■

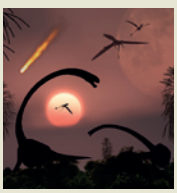
LIU JIANSHENG/XINHUA/PHOTOSHOT

Premier Li Keqiang pledged support for basic science at last week's National People's Congress in Beijing.

MORE ONLINE

MARK STEVENSON/STOCKTREK IMAGES/CORBIS

TOP NEWS



Sun's passages through dark-matter disk could trigger periodic extinctions go.nature.com/x6fx77

MORE NEWS

- A potential blood test for early Alzheimer's detection go.nature.com/nmaime
- Beautiful maths is like art to the brain go.nature.com/rtjypj
- River-basin geometry slowly shifts before it settles go.nature.com/5ytp2y

NATURE PODCAST

A medieval multiverse theory; a rare mineral from Earth's mantle; and wolves in Yellowstone nature.com/nature/podcast

The WHO recommends that adults have less sugar per day than is found in one glass of many soft drinks.

ALEPH STUDIO/SHUTTERSTOCK

NUTRITION

Storm brewing over WHO sugar proposal

Industry backlash expected over suggested cut in intake.

BY BRIAN OWENS

Scientists are gearing up for a battle with the food industry after the World Health Organization (WHO) moved to halve its recommendation on sugar intake.

Nutrition researchers fear a backlash similar to that seen in 2003, when the WHO released its current guidelines stating that no more than 10% of an adult's daily calories should come from 'free' sugars. That covers those added to food, as well as natural sugars in honey, syrups and fruit juice. In 2003, the US Sugar Association, a powerful food-industry lobby group based in Washington DC, pressed the US government to withdraw funding for the WHO if the organization did not modify its recommendations. The WHO did not back down, and has now mooted cutting the level to 5%.

"These are reasonable limits," says Walter Willett, head of nutrition at the Harvard School of Public Health in Boston, Massachusetts. "Five per cent of calories is just a bit less than in a typical serving of soda, and we have good evidence of increased risk of diabetes with that intake, which of course increases with greater intake."

But Marion Nestle, a nutrition researcher at New York University, predicts that grocery manufacturers are not going to take the proposal lying down. "If people follow this advice, that would be very bad for business," she says.

The WHO made its recommendations in draft guidelines that were released for public consultation on 5 March. In halving the 10% figure, it cited the need to fight obesity — worldwide incidence reached 11% in 2008 — and to prevent tooth decay. Five per cent of daily calories is equivalent to about 25 grams, or 6 teaspoons, of sugar. Many people around the world consume more than that — young adults in the United States, for example, get more than 14% of their calories from free sugars, according to the US Centers for Disease Control and Prevention in Atlanta, Georgia (see 'Sugar high').

The guidelines are based on a careful analysis of more than 120 scientific studies, summarized in two meta-analyses commissioned by the WHO (L. Te Morenga *et al. Br. Med. J.* **346,** e7492; 2013; P. J. Moynihan and S. A. M. Kelly *J. Dent. Res.* **93,** 8–18; 2014).

Jim Mann, a nutrition researcher at the University of Otago in New Zealand, who worked on one of the meta-analyses and helped to develop the guidelines, says that the science supporting a drop in sugar consumption has become more conclusive since 2003. But the biggest difference is in the process the WHO uses to produce its recommendations.

For the first time in the production of nutrition guidelines, the agency adopted the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system, a more formal, standardized approach to developing guidelines compared with a literature review. It requires a clear statement of the research question, uses the gold-standard methodology for literature review and meta-analysis developed by the Cochrane Collaboration — a non-profit group headquartered in Oxford, UK, dedicated to the systematic analysis of medical research — and weighs up biases and confounding factors before an expert committee develops recommendations. This painstaking process "doesn't give much leeway for opinions", says Mann.

His analysis showed a strong confirmation of the benefits of the 10% limit, especially for preventing tooth decay, with "good clues" that it would be worth going lower, although the evidence for that is weaker.

This is likely to form the focus of the sugar lobby's attacks, researchers say. Most industry groups have refused to comment until they have prepared submissions to the consultation, but a few have already criticized the lower limit. The US Sugar Association, for example, released a statement pointing out that the US Institute of Medicine and the European Food Safety Authority have said in the past (in 2005 and 2010, respectively) that there was no conclusive evidence to justify such a limit on free sugars.

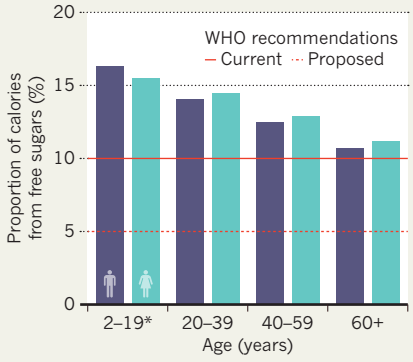
Industry submissions to the consultation are likely to be forceful. When the WHO recommended the 10% limit, it faced a ferocious attack on the credibility of its science from several camps — including the administration of then US President George W. Bush. The administration said that the WHO report did not meet US data-quality standards, was not properly peer-reviewed, and failed to separate scientific and policy recommendations.

Nestle thinks that if the WHO is willing to face that kind of pressure again, it must have confidence not only in its science, but also in the political climate. "There's so much evidence now that says that people would be healthier if they ate less sugar, it may be that things have changed," she says.

This time around, the WHO is taking steps to counter excessive lobbying. Anyone who wishes to submit a comment on the draft guidelines must first complete a declaration-of-interest form. And the organization says that it will stand firm against any push-back from the food industry. "If pressure comes to the organization, then we're very well equipped to resist that type of pressure," said Francesco Branca, director of the WHO's Department for Nutrition for Health and Development, at a press conference. ■

SUGAR HIGH

The latest figures show that average consumption of sugar by males and females in the United States exceeds that recommended by the WHO.



*Data for 2–19 from 2005–08; other groups from 2005–10.

SOURCE: CDC/NCHS, NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEYS

GEOSCIENCE

Global seismic network takes to the seas

Two systems could plug the ocean-sized gap in earthquake detection.

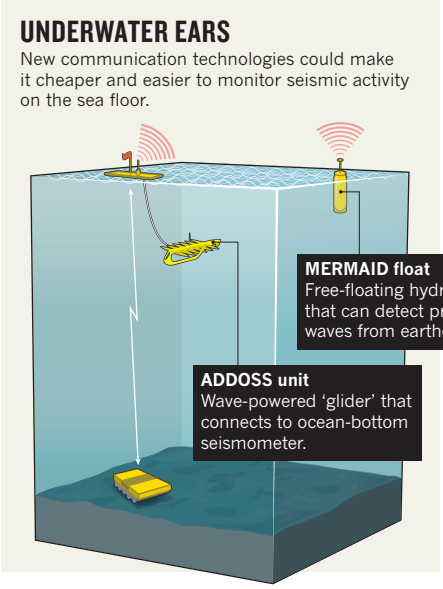
BY NICOLA JONES

More than 25 years after its inception, there is hope that the Global Seismographic Network (GSN) will finally live up to its name. The network's 150 or so seismic stations listen for signs of earthquakes and nuclear tests, and help geophysicists to image Earth's interior, but their scope is limited: they are all located on land.

The GSN's vast marine blind spot could soon be eliminated, however, thanks to new, relatively inexpensive equipment that will be field-tested between April and June. For John Orcutt, a geophysicist at the Scripps Institution of Oceanography in La Jolla, California, the prospect of truly global measurements is tantalizing. "Working out how the interior of the planet works is really hard when all your sensors are only on 30% of the planet's surface," he says.

The GSN, which is run jointly by the Incorporated Research Institutions for Seismology in Washington DC and the US Geological Survey (USGS), was originally intended to blanket the globe with sensors. But installing permanent seismic monitors on the sea floor proved to be too expensive. Thousands of metres of cable are needed to connect the monitors to surface buoys that transmit data in real time, and the bulky equipment must be deployed from costly research ships. "It's the ship time that kills you," says Jonathan Berger, a geophysicist at Scripps who has been involved with the GSN since its inception. Placing and maintaining 2,250 seafloor stations, spaced roughly 400 kilometres apart, would cost between US$700 million and $1 billion over five years, says Guust Nolet, a geophysicist at the University of Nice Sophia Antipolis in France.

Faced with such a steep price tag, researchers have made do with half measures. At any given time, there are a few hundred seismic stations temporarily deployed on the ocean floor, storing data until they can be picked up by ship — usually once a year. The largest number is overseen by the Ocean Bottom Seismograph Instrument Pool (OBSIP), which is funded by the US National Science Foundation (NSF). These temporary stations are useful for retrospective analyses, such as tracing how earthquakes echoed through the Earth and helping to calculate the location of molten plumes inside the mantle. But they cannot be used for anything that requires real-time data, such as earthquake monitoring.



For that purpose, a few nations, including Japan and Canada, have installed expensive wired arrays of offshore seismic stations that receive power and send data along fibre-optic cables. The United States will soon install its own array as part of the NSF's Ocean Observatories Initiative (see *Nature* **501,** 480–482; 2013). But for global monitoring, more practical and affordable options are now surfacing.

Berger recently began extended field trials of his Autonomously Deployed Deep-Ocean Seismic System (ADDOSS), which uses 'gliders' that convert wave motion into thrust. Comprised of a submerged portion and a surfboard-sized surface float equipped with solar panels and a satellite positioning system, the gliders are able to wirelessly retrieve data from seismometers on the ocean floor (see 'Underwater ears'). Built by Liquid Robotics of Sunnyvale, California, they are light enough to be installed and maintained by regular ships rather than specialized research vessels. If they experience a problem, the gliders can be programmed to 'swim' to shore. Berger says that his team has also designed, but not yet built, a sleek ocean-bottom seismometer that the gliders can tow to a research site.

"It is a technology that can enable things we have wanted to do for a long time for basic science and earthquake-hazard studies," says Thorne Lay, a seismologist at the University of California, Santa Cruz, who is not affiliated with the project. Ocean stations should be able to detect small offshore earthquakes that are missed by instruments on land, and they will yield discoveries about Earth's mantle, he says.

Berger's first long-term test of an ADDOSS station this winter was interrupted when a glider experienced problems. He will try again in May or June; if all goes well, he envisages deploying 20 such stations across the world's oceans, roughly 2,000 kilometres apart. The cost of each station — less than $200,000 — would be comparable to that of installing and maintaining one of the GSN's existing land-based seismic stations.

A parallel effort pursued by Nolet aims to provide even cheaper ocean coverage. The Mobile Earthquake Recorder in Marine Areas by Independent Divers (MERMAID) system is a set of free-floating buoys that drift with the current. Each buoy carries a hydrophone that can detect pressure waves from large or nearby earthquakes but cannot sense the motions of the sea floor. Nolet estimates that he could blanket the globe with 300 such devices for $24 million. Tests of four buoys in the Indian Ocean this winter proved that they can 'hear' through the noise of stormy seas.

In April, Nolet plans to deploy ten more devices to image the mantle plume that lies under the Galapagos Islands. Some previous efforts to map plumes have made use of OBSIP. "It's a great programme, but it's very expensive and it can't do everything," says Cecily Wolfe, a USGS seismologist based at the University of Hawaii in Honolulu who used the network to investigate the plume beneath Hawaii. Technology such as MERMAID or Berger's ADDOSS programme could one day do similar work more cheaply, and their measurements could also be combined with those of OBSIP to help researchers to recognize and filter out 'contamination' from seismic signals originating outside their survey zone, Wolfe says.

If the new technologies succeed, it will be a remarkable change for science, says Orcutt. But it will be a while before the systems reach their full potential, he adds. "We'll need a couple of decades of observation before things really start to come into focus." ■

ABENGOA BIOENERGY

The Abengoa cellulosic ethanol plant near Hugoton, Kansas, will start production this year.

ENERGY

Cellulosic ethanol fights for life

Pioneering biofuel producers hope that US government largesse will ease their way into a tough market.

BY MARK PEPLOW

On the flat plains of Kansas, a stack of gleaming steel towers and pipes stretches 16 storeys into the sky. More than 1,000 construction workers toiled to complete the ethanol plant near the town of Hugoton, and its owners expect it to join a fermented-fuel revolution.

But unlike most ethanol factories, in which yeast feeds on sugars in foodstuffs such as maize (corn) kernels, the Hugoton facility will make use of what has been, until now, agricultural waste: cellulose. Thousands of tonnes of corn stover — the leaves, stalks and husks left over after the maize harvest — are already waiting, stacked in square bales, at the 1.6-square-kilometre site. By June, the plant will begin processing the stover into ethanol, which will be blended with petrol and end up in vehicle fuel tanks.

The plant, which is owned by multinational company Abengoa of Seville, Spain, is one of three US facilities that should start commercial production of cellulosic ethanol in the next few months (the others are both in Iowa, one run by POET-DSM Advanced Biofuels and the other by DuPont). The industry has long promised that this second-generation biofuel will cut greenhouse-gas emissions, reduce US reliance on imported oil and boost rural economies. Yet just as the fuel is on the cusp of making it big, market forces and government policies could choke its progress. "This is going to be a very critical year," says Zia Haq, a chemical engineer and senior analyst at the US Department of Energy, which has helped to fund the plants. The challenges have already pushed some researchers and companies towards an alternative approach that converts cellulose into hydrocarbon fuels using chemical rather than biological processes.

"We don't have room for any more ethanol."

With more than 200 operating plants, the corn-ethanol industry is well established in the United States. Its dramatic growth has been driven by tax credits and the Renewable Fuel Standard (RFS), created by law in 2005 and extended in 2007. Administered by the US Environmental Protection Agency (EPA), the standard mandates annual increases in the volumes of various renewable fuels included in the country's fuel supply. In its early years, the law emphasized the production of corn ethanol, considered ripe for early commercialization.

Yet corn ethanol comes with problems. It offers only modest savings in greenhouse-gas emissions compared to petrol (see *Nature* **499,** 13–14; 2013). Production is vulnerable to poor harvests and can contribute to increased food prices because the maize must be grown on land that would otherwise be used for food. Tapping the storehouse of biomass left after the harvest is much less controversial. Ethanol made from corn stover produces at least 60% less greenhouse-gas emissions than petrol, and making it does not require any extra farmland.

Brewing such cellulosic ethanol, however, is hard. Producers must dismember large, indigestible molecules such as cellulose and hemicellulose to yield fermentable sugars. The process requires the biomass to be ground up and pretreated with acids. A cocktail of enzymes must then be applied to chop up the tough biological polymers inside — all before the yeast is added to the resulting sugars. Hence the scale of Abengoa's processing facility, much larger and more expensive than any corn-ethanol plant. According to the RFS, commercial production of cellulosic ethanol was meant to start around 2010, but that did not happen. With patchy investment backing, many companies have fallen by the wayside.

BLEND WALL

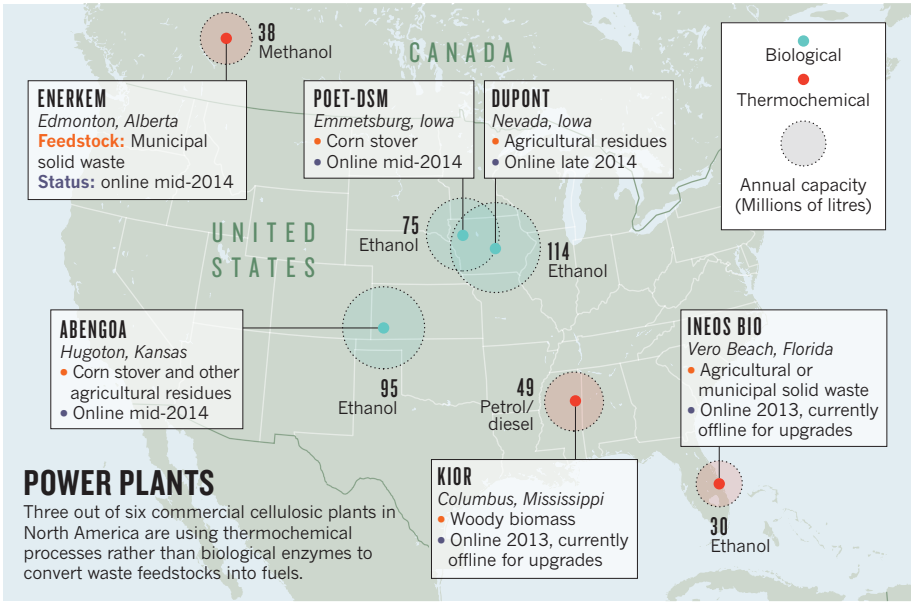
Part of the problem is that the ethanol market is already saturated. In 2012, the industry produced more than 50 billion litres of corn ethanol, comprising 10% of US transportation fuel — enough to completely satisfy demand for the E10 petrol blend that most vehicles now burn (see 'Hitting the wall'). This 'blend wall' puts an upper limit on the amount of ethanol

that the market can absorb. And it is closing in: demand for petrol has actually fallen, and there is growing interest in alternatives such as battery-powered cars. Cellulosic ethanol may now be arriving, but its timing is terrible. "We don't have room for any more ethanol," says Wallace Tyner, an agricultural economist at Purdue University in West Lafayette, Indiana.

Corn ethanol is now slightly cheaper than petrol, but cellulosic ethanol is more expensive than both. A cellulosic-ethanol plant's capital costs are roughly twice those of a corn-ethanol plant, says Tyner, and enzymes raise operational costs further. Unable to undercut its rivals, cellulosic ethanol will be heavily dependent on the RFS to clear its path to the pumps. Yet its delayed arrival has prompted the EPA to reduce the amount of cellulosic ethanol that refiners are required to blend into their petrol.

The RFS plan for this year originally called for 6.6 billion litres of cellulosic ethanol. But in November, the EPA proposed that the mandate should be reduced to 64 million litres, a mere trickle in comparison. A final ruling is expected in March or April. Groups working on renewable fuel, who say that producers will easily make more than 64 million litres once they get going, have cried out. "We think the EPA underestimated the capacity of the industry," says Christopher Standlee, executive vice-president of global affairs for Abengoa. The expensive excess ethanol might have to be sold at a loss on the open market, potentially crippling the fledgling industry.

SOURCE: CELLULOSIC BIOFUELS INDUSTRY PROGRESS REPORT 2012–2013 (AEC, 2012)



CAPACITY PROBLEM

Cellulosic-ethanol producers have several options to increase their market. First, they could break through the blend wall. All US vehicles produced in the past decade can run on a 15% ethanol–petrol blend — but consumers and distributors are mostly unconvinced, perhaps spooked by car-industry studies claiming that the fuel damages engines.

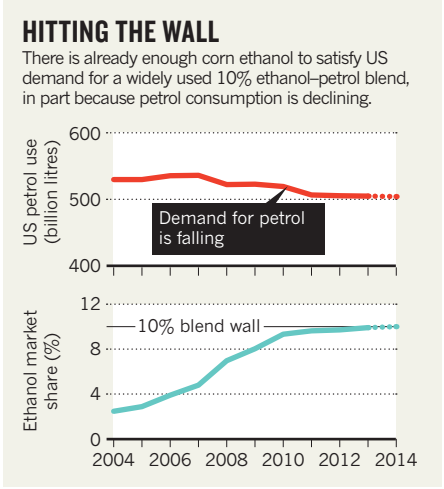
Another way over the wall might be exports to the European Union, which aims to make 10% of its transportation fuel renewable by 2020. And cellulosic ethanol could get cheaper with more efficient stover harvesting, beefier enzymes and cheaper pretreatments. The industry has already cut costs from as much as US$9 per gallon ($34 per litre) five or six years ago to close to $2 today, says Thomas Foust, director of the National Bioenergy Center, part of the National Renewable Energy Laboratory in Golden, Colorado.

But Tyner says that this approach can be squeezed only so far. He and others see more promise in a different approach to breaking up cellulose — a brute-force combination of temperature, pressure and chemistry. These thermochemical methods can produce either a crude bio-oil or a stream of carbon monoxide and hydrogen known as syngas. After further treatment and refining with the help of chemical catalysts, both can be turned into hydrocarbons such as petrol, diesel and jet fuel. Crucially, these 'drop-in' fuels, named because they can replace normal fuels with no adjustments to engines, have no blend wall to vault.

Thermochemical routes can also use lower-quality feedstocks, tearing through anything from wood chips to municipal solid waste. Enerkem, a company in Montreal, Canada, is starting its first commercial-scale plant to turn solid waste into syngas, in Edmonton. By April or May, it will be able to transform the syngas into methanol. Next year, it plans to convert the methanol into ethanol, and it says that will be cheaper than corn ethanol.

This is mostly because the feedstock is cheap. US landfill fees for solid waste are about $44 per tonne, not including transportation, so municipalities are keen to get companies such as Enerkem to take the waste off their hands, says Marie-Hélène Labrie, the company's vice-president for government affairs and communications. "Generally, we're being paid to take the feedstock."

SOURCE: EIA



Producing syngas also gives the company a lot of flexibility. If there are changes to policy mandates or the market, the system could be switched to making hydrocarbon fuels or higher-value chemical products. Enerkem plans to build similar plants in Mississippi and Quebec next year, and it is partnering with Waste Management in Houston, Texas — the biggest US landfill operator — to take its waste.

Research funding, too, is shifting to thermochemical methods, says Haq. "That doesn't mean we're abandoning cellulosic ethanol," he says. "But certainly, going forward, we're looking more seriously at hydrocarbon pathways."

Last year, an energy-department project to supply the US Navy with advanced biofuels provided funding for four facilities that will all use thermochemical methods to make drop-in fuels. Thermochemical processes are also key to the first two commercial cellulosic plants in the United States, which opened last year: KiOR in Columbus, Mississippi, and INEOS Bio near Vero Beach, Florida (see 'Power plants'). (Both plants are currently idle, pending upgrades.) Haq thinks that longer-lived catalysts will further reduce the costs of thermochemically produced cellulosic hydrocarbons in coming years.

But Standlee says that biology can still compete, by tackling ever-cheaper feedstocks. His company is betting that a new generation of enzymes can turn municipal waste into ethanol, and last July it opened a demonstration plant near Salamanca, Spain, to do just that. Abengoa hopes that this technology will eventually allow it to expand its US operations beyond the 'corn belt'.

Standlee says that, as long as the cellulosic industry is given time to mature — just as corn ethanol was — it can get back on the trajectory set out by the RFS. "If," he adds, "the EPA sticks with the programme." ■

GETTY

ISLANDS OF LIGHT

More than a billion people lack electricity, but now microgrids are powering up rural areas.

BY JEFF TOLLEFSON

In Haiti, the least-electrified country in the Western Hemisphere, some residents spend US$10 a month on candles and kerosene just to light their homes — roughly 125 times what those in the United States typically pay for the equivalent light. In India, many pay a premium to charge their mobile phones from car batteries at the local market. The Sun still dictates life for millions of Africans, and diesel generators burn through budgets on small Pacific islands. Around the world, nearly 1.3 billion people live without access to electricity, many of them far from the ever-expanding electric grid.

The quest is on to find the best way to bring clean power to rural areas. Mixing local development work with Silicon-Valley-style entrepreneurship, engineers, scientists and economists are setting up independent 'microgrids' that can be deployed quickly and cheaply one community at a time. Those leading such electrification schemes aim to create small-scale renewable-energy systems, building an archipelago of light across the developing world and helping remote communities to kick their dependence on fossil fuels.

Such efforts have often failed in the past, as subsidies lapsed or infrastructure collapsed. But today's entrepreneurs are better placed to succeed. A new generation of cheaper photovoltaic panels and wind turbines can be managed with simple smart-grid devices. The price of fossil fuels has soared over the past decade, making renewable energy more competitive. And the United Nations has set a goal of achieving universal access to electricity by 2030, providing political impetus.

"The ambition is there, and the economics are making a lot more sense now than they were a few years ago," says Richenda Van Leeuwen, executive director for energy access at the United Nations Foundation. But the challenge remains extreme. A 2012 analysis by the International Energy Agency projects that, on the basis of current plans, the percentage of people without access to electricity will fall from 19% in 2010 to 12% in 2030 — leaving nearly 1 billion people still in the dark. Achieving universal energy access would mean increasing investments from a projected $14 billion to $49 billion a year, the agency says. Centralized grids are expected to provide only about 30% of the solution in rural areas.

Among the projects already on the go are a few bright spots with lessons to teach about technologies and business models that could help to light the world.

TAMKUHA, INDIA

When a pair of young Indian entrepreneurs flipped the switch to electrify the remote agricultural village of Tamkuha in 2007, the power flowed from rice husks. Gyanesh Pandey and Ratnesh Yadav knew that photovoltaic panels were too expensive for their plans, and there wasn't a lot of wind blowing through this town of roughly 2,000 people. But their home state

Around the world, more than a billion people rely on candles or kerosene to light their homes at night.

of Bihar has rice in abundance.

Trained in electrical engineering at Rensselaer Polytechnic Institute in Troy, New York, Pandey sketched out a plan with his long-time friend Yadav. Working with a grant of roughly $12,000 from the Indian Ministry of New and Renewable Energy, the duo invested more than $40,000 of their own money to purchase and modify a gasifier to turn rice husks into biofuel, buy a 32-kilowatt generator, and run power lines through the village.

Within five months, the residents of Tamkuha had enough electricity to charge their mobile phones and fend off darkness with two compact fluorescent light bulbs per household for 6–8 hours a night. Pandey and Yadav formed Husk Power Systems with Manoj Sinha, who studied business at the University of Virginia in Charlottesville, and the company now has more than 80 mini-power plants serving some 200,000 people in India, Uganda and Tanzania.

Success in Tamkuha proved that even poor customers will pay 100 rupees ($1.60) per month or more for minimal power, in a country where rural households often survive on $15–80 a month. The rates are higher than in urban centres, but customers typically save overall because they purchase less kerosene. In 2007, says company president Sinha, nobody believed that Husk Power could create a viable business."But when we scaled up to more than 300 villages, people started believing in the model."

The opportunities in India are huge. Although the percentage of people without access to electricity in 2011 was only 25% — much lower than the 80–90% rates seen in some African countries — that still left a record 300 million people without power in a single country. The government has invested money and attention in the problem, and those efforts have slashed the number of people without a connection to the main grid by more than half in the past decade. But the country is struggling to supply enough power to feed all those lines, and to hook up the most remote communities.

Husk Power has become one of the world's largest microgrid developers. And it is dreaming big, targeting 5 million customers within five years in India and east Africa. With the cost of photovoltaic panels falling, the company is building solar microgrids and pairing them with storage batteries to meet evening demand. And it is experimenting with a solar–biomass hybrid power plant intended to provide power around the clock.

"Where Husk is going is very positive," says Daniel Kammen, an energy researcher at the University of California, Berkeley. "It is not fixating on one technology, it is fixating on solutions."

But problems may lie ahead. In some areas, small providers such as Husk Power compete with the expanding central grid, leaving some villages with two suppliers. The microgrids tend to be more reliable, but are also more expensive, because subsidies usually go into capital construction costs rather than towards keeping electricity rates low. Kammen says governments and companies should agree on some basic industry standards for regulation and finance, so that investment in microgrids — the only solution for some areas — expands rather than being undermined.

TOKELAU, SOUTH PACIFIC

The Sun shone brightly on Tokelau as a cargo ship pulled into harbour in June 2012, bringing the tiny trio of South Pacific islands their largest delivery ever. On board were more than 4,000 solar panels and 1,000 storage batteries, as well as innumerable nails and screws. "We thought the island was going to sink," jokes energy minister Foua Toloa. The cargo gave Tokelau the moral high ground in the battle to halt global warming: it has been widely billed as the first nation to accomplish a sweeping shift from fossil fuels to renewable energy.

Tokelau, like almost all small island nations, used to rely on diesel-powered generators to meet the needs of its 1,400 residents. In its first full year of operation, the new 1-megawatt solar system met roughly 93% of the nation's electricity demand. Today, Tokelau has reduced its annual fuel bill by about $800,000, which more than covers payments on the loan it received from the government of New Zealand for the microgrid. "We're very proud," Toloa says. "We are challenging the world and the big emitters of greenhouse gases to equal or better what Tokelau has done."

A handful of Caribbean islands has signed up to that challenge with the help of the Carbon War Room, a Washington DC-based advocacy group founded by British entrepreneur Richard Branson. The Caribbean island of Aruba, where wind power currently provides 12% of demand, led the way in March 2012 with a commitment to eliminate fossil-fuel use by 2020. But with 109,000 residents and regular demand for roughly 100 megawatts of power, Aruba's challenge is much bigger than Tokelau's. "This is actually a very interesting proving ground to test out ambitious levels of renewables and energy efficiency," says Amory Lovins, co-founder of the Rocky Mountain Institute in Snowmass, Colorado, which co-hosted a clean-energy summit for Caribbean nations with the Carbon War Room in February. Lovins notes that the lessons learned about balancing energy supply and demand could help with the management of mainland power, too. Some US states, including New York, are exploring ways to divide the main grid into electricity 'islands' that could be isolated in the face of large-scale outages. And, Lovins adds, island projects such as that on Aruba may help to convince the world that reliable power systems can be built almost entirely from renewables.

Centralized grids are expected to provide only about 30% of the solution in rural areas.

Tokelau has not yet achieved its goal of going 100% renewable. The old diesel generators still occasionally kick in to charge batteries during the rainy season, and many residents rely on imported gas for cooking. The nation's government is planning to help residents to purchase more efficient appliances or make the switch to electric cookers. Air conditioners, considered an unnecessary luxury on the island, have already been banned for government use. And if the economics work out, as early as next year the country hopes to begin producing coconut oil to power the generators when the Sun doesn't shine. "We've got plenty of extra coconuts," Toloa says.

SINE MOUSSA ABDOU, SENEGAL

Residents in the village of Sine Moussa Abdou once had to trek ten kilometres to a neighbouring village to charge their mobile phones, paying fees as high as $110 per kilowatt hour — the average US rate is 12 cents. Those with televisions hauled a car battery to be recharged. The village's 900 residents now pay about $1.40 per kilowatt hour for power delivered to their homes through a microgrid built in 2009. The company in charge of supplying the power — a combination of wind, solar and diesel — says all of the students in the village school passed their annual exam for the first time one year after electrification, thanks to having enough light at night to study by.

The project is just one of many seeking to solve the massive energy problem in sub-Saharan Africa, where nearly 600 million people — more than two-thirds of the population — lack access to electricity (see 'In the dark'). But it is an innovative example of public–private partnership that many observers are watching closely.

The project, a partnership of Inensus in Goslar, Germany, and Matforce, based in Dakar, Senegal, was divided into two parts: international grants were used to wire up the village, but the power generation and supply is entirely unsubsidized. Inensus uses smart meters to track customers' usage, and asks users to pay for weeks' worth of power in advance, offering a discounted rate to those who predict and commit six months ahead. That information helps to keep costs and emissions down by ensuring that the wind and solar systems can cope, and that the diesel generators are not leaned on too heavily; they typically mop up the last 10–20% of demand.

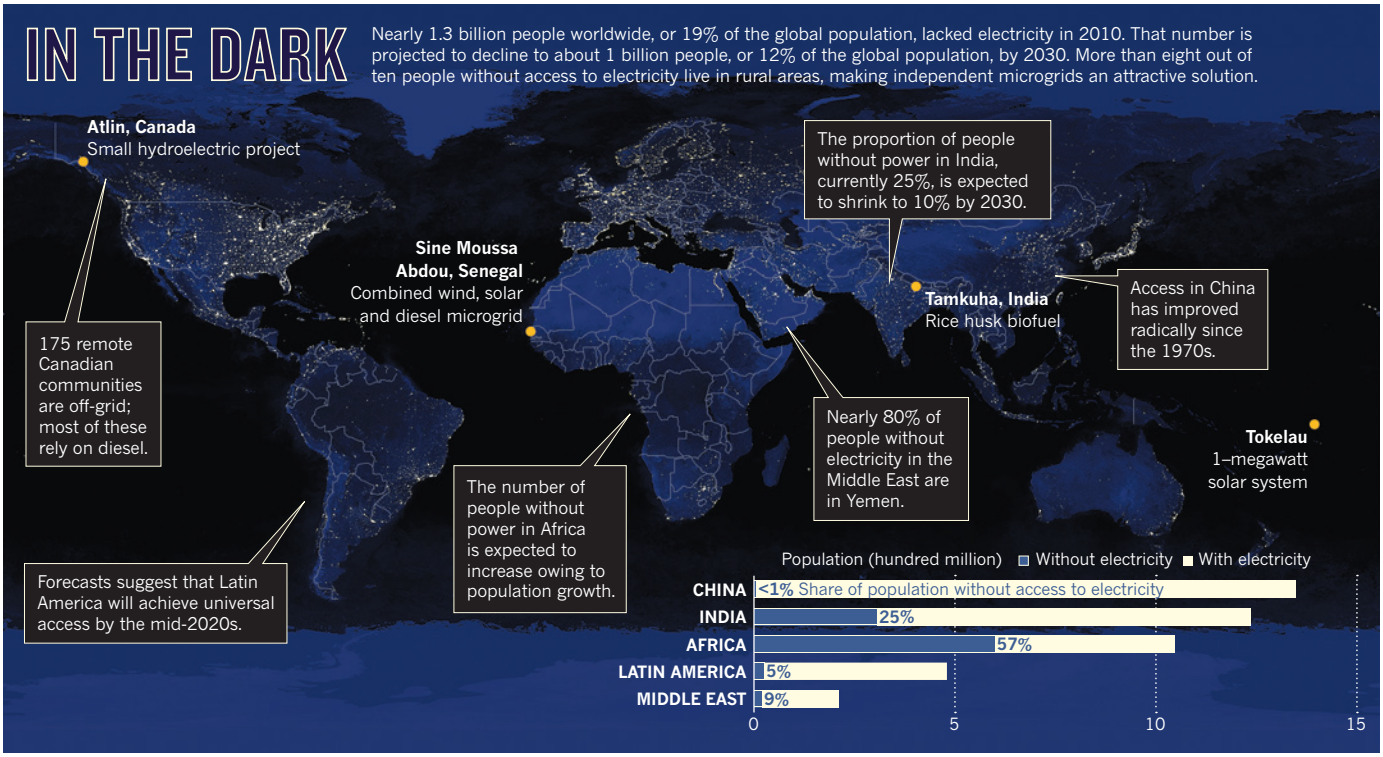


IMAGE: CRAIG MAYHEW AND ROBERT SIMMON, NASA GSFC/DATA SOURCE: IEA, WORLD ENERGY OUTLOOK 2013

Although the cost of the power is more than three times what urban customers might face, the business model is designed to promote sustainability and flexibility. Nico Peterschmidt, managing director of Inensus, foresees a scenario in which local companies own a grid and contract out the supply of power, fostering competition and freeing up companies and communities to shop around.

The company is now expanding the project into five nearby villages and has launched a larger project targeting 16 villages and 82,000 people in Tanzania. Peterschmidt says the Tanzanian government has perhaps the most advanced microgrid policy in the world, including a simple subsidy of $500 per connection for grid infrastructure, which covers the bulk of up-front costs. The biggest challenge, he says, is convincing governments to abandon fixed electricity rates that do not allow for company profit. "If we can overcome that, we can accelerate the private sector to provide energy access," he says. Kammen notes that there is generally enough commercial competition and watchdog activity to prevent price abuses, with most projects providing electricity at or below the price of diesel power.

The trick, notes Pepukaye Bardouille, an energy analyst who tracks microgrids at the International Finance Corporation in Washington DC, is to balance a nation's desire to attract profitable industry with the wish for their poorest people to get electricity. "Are we trying to promote commercially viable businesses? Or are we trying to promote access at any cost?" she asks. "Sometimes those two do not overlap."

Dean Cooper, an energy-finance specialist at the United Nations Environment Programme (UNEP) in Paris, says that UNEP is working on microgrid demonstrations in different countries to determine which policies and models work best. At present, it is too early to say which will win out. "All of the business models can be scaled up on paper," says Bardouille. "In practice, it is harder to deliver."

ATLIN, CANADA

For years, the only source of power in Atlin, an old mining town of some 400 people in the northwest corner of British Columbia, was diesel generators. The steady drone and smelly fumes were a constant reminder that money was going up in smoke, and members of the Taku River Tlingit First Nation, who make up 25% of the town's population, were determined to find an alternative. After experimenting with a wind turbine, which buckled under ice and wind in the winter of 2002–03, the tribal band settled on a small hydroelectric project. With $15 million from grants, community funds and loans, the Atlin hydroelectric project began generating 2.1 megawatts of power on 1 April 2009.

Atlin is enjoying the benefits. Eliminating diesel prevented more than 5,000 tonnes of greenhouse-gas emissions last year. And, because the First Nation owns the hydroelectric station, the money that residents pay for power stays at home. "We are paying our loans, but there's a little bit extra that is benefiting the community," says Stuart Simpson, general manager of the Atlin Tlingit Development Corporation.

There are 175 aboriginal or northern off-grid communities in Canada, most of which rely on diesel. The Taku River Tlingit First Nation was one of the first to switch to home-owned hydropower, and others are looking to follow.

A new wave of small-scale hydroelectric development aiming to supply both off-grid and on-grid energy in British Columbia has spurred controversy about the potential ecological impacts of such projects. The Vancouver-based Wilderness Committee, for one, has voiced concerns about possible disturbances to grizzly bear habitat and salmon-bearing rivers. But the actual effects are hard to tease out. A review released in January by the Pacific Salmon Foundation, a conservation organization in Vancouver, found no conclusive evidence about effects on fish.

Nigel Protter, executive director of the BC Sustainable Energy Association, says modern hydroelectric projects can, if well-conceived and implemented, improve the local ecosystem; the Atlin project, for example, built a fish ladder to help graylings get around the small dam, and Simpson says that fish counts have increased. The problem, Protter says, is that many rural communities in the developed world want more power than their small rivers can provide without the construction of dams to store up water. "Storage often creates additional environmental and social impacts."

Atlin has all the power it needs for now; it is even considering expanding its project to tie into the main electric grid and to export power to the northern Yukon territory. "In 20 years, when we have this bank loan paid off, we'll have a couple million coming into the community each year," Simpson says. "This is really about our grandkids." ■

Jeff Tollefson *covers climate, energy and the environment for* Nature.

Legend of the wolf

Predators are supposed to exert strong control over ecosystems, but nature doesn't always play by the rules.

BY EMMA MARRIS

The return of grey wolves to the western United States has sparked debate over their role in structuring ecosystems.

In the summer of 2008, Kristin Marshall was driving through Yellowstone National Park in Wyoming. Marshall, a graduate student at the time, had come to the park to study willow shrubs — specifically, how much they were being eaten by elk.

She pulled to the side of the road and was preparing to hike to one of her study plots when she ran into two sisters from the Midwest, who were touring the park. The women asked what Marshall was doing and she said, "I am a researcher. I am working in that willow patch down there."

The tourists gushed: "We watched all about the willows on this nature documentary. We hear that all the willows are doing so much better now because the wolves are back in the ecosystem." That stopped Marshall

short. "I didn't want to say, 'No, you are wrong, they aren't actually doing that well.'"

Instead, she said: "The story is a probably a little more complicated than what you saw on the nature documentary." That was the end of the conversation; the tourists seemed uninterested in the more-complicated story of how beavers and changes in hydrology might be more important than wolves for willow recovery. "I can't say I blame them," says Marshall, now an ecologist with the US National Oceanic and Atmospheric Administration in Seattle, Washington. "What you see on TV is captivating."

On television and in scientific journals, the story of how carnivores influence ecosystems has seized imaginations. From wolves in North America to lions in Africa and dingoes in Australia, top predators are thought to exert tight control over the populations and behaviours of other animals, shaping the entire food web down to the vegetation through a 'trophic cascade'. This story is popular in part because it supports calls to conserve large carnivores as 'keystone species' for whole ecosystems. It also offers the promise of a robust rule within ecology, a field in which researchers have yearned for more predictive power.

"Predators can be important, but they aren't a panacea."

But several studies in recent years have raised questions about the top-predator rule in the high-profile cases of the wolf and the dingo. That has led some scientists to suggest that the field's fascination with top predators stems not from their relative importance, but rather from society's interest in the big, the dangerous and the vulnerable. "Predators can be important," says Oswald Schmitz, an ecologist at Yale University in New Haven, Connecticut, "but they aren't a panacea."

PREDATORS ON TOP

In the early years of ecology, predators did not get so much respect. Instead, researchers thought that plants were the dominant forces in ecosystems. The theory was that photosynthesis from these primary producers determined how much energy was available in an area, and what could live there. Bottom-up control was all the rage.

Interest in top-down trophic cascades emerged in 1963, when ecologist Robert Paine of the University of Washington in Seattle started to exclude predators from study plots at his coastal research site. He pried predatory starfish off intertidal rocks and hurled them into deeper waters. Without the starfish to control their numbers, mussels eventually carpeted the plots and kept limpets and algae from taking hold in the region. A new ecosystem emerged (see *Nature* **493,** 286–289; 2013).

After this and other aquatic studies, the conventional wisdom in the field was that top-down trophic cascades happened only in rivers, lakes and the sea. An influential 1992 paper[1] by Donald Strong at the University of California, Davis, asked: "Are trophic cascades all wet?" As if in answer, ecologists began looking for similar carnivore stories on land.

They soon found them. In 2000, a review[2] tallied 41 terrestrial studies on trophic cascades, most of which showed that predation had significant effects on the number of herbivores in an area, or on plant damage, biomass or reproductive output. These studies were all on small plots involving small predators: birds, lizards, spiders and lots of ants.

Research on terrestrial trophic cascades moved to much larger scales with the work of John Terborgh and William Ripple. In 2001, Terborgh, an ecologist at Duke University in Durham, North Carolina, reported[3] on dramatic ecosystem changes that came after a dam was built in Venezuela. Flooding from the dam created islands that were too small to support big predators such as jaguars and harpy eagles. The population densities of their prey — rodents, howler monkeys, iguanas and leaf-cutter ants — boomed to 10–100 times those on the mainland. Seedlings and saplings were devastated.

In the same year, Ripple, an ecologist at Oregon State University in Corvallis, published a key paper[4] on the most famous, and probably the best-studied, example of a terrestrial carnivore structuring an ecosystem: Yellowstone's wolves. The ecosystem offered a natural experiment because the US National Park Service had the park's exterminated wolves (*Canis lupus*) by 1926 and then reinstated them in the 1990s, after public sentiment and ecological theory had shifted. In 1995, 14 wolves from Alberta, Canada, were introduced into the park. Seventeen from British Columbia followed in 1996. By 2009, there were almost 100 wolves in 14 packs in the Yellowstone area. (That number is now down to 83 in 10 packs.)

During the years when there were no wolves, ecologists grew increasingly worried about the aspen trees (*Populus tremuloides*) in the park. It seemed that intensive browsing by Rocky Mountain elk (*Cervus elaphus*) was preventing trees from reaching adult height, or 'recruiting'. In the early twentieth century, aspen covered between 4% and 6% of the winter range of the northern Yellowstone herd of elk; by the end of the century, they accounted for only 1% (ref. 4).

When Ripple and his co-authors checked aspen growth against the roaming behaviour of wolves in three packs, they found that aspen grew tallest in stream-side spots that saw high wolf traffic. That pattern hinted at an indirect behavioural cascade: rather than limiting browsing by reducing elk populations throughout the park, wolves apparently made elk more skittish and less likely to browse in the tightly confined stream valleys, where prey have limited escape routes (see 'The tangled web'). A 2007 study[5] by Ripple and Robert Beschta, also of Oregon State, seemed to strengthen the behavioural-cascade hypothesis. It found that the five tallest young aspen in stream-side stands where there were downed logs — a potential trip hazard for elk — were taller than the five tallest young aspen in stands away from streams or without downed logs.

Similar evidence of indirect wolf effects emerged from a study of willows. In 2004, Ripple and Beschta found[6] that the shrubs were returning in narrow river valleys, where the researchers thought that the chances of wolves attacking elk were greatest.

More recently, Ripple has been documenting the regrowth of cottonwood trees. "When we look around western North America, we see a big decrease in tree recruitment after wolves were removed. And when wolves returned to Yellowstone, the trees started growing again. It is just wonderful to walk through that new cottonwood forest."

TALES FROM TREES

But some ecologists had their doubts. The first major study[7] critical of the wolf effect appeared in 2010, led by Matthew Kauffman of the Wyoming Cooperative Fish and Wildlife Research Unit in Laramie. When researchers drilled boreholes into more than 200 trees in Yellowstone and analysed growth patterns, they found that the recruitment of aspen had not ended all at once. Some trees had reached adult size as late as 1960, long after the wolves had gone. And some stands had stopped growing new adults as early as 1892, well before the wolves left. The aspen petered out over decades, as elk populations slowly grew, suggesting that the major influence on the trees is the size of the elk population, rather than elk behaviour in response to wolves. And although wolves influence elk numbers, many other factors play a part, says Kauffman: grizzly bears are increasingly killing elk; droughts deplete elk populations; and humans hunt elk that migrate out of the park in winter.

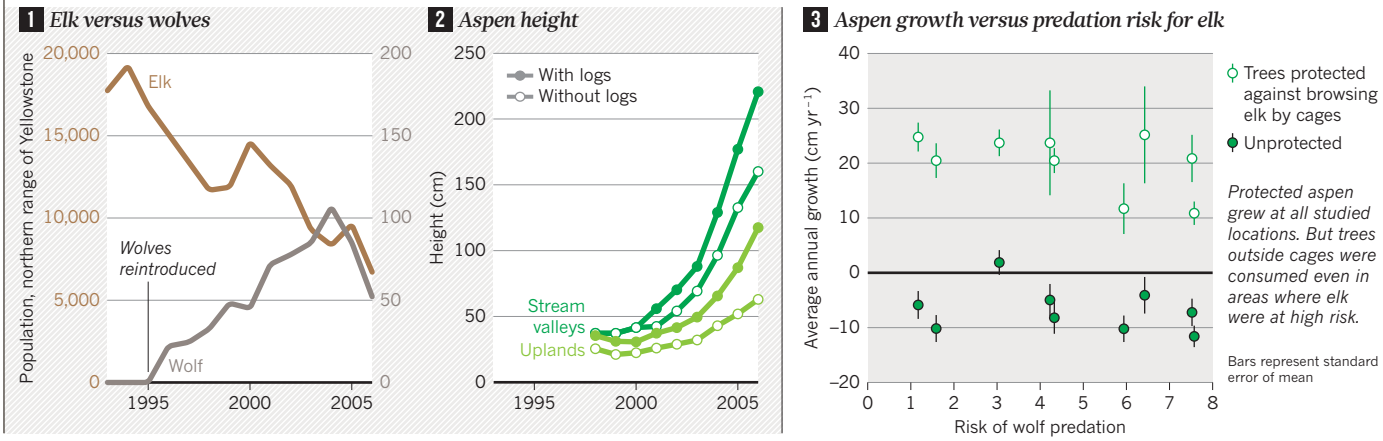
When Kauffman and his colleagues studied[7] aspen in areas where risk of attack by wolves was high or low, they obtained results different from Ripple's. Rather than look at the five tallest aspen in each stand, as Ripple had done, they tallied the average tree height and used locations of elk kills to map the risk of wolf attacks. By these measures, they found no differences between trees in high- and low-risk areas.

Questions have also emerged about the well-publicized relationship between wolves and willows. Marshall and two colleagues investigated the controls on willow shrubs by examining ten years' worth of data

NATURE.COM
For a podcast on predators, visit:
go.nature.com/x1npp3

The tangled web

Researchers disagree on whether the return of wolves to Yellowstone National Park (**1**) sparked a resurgence of aspen trees by limiting browsing by elk. One study found that aspen grow better in stream areas with fallen trees (**2**), where elk may feel most vulnerable to wolves. But another study found that aspen fare poorly even in areas where elk are most at risk from wolves (**3**).



SOURCE: 1 & 2: REF. 5; 3: REF. 7

from open plots and plots surrounded by cages to keep the elk out. Her team found[8] that the willows were not thriving in all the protected sites. The only plants that grew above 2 metres — beyond the reach of browsing elk — were those in areas where simulated beaver dams had raised the water table.

If beavers have a key role in helping willows to thrive, as Marshall's study suggests, the shrubs face a tough future because the park's beaver populations have dropped. Researchers speculate that the removal of wolves in the 1920s allowed elk to eat so much willow that there was none left for the beavers, causing an irreversible decline.

"The predator was gone for at least 70 years," says Marshall. "Removing it has changed the ecosystem in fundamental ways." This work suggests that wolves did meaningfully structure the Yellowstone ecosystem a century ago, but that reintroducing them cannot restore the old arrangement.

Arthur Middleton, a Yale ecologist who works on Yellowstone elk, says that such studies have disproved the simple version of the trophic cascade story. The wolves, elk and vegetation exist in an ecosystem with hundreds of other factors, many of which seem to be important, he says.

DINGO DEBATE

Another classic example of a trophic cascade has come under attack in Australia. The standard story there is that the top predator, the dingo (*Canis lupus dingo*), controls smaller introduced predators such as cats and foxes, allowing native marsupials to thrive. But Ben Allen, an ecologist at the Department of Agriculture, Fisheries and Forestry in Toowoomba, has compared[9] areas where dingoes are poisoned with areas where they are left alone, and found no difference in marsupial abundance. He is quite cynical, he says, about "this idea that top predators are wonderful for the environment and will put everything back to the Garden of Eden".

Allen's opponents counter that he has failed to show that the poisoning regimens actually reduce dingo population densities. Chris Johnson, an ecologist the University of Tasmania in Hobart, says he is "very critical" of Allen's experimental design and methods. The dingo effect is real, says Johnson.

Ripple is not worried about these debates, which he views as quibbling over details that do not undermine the overall strength of the tropic-cascade hypothesis. In fact, when he published a major review[10] this year of the effects that predators exert over ecosystems, he left out studies critical of the wolf and dingo trophic-cascade theories; he says that there was no room for them in the space he had to work with. Ripple is particularly concerned with documenting the impacts of Earth's top carnivores because so many are endangered. "We are losing these carnivores at the same time that we are learning about their ecological effects," he says. "It is alarming, and this information needs to be brought forth."

The debate has been harsh at times, but in quieter moments the different factions all tend to talk in similar terms about the great complexity of ecosystems and the likelihood that the truth lies somewhere in the middle. James Estes, an ecologist at the University of California, Santa Cruz, and one of the fathers of the trophic-cascade idea, says that the evidence for cascades mediated by changes in animal behaviour rather than by changes in animal number is "thin", at the moment — and that many of the effects that have been documented are spotty and badly need to be rigorously mapped out. Still, he adds, "When all is said and done, and everyone is dead 100 years from now, Bill [Ripple] will be closer to right".

Although Ripple stresses the role of the top carnivores, he agrees they are not the end of the story. "I believe in the combination of top-down and bottom-up, working in unison," he says. "They are both playing out on any given piece of ground and the challenge will be to discover what determines their interactions and relative effects."

Schmitz has some thoughts on how to do that. His own smaller-scale work on invertebrates has convinced him that neither bottom-up nor top-down theories adequately capture the story of ecosystems. He is starting to look at the middle players, such as elk, beavers and grass-eating grasshoppers. These herbivores, he says, integrate influences from both the top (such as predation pressure) and the bottom (such as the nutritional quality of plants). "It is not really bottom-up or top-down but trophic cascades from the middle out," he says. "That is where we will evolve. It is knowing what the middle guy is going to do that gives you the predictive ability."

It remains to be seen whether theories such as this middle-out idea will grip researchers and the public as much as the theory of top-down cascades. Many researchers have doubts. They worry that tales of predators shaping their ecosystems are so attractive that they have unrivalled control over discourse. "Everyone likes to think of the big wolf or the big bear looking after the environment," says Allen. "We do love a good story." ■ SEE EDITORIAL P.139

Emma Marris *is a freelance writer in Klamath Falls, Oregon.*

1. Strong, D. R. *Ecology* **73,** 747–754 (1992).
2. Schmitz, O. J., Hambäck, P. A. & Beckerman, A. P. *Am. Nat.* **155,** 141–153 (2000).
3. Terborgh, J. *et al. Science* **294,** 1923–1926 (2001).
4. Ripple, W. J., Larsen, E. J., Renkin, R. A. & Smith, D. W. *Biol. Conserv.* **102,** 227–234 (2001).
5. Ripple, W. J. & Beschta, R. L. *Biol. Conserv.* **138,** 514–529 (2007).
6. Ripple, W. J. & Beschta, R. L. *Forest Ecol. Management* **200,** 161–181 (2004).
7. Kauffman, M. J., Brodie, J. F. & Jules, E. S. *Ecology* **91,** 2742–2755 (2010).
8. Marshall, K. N., Hobbs, N. T. & Cooper, D. J. *Proc. R. Soc. B* **280,** 20122977 (2013).
9. Allen, B. L. in *Wolves: Biology, Behavior and Conservation* (eds Maia, A. P. & Crussi, H. F.) Ch. 3 (Nova, 2012).
10. Ripple, W. J. *et al. Science* **343,** 1241484 (2014).

COMMENT

POLICY New Zealand's science adviser shares fruits of his experience **p.163**

TREES Foundational forestry book reviewed 350 years on **p.166**

FILM Higgs documentary gets under the skin of day-to-day discovery **p168**

WARMING China's climate-change reports should be taken more seriously **p.169**

L'IMAGE DU MONDE BY GOSSUIN DE METZ, BIBLIOTHÈQUE NATIONALE DE FRANCE

A thirteenth-century depiction of the geocentric cosmos.

A medieval multiverse

Ideas in a thirteenth-century treatise on the nature of matter still resonate today, say **Tom C. B. McLeish** and colleagues.

Earlier this year we submitted an unusual paper[1] to a scientific journal. What is unusual about it is not the topic — computations of how interactions between light and matter in the primordial Universe affected large-scale cosmic structures — but what inspired it. The paper draws on ideas in a medieval manuscript by the thirteenth-century English scholar Robert Grosseteste.

De Luce (*On Light*), written in 1225 in Latin and dense with mathematical thinking, explores the nature of matter and the cosmos. Four centuries before Isaac Newton proposed gravity and seven centuries before the Big Bang theory, Grosseteste describes the birth of the Universe in an explosion and the crystallization of matter to form stars and planets in a set of nested spheres around Earth.

To our knowledge, *De Luce* is the first attempt to describe the heavens and Earth using a single set of physical laws. Implying, probably unrealized by its author, a family of ordered universes in an ocean of disordered ones, the physics resembles the modern 'multiverse' concept.

Grosseteste's treatise was translated and interpreted by us as part of an interdisciplinary project led by Durham University, UK, that includes Latinists, philologists, medieval historians, physicists and cosmologists (see ordered-universe.com). Our experience shows how science and humanities scholars working together can gain fresh perspectives in both fields. And Grosseteste's thesis demonstrates how advanced natural philosophy was in the thirteenth century — it was no dark age.

EARLY INSIGHTS

Many coffee-table histories of science maintain that the natural philosophy of the medieval centuries constituted a scientific dead end — burrowing ever deeper into alchemy and astrology. A closer examination reveals a more nuanced story. Preserved on vellum manuscripts, written in coded medieval Latin and enveloped in unfamiliar metaphysics it may be, but the science of the twelfth and thirteenth centuries constitutes a crucial stage in the history of thought.

By the late twelfth century, Aristotle's observation-oriented science had burst afresh onto the European scene, transmitted in a long series of cross-cultural translations from Greek to Arabic to Latin. Great ▶

questions arose in the minds of scholars such as Grosseteste, Averroes (in Cordoba) and Gerard of Cremona (in Toledo). What is colour? What is light? How does the rainbow appear? How was the cosmos formed? We should not underestimate the imaginative work needed to conceive that these questions were, in principle, answerable.

Grosseteste (*c.*1175–1253) rose from obscure Anglo–Norman origins to become a respected theologian and Bishop of Lincoln. He was one of the first in northern Europe to read the newly translated scientific works of Aristotle, attempting to take forward the big questions of what we can know about the natural world (ontology) and how we know it (epistemology). The late thirteenth-century philosopher Roger Bacon called him "the greatest mathematician" of his time. Grosseteste's work on optical physics influenced mathematicians and natural philosophers for generations, notably in Oxford during the fourteenth century and in Prague during the fifteenth.

Exploring the scientific thought of the thirteenth century is inherently interdisciplinary, requiring knowledge of Latin, history and philosophy, as well as of mathematics and science. Our collaboration at Durham began in 2008, following a seminar on Grosseteste by one of us, Tom McLeish, a physicist who had become interested in the thirteenth-century thinkers after hearing talks at Leeds University, UK, by historian James Ginther of Saint Louis University in Missouri.

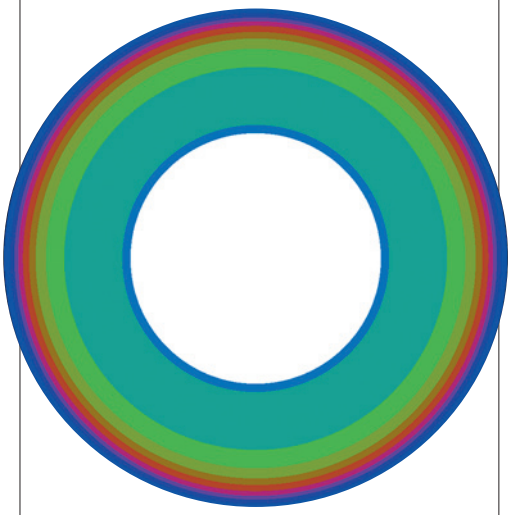
Intrigued, medieval scholars at Durham, including Giles Gasper, recruited other Grosseteste specialists, including Cecilia Panti at the University of Rome Tor Vergata, Neil Lewis at Georgetown University in Washington DC, and the Latinist Greti Dinkova-Bruun at the Pontifical Institute of Mediaeval Studies in Toronto, Canada. Before tackling *De Luce*, we honed our skills[2,3] on two simpler short works by Grosseteste: *De Colore*, on colour theory, and *De Iride*, on the rainbow, aided by colour psychophysicist Hannah Smithson at the University of Oxford, UK, and Durham optical physicist Brian Tanner.

LIGHT WORK

Grosseteste's *De Luce*, available in English since the 1940s, opens by addressing a problem with classical atomism: why, if atoms are point-like, do materials have volume? Light is discussed as a medium for filling space. Grosseteste's recognition that matter's bulk and bulk stability requires subtle explanation was impressive. Even more intriguing was his use of mathematics to illuminate his physics.

A finite volume, he writes, emerges from an "infinite multiplication of light" acting on infinitesimal matter. He draws an analogy to the finite ratio of two infinite sums, claiming that $(1+2+4+8+\ldots)/(0.5+1+2+4+\ldots)$ is equal to 2. He does not articulate carefully the idea of the limits one needs to make this rigorous, but we know what he means — simultaneously adding to both numerator and denominator keeps the ratio finite.

The third remarkable ingredient of *De Luce* to modern eyes is its universal canvas: it suggests that the same physics of light and matter that explains the solidity of ordinary objects can be applied to the cosmos as a whole. An initial explosion of



A simulation of Robert Grosseteste's nine-sphere universe.

a primordial sort of light, *lux*, according to Grosseteste, expands the Universe into an enormous sphere, thinning matter as it goes. This sounds, to a twenty-first-century reader, like the Big Bang.

Then Grosseteste makes an assumption: matter possesses a minimum density at which it becomes 'perfected' into a sort of crystalline form. Today, we would call this a phase transition. The perfection occurs first at the thinnest outer edge of the cosmos, which crystallizes into the outermost sphere of the medieval cosmos. This perfect matter radiates inward another sort of light, *lumen*, which is able to push matter by its radiative force, piling it up in front and rarefying it behind. An analogous process in today's physics is the inward propagation of shock waves in a supernova explosion.

Like a sonata returning to its theme, that finite ratio of infinite sums reappears, this time as a 'quantization condition' — a rule that permits only discrete solutions such as the energy levels in atoms — that limits matter to a finite number of spheres. Grosseteste needed to account for nine perfect spheres in the medieval geocentric cosmos: the 'firmament', the fixed stars, Saturn, Jupiter, Mars, the Sun, Venus, Mercury and the Moon. By requiring that the density is doubled in the second sphere and tripled in the third, and so on, a nested set of spheres results.

In an impressive final stroke of unification, he postulates that towards the centre of the cosmos, the remaining unperfected matter becomes so dense and the inwardly radiating *lumen* so weak, that no further perfection transitions are possible. He thus accounts for the Aristotelian distinction between the perfect heavens and the imperfect Earth and atmosphere.

MODERN TOOLS

To our knowledge, *De Luce* is the first worked example showing that a single set of physical laws might account for the very different structures of the heavens and Earth, hundreds of years before Newton's 1687 appeal to gravity to unite the falling of objects on Earth with the orbiting of the Moon. Our translation has also cleared up a misconception in some previous studies that the light in Grosseteste's treatise travelled both inwards and outwards.

To explore the consequences of the physics in the treatise further, and to urge a closer and more careful reading of the text, the science team turned to modern tools. *De Luce* is remarkably precise in its formulation of physics — had Grosseteste had access to the mathematics of calculus and the computing power we have today, it would have been natural to apply them.

We identified six physical 'laws' in the manuscript, including the interaction of light and matter, the critical criteria for perfection, and the re-radiation and absorption of *lumen*. We wrote down these laws mathematically, including modern concepts such as opacity, which were consistent with the text although not described explicitly in it. Then we computed the resulting equations, expressed in differential form, in three-dimensional spherical symmetry.

To assess the range of possible solutions to these novel equations, and out of curiosity, Durham cosmologist Richard Bower then computed the space of possible medieval universes by varying the values of four parameters: the gradient of the initial 'Big Bang' matter distribution, the coupling strength of light and matter, the opacity of impure matter and the transparency of the perfected spheres.

A rich set of solutions emerged. A narrow set of parameters did indeed produce the series of celestial perfected spheres and, within the Moon's orbit, a further four spheres corresponding to fire, air, water and earth — as the medieval world view demanded. But most choices of the four parameters yielded no spheres, or a disordered mess of hundreds of concentric spheres with no radial pattern to their densities. Other possible model universes contained infinite numbers of spheres, some with unbounded density. The project had unwittingly stumbled on a medieval multiverse.

The possible existence of more than one

REF. 1

universe was indeed a live issue of the period, and a highly contentious one — appearing, for example, in the Papal edict of 1277 that banned a list of scientific teachings. But it was a debate that Grosseteste apparently chose to avoid. None of his surviving treatises discusses the possibility of other forms of universe, however close he came to implying it in his cosmogony.

Of course we know now, thanks to telescope observations from the early seventeenth century onwards, that a geocentric cosmos is untenable. But in 1225, it was the simplest theory consistent with the observations. Grosseteste's effort to give a physical account of its origin is an impressive achievement, but it also reminds us of the limitations of our own current cosmological theory, with its reliance on intangible factors such as 'dark matter' and 'dark energy'.

SCIENTIFIC VALUE

The translation of *De Luce* is an exemplar of the importance of collaborations between the arts and sciences, of thinking and learning together in new ways, and a reminder that the intellectual tradition we now call science has a long and rich history.

Both the scientists and the humanities scholars in our collaboration found working together enriching and transformational; it forced us to engage with different ideas and problems. There were challenges: getting used to each other's methodologies and approaches took time and patience. And our expectations changed. At the start of the project we had hoped for a sharper understanding of the text; we were surprised when new science emerged as well.

What next for the collaboration? The Durham-led team has examined three of Grosseteste's science works in detail so far. There are at least another ten to explore, including a work on the origin of sounds (*De Generatione Sonorum*). The scientific writings of Grosseteste's immediate predecessors, Alfred of Sareshull and Alexander Neckham, and his successors, including Bacon, could hold similar insights into the evolution of ideas.

Funding for such interdisciplinary work, however, remains a problem. In the United Kingdom, none of the scientific research councils offered grants for such a project. In the end, we were funded by the Arts and Humanities Research Council. The US granting system is similarly biased. The European Research Council and philanthropic sources, such as the Wellcome Trust, do fund science–arts projects, but in our experience it is easier to obtain funding for science and social science collaborations than for science and humanities partnerships.

Because projects such as ours can be of significant scientific and cultural value, scientific granting agencies should consider funding arts and sciences projects or partnering with arts and humanities councils to translate other early scientific works, for example.

The eight-century journey from Grosseteste's cosmological ideas to our own offers a rich illustration of the slow evolution in our understanding, and of the delight to be found in reaching out into nature with our imagination. ■

Tom C. B. McLeish *is in the Department of Physics, Durham University, UK. His book* Faith and Wisdom in Science *is published in May.* **Richard G. Bower** *and* **Brian K. Tanner** *are in the Department of Physics, Durham University, UK.* **Hannah E. Smithson** *is in the Department of Experimental Psychology, University of Oxford, UK.* **Cecilia Panti** *is in the Department of Corporate Governance Philosophy, University of Rome Tor Vergata, Italy.* **Neil Lewis** *is in the Department of Philosophy, Georgetown University, Washington DC, USA.* **Giles E. M. Gasper** *is in the Institute of Medieval and Early Modern Studies and in the Department of History, Durham University, UK.*
e-mail: t.c.b.mcleish@durham.ac.uk

1. Bower, R. G. *et al.* Preprint at http://arxiv.org/abs/1403.0769 (2014).
2. Smithson, H. E. *et al. J. Opt. Soc. Am.* **29,** A346–A352 (2012).
3. Smithson, H. E. *et al. J. Opt. Soc. Am.* **31,** A341–A349 (2014).

The art of science advice to government

Peter Gluckman, New Zealand's chief science adviser, offers his ten principles for building trust, influence, engagement and independence.

In 2009, I was appointed as the first science adviser to the Prime Minister of New Zealand. The week I was appointed coincided with the government announcement that the New Zealand food industry would not be required to add folate to flour-based products to help to prevent neural-tube defects in newborns, despite an earlier agreement to do so. As it happens, this is an area of my own scientific expertise and, before my appointment, I had advised the government that folate supplementation should occur. But various groups had stirred considerable public concern on the matter, about health risks and about medicalizing the food supply.

Thus, in my first media interview as science adviser I was asked how I felt about my advice not being heeded. I pointed out that despite strong scientific evidence to support folate supplementation, a democratic government could not easily ignore overwhelming public concern about the food supply. The failure here was not political; rather, it was the lack of sustained and effective public engagement by the medical-science community on the role of folate in the diet. As a result, the intervention did not get the social licence necessary to proceed.

Five years on, I am still in the post. I have come to understand that the primary functions and greatest challenges for a science adviser are providing advice not on straightforward scientific matters, but instead on issues that have the hallmarks of what has been called post-normal science[1]. These issues are urgent and of high public and political concern; the people involved hold strong positions based on their values, and the science is complex, incomplete and uncertain. Diverse meanings and understandings of risks and trade-offs dominate.

Examples include the eradication of exogenous pests in New Zealand's unique ecosystems, offshore oil prospecting, legalization of recreational psychotropic drugs, water quality, family violence, obesity, teenage morbidity and suicide, the ageing ▶

population, the prioritization of early-childhood education, reduction of agricultural greenhouse gases, and balancing economic growth and environmental sustainability.

My own experience is of a Westminster-style parliamentary democracy in a small advanced economy. In this context, I have distilled a set of ten principles to guide my work. Other countries have different forms of government and different cultural histories of public reason; high-level scientific advice may be provided by individuals, councils or academies, or a combination. Nevertheless, I think my guidelines are relevant to all those providing advice to senior levels of government.

"Policy-makers and elected officials rightly guard their responsibility to define policy."

These principles differ a little from those that might guide individual researchers and academics in attempting to influence policy in areas of their own interest and expertise[2,3]. Crucially, science advisers are obliged to advise in the context of the policy process. This means elucidating the evidence-informed options, rather than simply advocating a course of action.

TOP TEN

Maintain the trust of many. The science adviser must sustain in parallel the trust of the public, the media, policy-makers, politicians and the science community. This is especially true in times of crisis and is no small challenge. Food-safety panics such as foot-and-mouth disease and Creutzfeldt–Jakob disease (CJD) catalysed a strengthening of the science-advisory system in the United Kingdom, enhancing the roles of departmental science advisers. The aftermath of the 2011 nuclear meltdown in Fukushima is causing Japanese officials to take a critical look at advisory practices[4].

In my case, it was an earthquake that tested trust. In early 2011, Christchurch in New Zealand experienced the second of two major earthquakes in six months. The events had devastating consequences, including nearly 200 deaths, and effectively destroyed our second-largest city. This cluster of quakes was unusual and led to seismologists publicly competing in their interpretation of the nature of the fault-lines and future risks. This confused the public and policy-makers.

It took considerable dialogue with the scientists for them to understand the need for simple and consistent communication, and to accept that erudite and, in many ways, self-serving scholarly discourse did not belong on the front page of newspapers every day. What was needed was clear communication of the knowns and unknowns.

Worse, because the earthquake happened on the day of a full Moon, a popular astrologer got primetime coverage when he predicted another big earthquake would occur one month later when the Moon and Sun would next be in alignment. Panic set in. We worked with New Zealand's Science Media Centre to calm the public while acknowledging seismic uncertainty.

Protect the independence of advice. The advisory role should be structured so as to protect its independence from both political interference and premature filtering in the policy process. There is inevitably a tension between such independent advice and departmental policy processes, and it takes considerable diplomacy to create a trusted partnership between an external adviser and departmental officials. The terms of my appointment protect my independence in that I continue to be an employee of my academic institution, seconded to the prime minister, and my advisership is not tied to the electoral cycle. That said, an adviser must recognize that publicly disturbing the democratic process could mean losing the trust of the elected leader and thus any potential for influence. So there are issues on which a national scientific academy or a panel may be best placed to advise or to be seen supporting the individual adviser.

Report to the top. Scientific advice must be available directly — uncensored — to the head of government or the head of the relevant department. Indeed, the questions for which advice is most often sought tend to be politically sensitive and cut across individual portfolios.

In New Zealand, for instance, the economic importance of land-based primary industries must be balanced against maintaining our ecosystems and the eco-tourism industry built on them. These concerns are the responsibility of separate ministries, whose respective takes on environmental impact are inevitably framed by their mandates. The adviser's perspective transcends these.

Distinguish science for policy from policy for science. Science advising is distinct from the role of administering the system of public funding for science. There is potential for perceived conflict of interest and consequent loss of influence if the science adviser has both roles. There is a risk that the adviser comes to be perceived as a lobbyist for resources, or that the role becomes restricted to the ministry that manages the national research funding. Yes, a science adviser should have input into science policy, but there is a delicate balance to strike.

Early in my appointment, an unnecessary tension was created by media portrayals that I, rather than the relevant ministry, was the key influence on science policy. This is not so — nor did I want it to be — but communication with that ministry became strained for some time, denting my effectiveness.

Expect to inform policy, not make it. Science advice is about presenting a rigorous analysis of what we do and do not know. Alone, it does not make policy. There are many other appropriate inputs to policy, including fiscal considerations and public opinion. Policy-makers and elected officials rightly guard their responsibility to define policy — and this means choosing between

KYODO/REUTERS

Post-normal science: Christchurch, New Zealand, after its second earthquake in six months.

A gas-production station near Mount Taranaki in New Zealand's North Island.

DAVID WALL/ALAMY

options with different trade-offs. This is not the domain of a science adviser. Being explicit about this[5] has eased my capacity to establish and sustain trust broadly across government and the policy community.

Give science privilege as an input into policy. While acknowledging the other relevant inputs into policy formation, we need to demonstrate why science should hold a privileged place among the 'types of knowledge' that may be meaningful to a politician. These include social tradition and popular belief. The privilege of science-derived knowledge comes from its set of standard procedures — for example, replication and peer review — that limit the influence of beliefs and dogma. The other inputs into policy are value-intensive, and rightly so.

Recognize the limits of science. Science can increasingly address complex questions over which policy-makers and elected officials agonize. But scientists must not overstate what is or can be known, even though the shift from a view of science as a source of certainty to a source of probability can frustrate and confuse decision-makers and the public. How many politicians or issues advocates have claimed that they can find a scientist to back any position as, indeed, at least one did in the folate debate? This attitude reflects the dangerous temptation to use science to justify value-based beliefs[6] and a lack of literacy about what science is (a process)[7]. For example, much of the debate about climate change is not primarily about the data. Rather, it is about intergenerational economic interests.

Act as a broker not an advocate. Trust can be earned and maintained only if the science adviser or advisory committee acts as a knowledge broker, rather than as an advocate[6] — often a subtle distinction. When formal science advice is perceived as advocacy, trust in that advice and in the adviser is undermined, even if the advice is accepted. For example, exaggerated presentations about the causes of storms and floods can erode the credibility of the underlying argument about global warming. My own academic research has been focused on the developmental origins of obesity, and I have had to be careful ensure a balance of advice. Even so, where there is strong advocacy for other approaches, suspicion about the balance of my advice is hard to avoid (see go.nature.com/syxyee).

Engage the scientific community. The science adviser must know how to reach out to scientists for the appropriate expertise, and help them to enact their social responsibility in making their knowledge accessible and understandable, and in being more self-aware about when they might be acting as advocates. These issues are encapsulated in the recently updated, groundbreaking Code of Conduct for Scientists[8], which directly implies a distinction between brokerage and advocacy, published by the Japanese Council of Science.

Engage the policy community. The role of the science adviser is often less about providing direct technical expertise than it is about nudging attitudes and practices to enhance both the demand for and the supply of evidence for public policy.

Why? Because sceptics in the policy community are surprisingly prevalent. In 2012, I surveyed how our public-service personnel use evidence in making policy[4]. Several ministries stated that their job was to design policy that met the minister's requirements, not to advise on policy options on the basis of available evidence. Studies in Canada and Australia[9] found similar results.

GOOD ADVICE

These principles that guide my own work probably apply to most models of science advice. The use of advisers, advisory councils or academies need not be mutually exclusive. Different approaches suit different purposes and are the product of a country's culture, history, political and social structures and approaches to civic reason[10].

In my experience, achieving the culture change that encourages the better use of scientifically derived evidence in government relies on a level of trust that may be best achieved by one-to-one relationships with senior members of the executive government. In crises, such relationships are essential. By contrast, for complex and chronic issues, I believe that advisory committees or academies have a crucial part to play.

Happily, these matters are increasingly being discussed. In August this year, the first global conference of academics and practitioners of science advice to governments will take place in Auckland, New Zealand (see www.globalscienceadvice.org). I hope that two days of discussion between thought leaders from around the world about principles, methods, tensions and solutions from myriad contexts will make an important contribution to this rapidly growing field. ■

Peter Gluckman *is chief science adviser to the Prime Minister of New Zealand. He is also professor at the Liggins Institute, University of Auckland, New Zealand.*
e-mail: csa@pmcsa.org.nz

1. Funtowicz, S. O. & Ravetz, J. R. *Futures* **25,** 739–755 (1993).
2. Sutherland, W. J., Speigelhalter, D. & Burgman, M. A. *Nature* **503,** 335 (2013).
3. Douglas, H. *Science, Policy and the Value-Free Ideal* (Univ. Pittsburgh Press, 2009).
4. Doubleday, R. & Wilsdon, J. *Nature* **485,** 301–302 (2012).
5. Gluckman, P. D. *Towards Better Use of Evidence in Policy Formation: A Discussion Paper* (Office of the Prime Minister's Science Advisory Committee, 2011); available at http://go.nature.com/ntrw7u.
6. Pielke, R. A. Jr *The Honest Broker: Making Sense of Science in Policy and Politics* (Cambridge Univ. Press, 2007).
7. Gluckman, P. D. *The Role of Evidence in Policy Formation and Implementation* (Office of the Prime Minister's Science Advisory Committee, 2013); available at http://go.nature.com/kmxmbx.
8. Science Council of Japan *Statement: Code of Conduct for Scientists – Revised Version* (SCJ, 2013); available at http://go.nature.com/nhrnbb.
9. Hickey, G. M., Forest, P., Sandall, J. L., Lalor, B. M. & Keenan, R. J. *Sci. Pub. Pol.* **40,** 529–543 (2013).
10. Jasanoff, S. *Science and Public Reason* (Routledge, 2012).

SHAKESPEARE BIRTHPLACE TRUST

Royal Society co-founder John Evelyn, painted in around 1650 by Dutch artist Adriaen Hanneman.

IN RETROSPECT

Sylva

Gabriel Hemery celebrates the 350th anniversary of John Evelyn's treatise on the science and practice of forestry.

Three hundred and fifty years ago, London's recently formed Royal Society — the body at the heart of the Enlightenment — published its first book. It was written not by Robert Boyle, Isaac Newton or any of the other luminaries of seventeenth-century experimental philosophy, but by another founding member of the society: the prodigious public servant John Evelyn (1620–1706). And its subject was not anatomy, astronomy, chemistry or optics, but forestry. *Sylva* is a practical treatise on silviculture and an enduring classic, published in four editions during Evelyn's lifetime and posthumously in a further six, up until 1825. It remained the dominant forestry treatise in English until the nineteenth century and, thanks to its rich language, remains a favourite among tree experts. It inspired my own forthcoming book with Sarah Simblet, *The New Sylva* (Bloomsbury, 2014; www.newsylva.com).

Sylva was a response to an early venture of the Royal Society. Various committees had been formed to help to organize experiments and produce reports, and one of the first set out to respond to the Navy Royal's concerns about timber shortages arising from the degraded state of the nation's forests.

Evelyn took the lead in this venture, presenting a paper to the society in October 1662 that was imprinted as a book some 18 months later. Evelyn wrote in his diary — alongside that of his contemporary and friend Samuel Pepys, a record of significant historical importance — on 16 February 1664 that "I presented my 'Sylva' to the Society; and next day to his Majestie [Charles II], to whom it was dedicated; also to the Lord Treasurer and the Lord Chancellor".

Sylva; Or a Discourse of Forest-Trees, and the Propagation of Timber in His Majesties Dominions
JOHN EVELYN
Royal Society: 1664.

Sylva encouraged the nation's landowners to plant more trees and care for their forests, in the interests of the strategic defence of a nation reliant on 'wooden walls' — that is, the navy. It inspired considerable interest in tree planting in Britain, both as new forests on private estates, and on city streets and in formal gardens. Evelyn wrote on everything from London smog (*Fumifugium*, 1661) and salad (*Acetaria*, 1699) to soils (*A Philosophical Discourse of Earth*, 1676), and served as Commissioner for the Privy Seal and as Treasurer to the Royal Hospital for Seamen at Greenwich. But *Sylva* and his diary comprise his greatest legacy.

CULTIVATING VARIETY

As the wellspring of the Enlightenment, the seventeenth century witnessed considerable botanical discovery and geographical exploration in the New World and the Far East. German naturalist Engelbert Kaempfer, for example, was the first European to describe the maidenhair tree (*Ginkgo biloba*), which he saw in Japan. Evelyn advocated introducing new tree species to Britain, where diversity was limited to 60 native species. As he wrote, it was important "to promote the culture of such plants and trees (especially timber) as may yet add to those we find already agreeable to our climate in England" (this and other quotes taken from the 1776 edition of *Sylva*).

Evelyn was born into a family whose wealth was founded on gunpowder. He attended Balliol College at the University of Oxford but never graduated, prevented by his father's ailing health and the rumblings of the English Civil War. During the Interregnum, Evelyn travelled widely in Europe, returning to England in 1652 considerably better educated in areas such as anatomy. Notably, he now had a strong interest in horticulture after witnessing continental garden design, and had collected specimens and seeds of exotic plants. At Sayes Court near London, he transformed the garden, introducing a European formality merged with traditional English informality. In the same decade, Evelyn began to write his vast gardening treatise *Elysium Britannicum*, which he worked on for much of his life but never completed. It was largely down to his

KING'S COLLEGE LONDON

The Cawthorne oak in winter, as depicted in the 1801 edition of John Evelyn's *Sylva*.

prowess as a garden designer and plantsman that he was asked to lead the enquiry on the state of the nation's forests and their care.

Evelyn's personal motto was *omnia explorate; meliora retinete* (explore everything; keep the best), and in *Sylva* he adhered closely to this ambition. He described in detail the tree species of "greatest use, and the fittest to be cultivated", dwelling mostly on oak. As many as 2,000 oak trees were required for each navy ship. Of its wood, he notes that "though some trees be harder, as Box, Cornus, Ebony, and divers of the Indian woods; yet we find them more fragil, and not so well qualified to support great incumbencies and weights, nor is there any timber more lasting". After oak, he gave ash, elm and pine greatest prominence, given their utility in shipbuilding, construction and everyday life.

He discusses the natural environment — air, soil and water — and tree-nursery and forest management, tree diseases, and the cultural significance of trees and forests. He details how to collect seeds, raise young plants, prune (often improving the healing of a cut with cow dung) and optimize timber use. He relies heavily on the wisdom of ancient philosophers, such as Pliny the Elder, melded with the contemporary and practical silviculture practised by the landed gentry. He also includes many medicinal remedies — for example, ash for toothache, or box for venereal diseases — although he admits that "quacking is not my trade; I speak only here as a plain husband-man, and a simple forester".

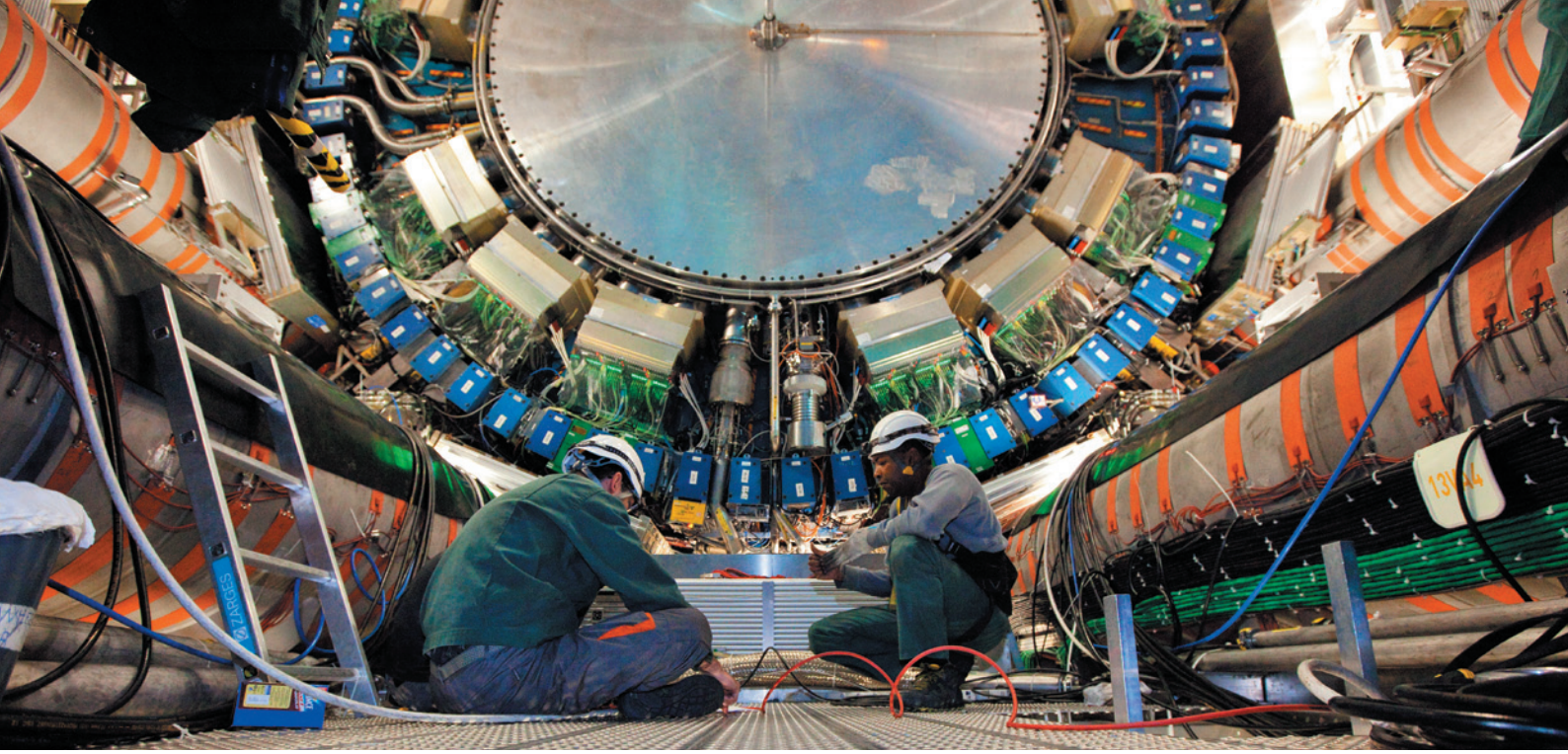
Evelyn inspired landowners to plant more forest trees, yet such is the lag between vision and fruition in forestry that the oak and other productive forest species intended for shipbuilding were eventually to support other industries, especially as pit props for coal mining. He also sought to ensure the protection of Britain's forests, but it was not until the mid-eighteenth century that an Act of Parliament offered them formal protection. Despite both afforestation and conservation, Britain's forested area continued to dwindle. It reached an all-time low of 5% at the start of the twentieth century, and in response the Forestry Commission was formed in 1919 to coordinate an afforestation programme aiming to create a strategic reserve of timber for the nation.

"Sylva was a response to the Navy Royal's concerns about timber shortages."

Like many foresters, Evelyn had foresight and ambition that echoed well beyond his own lifetime. Society has finally come to appreciate the functions of forest soils in the carbon cycle, the role of the world's forests in combating climate change, the importance of the world's forests and their associated biodiversity, and the role that trees have in maintaining human wellbeing. We are just beginning to realize the true potential of renewable materials made from woody biomass. Skyscrapers up to 30 stories high and of mass timber construction are being considered. Nanocrystalline cellulose made from wood pulp — a material stronger than steel — is being used to replace synthetic materials, such as the plastics in car manufacture and conventional ballistic material in bullet-proof vests. Evelyn planted the concept of a wood culture, but it is maturing only in the early twenty-first century.

Balancing our demand for nature's wonder-material with the need to protect Earth from our industry, to grow food for our ever-increasing population, and to address the problems posed by pests and pathogens spread by global trade, presents an enormous challenge. The delightful prose and practical advice in *Sylva* continue to inspire 350 years on. ■

Gabriel Hemery *is a silvologist, author and photographer. He is chief executive of the Sylva Foundation. His first book,* The New Sylva, *will be published in April.*
e-mail: gabriel@sylva.org.uk



Engineers work to reconfigure CERN's ATLAS detector, one of two experiments to co-discover the Higgs boson.

CLAUDIA MARCELLONI/CERN

PARTICLE PHYSICS

Higgs on the big screen

Alexandra Witze savours a behind-the-scenes look at physics's most famous arrival.

How could there possibly be anything fresh to say about the Higgs boson, the subatomic particle whose 2012 discovery sparked a Nobel prize, a slew of popular books, an exhibition and even a zombie movie and a rap song? Remarkably, physicist-turned-filmmaker Mark Levinson pulls it off in a documentary about the Higgs shot where the particle was discovered.

It is hard to get distinctive footage when camera crews have been crawling over the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, for years. Levinson's edge is that he filmed on and off between 2008, when the LHC launched, and 2012. The *Particle Fever* team shot almost everything that counted: from the champagne-popping celebrations at the first beam of circulating protons, to the dirty and disfigured wrecks of the superconducting magnets that blew a week later, crippling the machine for months.

The film's other fresh take is its choice of protagonists. You won't hear much about the superstar researchers. Rather, the characters are everyday experimentalists and theoreticians caught up in the race to discovery. One of the most endearing is Monica Dunford, a talkative US postdoc who is forever throwing on a hard hat and taking a wrench to bits of the LHC that are not working. Hearing that one of the machine's detectors is five storeys tall is one thing — watching Dunford's face light up as she takes it all in is another.

Another key figure is CERN's Fabiola Gianotti, the former would-be philosopher who became spokeswoman for one of the two major Higgs-hunting experiments in 2009.

A parallel storyline follows a handful of theoretical physicists waiting for the results. The most eloquent is David Kaplan, the physicist who produced the film. It is Kaplan who handles the explainers on the Higgs during an academic lecture that efficiently dispenses scientific background. And it is Kaplan who films himself driving in the middle of the night to a party in Princeton, New Jersey, to watch the unveiling of the particle, missing a highway exit in his excitement. But Nima Arkani-Hamed steals the show on several levels. Slightly distracted and ever garbed in a red-and-black striped rugby shirt, he talks about his family escaping revolutionary Iran, and the solace that physics provided. Pacing late at night at the Institute for Advanced Study in Princeton, Arkani-Hamed is a kind of physics everyman, standing in for all the scientists straining to catch the news from the LHC.

The film's only stumble is a segment involving controversial ideas about whether many universes might co-exist — the multiverse. As data start to flow back from the LHC, tantalizing hints of the Higgs emerge, but the particle's mass is initially unclear. The film pits two possible Higgs masses against one another, each representing one possible explanation — either that the multiverse exists, or that every particle in the Universe has a shadowy 'supersymmetric' partner. This strategy provides tension but overemphasizes the possibility that a particular Higgs mass supports the multiverse idea. It might have been better to play up the competition between the two main LHC instrument teams. This drama is not apparent in *Particle Fever*, even in the climactic scene in which Gianotti and her counterpart on the second experiment unveil the Higgs discovery at an unforgettable seminar.

That is a quibble, however. The sense of scientific drive is palpable throughout the film, and even the coverage of the discovery seminar is fresh. Rather than chase the CERN director around the auditorium where the announcement was made, Levinson shows us Savas Dimopolous, a theorist at Stanford University in California, trying to talk his way past the guards at the auditorium door to be present for the historic moment. We see Dunford huddled in front of a computer screen, eyes fixed on the video feed as Gianotti announces that the Higgs has been found.

And we see a baby erupt into tears in the hallway outside the CERN auditorium — like the Higgs itself, a noisy newborn and a bringer of joy. ■

Particle Fever
DIRECTED BY MARK LEVINSON
On limited cinema release from 5 March 2014.

Alexandra Witze *is a correspondent for* Nature *in Boulder, Colorado.*

Correspondence

China must publicize its emissions reports

As China is the world's largest energy consumer and carbon dioxide emitter, the future trajectory of its carbon emissions will play a crucial part in global mitigation plans. However, China's National Assessment Reports on Climate Change, issued in 2007 and 2011, are limited in scope and not widely disseminated or cited. These shortcomings must be rectified before the next report is released this year.

Compared with the assessment reports by the Intergovernmental Panel on Climate Change (IPCC), China's have had scarcely any impact. The IPCC report is cited thousands of times in United Nations official documents, whereas China's seem to be barely mentioned even in Chinese-government documents.

In our view, this low impact reflects the general lack of public interest in climate change in China, the paucity of media coverage and scholarly study, and the insufficient efforts by central and local government to reduce energy consumption and greenhouse-gas emissions.

China's next assessment report needs to be more widely promoted, include more high-quality research results, and objectively evaluate current policies to tackle climate change in the country. China should learn from the IPCC and open the way to international collaboration in preparing and promoting the country's assessment reports.

Yuan-Feng Wang, Yu-Rong Zhang *Beijing Jiaotong University, China.*
cyfwang@bjtu.edu.cn

Biomedicine must look beyond *P* values

Establishing statistical validity for study findings goes beyond a consideration of *P* values alone (R. Nuzzo *Nature* **506,** 150–152; 2014). In the era of big data, we now have many biological measures available for assessing how likely findings are to be true positives.

This more-comprehensive approach has long been used by epidemiologists to address concerns about bias and causality: for example, in investigations of possible components of hypothetical disease-causing pathways (L. H. Kuller *et al. Am. J. Epidemiol.* **178,** 1350–1354; 2013). A way of inferring a causal association is to apply Hill's criteria, which seek ties between many factors, such as dose response, temporality and disease exposure (A. B. Hill *Proc. R. Soc. Med.* **58,** 295–300; 1965).

Advances in genomics and systems biology enhance our capacity for such investigations. We can now determine whether findings operate in a specific genotype context or fit biologically plausible pathways or networks — as was done in a re-evaluation of results from a genome-wide association study for multiple sclerosis (International Multiple Sclerosis Genetics Consortium *Am. J. Hum. Genet.* **92,** 854–865; 2013).

Anne-Louise Ponsonby *Murdoch Childrens Research Institute, Parkville, Victoria, Australia.*
anne-louise.ponsonby@mcri.edu.au
Terence Dwyer *International Agency for Research on Cancer, Lyon, France.*

Road maps of no use to some physicists

We do not believe that it is feasible for a single organization to draw up a 'road map' for future light- and neutron-source facilities in the way that CERN does for the particle-physics community (see P. G. Radaelli *Nature* **505,** 607–609; 2014). The diversity of the communities that use these facilities makes the centralization of scientific priorities impossible.

A range of research fields will benefit from Europe's investments in the X-ray Free Electron Laser and the European Spallation Source (ESS), and hundreds of scientists are collaborating to define ESS's capabilities and instrumentation (see go.nature.com/ip6afc).

Cassandras abound at the start of any large project that pushes the technological envelope, but the ESS is expected to boost scientific performance for neutron studies by as much as 300-fold, opening up entirely new fields of science.

Because the ESS forms part of a wider network of neutron and light sources, the discussion needed in Europe is how to leverage and integrate these sources most effectively to improve our research and economic environment.

Aleksandar Matic *Chalmers University of Technology, Gothenburg, Sweden.*
Peter Böni *Technische Universität München, Germany.*
Adrian Goldman *University of Leeds, UK.*
a.goldman@leeds.ac.uk

End education meddling in Nepal

The Nepalese government's backtracking last month on its political patronage of academia (*Nature* **506,** 279; 2014) raises hopes that appointments to top university positions will soon be made by non-political committees of scholars. Such a move would ensure the selection of senior academics on merit, rather than political affiliation; these individuals would help to attract much-needed funding for education and research.

The Nepalese government needs to create an educational system that is free of political meddling and nepotism. The country's young scientists will return from abroad only when a proper infrastructure is in place that will enable them to implement their skills and to realize their ambitions in a merit-based society.

Kosh P. Neupane *Tufts University, Medford, Massachusetts, USA.*
koshalnp@hotmail.com

Fight floods on a global scale

Accurate digital elevation models (DEMs) created using airborne lidar have transformed regional flood modelling and forecasting. At continental and global scales, however, the best-available DEMs come from satellite images and are too crude for simulating flooding — and its related risks to public health, biogeochemical cycling and wetland ecology. We would like to see industry, governments and humanitarian agencies come together to support the development of a global DEM with higher resolution and accuracy.

Current global DEMs cannot resolve the detail of terrain features that control flooding. More-effective flood-hazard maps could be created by obtaining high-resolution stereo images from satellites, combined with the latest advances in flood modelling using supercomputers. By 2050, worldwide annual losses due to flooding are predicted to reach US$1 trillion (S. Hallegatte *et al. Nature Clim. Change* **3,** 802–806; 2013). A global-scale DEM would have an enormous impact on finance (such as flood re-insurance), humanitarian services (such as disaster relief) and scientific research.

The advanced global DEM would use existing lidar data and stereo satellite images, and new lidar elevation data would be acquired on board disaster-relief aircraft or on drones deployed over flood plains. The operation costs would therefore be substantially cheaper than most satellite missions.

Guy J.-P. Schumann* *NASA Jet Propulsion Laboratory, Pasadena, California, USA.*
guy.j.schumann@jpl.nasa.gov
**On behalf of 4 co-signatories; see go.nature.com/j1pchz for full list.*

Air pollution and forest water use

ARISING FROM T. F. Keenan *et al.* *Nature* **499,** 324–327 (2013)

Forests in North America and northern Europe increased their water-use efficiency (WUE)—the ratio of photosynthetic CO_2 uptake to water loss through evapotranspiration—over the last two decades, according to a recent Letter[1]. Keenan *et al.* attribute the rising WUE to fertilization by increasing levels of atmospheric CO_2 (ref. 1), although biosphere models predict this effect to be much smaller than the observed trend. Here, I show that falling concentrations of ozone and other phytotoxic air pollutants, which were not considered in ref. 1, may explain part of the WUE trend. Future efforts to reconcile biosphere models with field data should, therefore, use integrated modelling approaches that include both air quality and CO_2 effects on forest growth and water use. There is a Reply to this Brief Communication Arising by Keenan, T. F. *et al.* *Nature* **507,** http://dx.doi.org/10.1038/nature13114 (2014).

Tree injuries caused by ozone, the most phytotoxic air pollutant—including visible foliar injury, reduced photosynthesis and diminished biomass—depress global ecosystem productivity[2] and are well documented in field observations from North America and Europe[3,4]. Ozone enters leaves through stomata and causes internal oxidative stress and membrane damage that reduce photosynthetic CO_2 assimilation[5,6]. During ozone injury, transpiration usually falls less than does photosynthesis, but transpiration can sometimes rise because of ozone injury to stomata[6–8]. In either case WUE declines.

Surface ozone concentrations during the summer growing season have fallen significantly in eastern North America and modestly in northern Europe owing to emission controls on vehicles and industrial sources of ozone precursors[9,10]. Figure 1 shows ozone trends in regions around the rural forest sites analysed in ref. 1, evaluated as summer daytime-mean mole fraction (Fig. 1a) and as the accumulated concentration over a threshold of 40 nmol mol^{-1} (AOT40, defined as in the literature[11,12]), which is a common predictor for plant injury (Fig. 1b). I calculated both ozone metrics using only rural sites—from the US Clean Air Status and Trends Network (CASTNET; http://epa.gov/castnet) and the European Monitoring and Evaluation Programme (EMEP; http://www.nilu.no/projects/ccc/emepdata.html)—reporting at least 14 years of hourly ozone data during the period 1995–2010 (Fig. 1c and d). By either metric, ozone significantly decreased at all sites in the midwestern USA ($n = 11$, $P < 0.001$–0.02 for Kendall's τ test) and northeastern USA ($n = 5$, $P = 0.001$–0.004). For averages over all sites within each region, AOT40 fell by half in the period 1995–2010 in both regions ($P < 0.002$). Over northern Europe most sites had negative trends, but with smaller magnitudes, consistent with other recent analyses[10].

The first-order effect of these ozone trends in the Midwest, using sensitivities for broad-leaf trees[6,12,13], would be a 0.6% annual increase in biomass accumulation and a 0.3% annual improvement in WUE. In addition, partial closure of stomata in response to rising CO_2 (ref. 14) and rising vapour pressure deficit[1] reduces leaf uptake of ozone by approximately 0.9% per year regardless of ozone trends. Combining all these effects, improvements in ozone air quality over the period 1995–2010 probably increased forest WUE by approximately 0.33% per year in the midwestern USA and slightly less in the northeastern USA. Using the range of ozone sensitivities reported for tree species[12,13,15], the ozone effect on WUE in the midwestern USA could be 0.1–0.8% per year. This predicted ozone effect is about one-sixth of the observed WUE trend (2% per year, calculated from the Supplementary Information to ref. 1) and larger than the mean simulated effect of CO_2 fertilization in the terrestrial biosphere models surveyed by ref. 1. Measuring ozone mole fractions and fluxes into the forest canopy simultaneously with

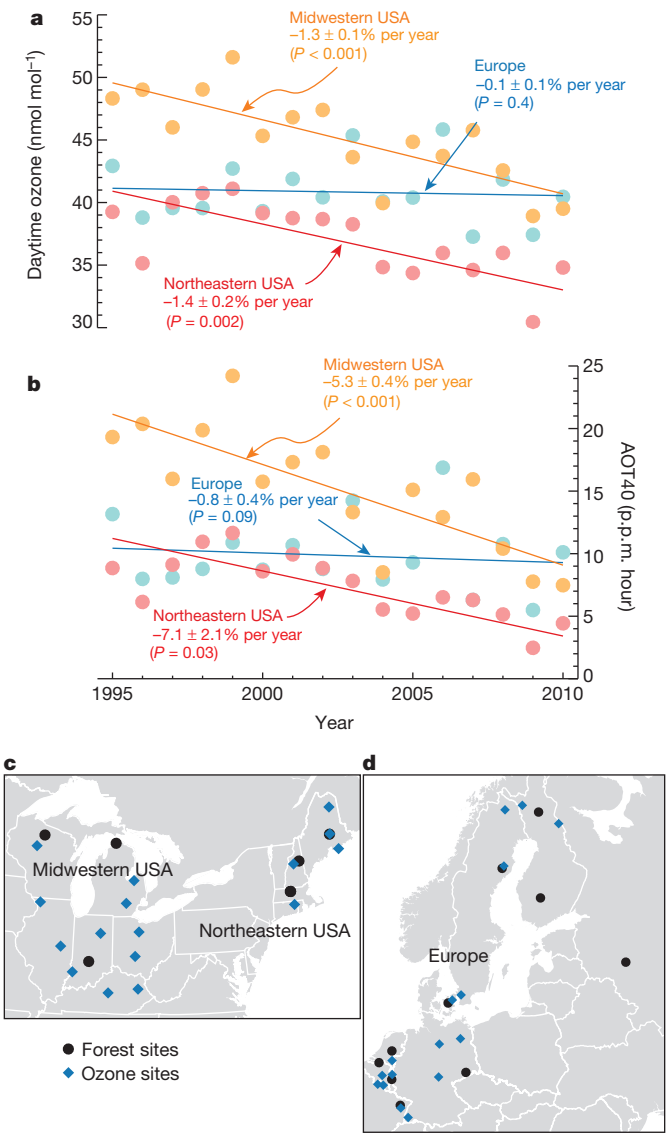


Figure 1 | Trends in ozone exposure metrics that correlate with tree injury. **a**, Daytime-mean ozone mole fraction; **b**, AOT40. Both metrics are calculated in April–September of each year during the hours 8:00–20:00 (local time) at rural sites in the USA (**c**) and Europe (**d**) near forest stations that monitor WUE. Lines show the mean trends (Sen's method) averaged across all stations within each region (±1 standard error, *P* values from Student's *t*-test). The unusually high mole fraction and AOT40 values in Europe in 2003 and 2006 were caused by extreme heatwaves.

WUE should constrain the effect, but the variability of WUE trends across sites and years illustrates that ozone data from multiple forests and many years are necessary to obtain robust results. In addition to the decline in ozone concentration, the concentrations of the air pollutants NO_x and SO_2, which also harm WUE both individually and through synergistic effects with ozone, have fallen quickly but the effects are not included here[11]. Thus, the benefits of improved air quality to forest productivity and WUE may be larger than I have estimated. Keenan *et al.*[1]

suggest that current terrestrial biosphere models underestimate the impact of CO_2 fertilization on WUE. The calculations here show that ozone trends help to reconcile the large differences between models and observations.

Methods

I calculated photosynthesis reductions from ozone AOT40 trends (−0.8 parts per million (p.p.m.) hours per year, where 1 p.p.m. = 1 μmol mol^{-1}, for the midwestern USA) using empirical correlations with ozone exposure for young broad-leaf trees (−0.7% per p.p.m. hour, for beech, birch and maple)[12,13]. Other tree species may be more (poplar) or less (conifers, oak) sensitive: −1.8% to −0.2% per p.p.m. hour (refs 12 and 15). Ozone-induced WUE changes are half those of photosynthesis and of the same sign[6]. Rising CO_2 (2 p.p.m. per year) and rising vapour pressure deficit (11 Pa per year; ref. 1) reduced stomatal conductance and ozone uptake by approximately 0.4% per year and 0.5% per year, respectively, based on empirical sensitivity factors[14] (conductance changes are −0.2% per p.p.m. of CO_2 and −0.05% per Pa).

Christopher D. Holmes[1]
[1]Department of Earth System Science University of California, Irvine, California 92697, USA.
email: cdholmes@uci.edu

Received 1 October 2013; accepted 22 January 2014.

1. Keenan, T. F. *et al.* Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **499,** 324–327 (2013).
2. Sitch, S., Cox, P. M., Collins, W. J. & Huntingford, C. Indirect radiative forcing of climate change through ozone effects on the land-carbon sink. *Nature* **448,** 791–794 (2007).
3. Chappelka, A. H. & Samuelson, L. J. Ambient ozone effects on forest trees of the eastern United States. *New Phytol.* **139,** 91–108 (1998).
4. Skärby, L., Ro-poulsen, H., Wellburn, F. A. M. & Sheppard, L. J. Impacts of ozone on forests: a European perspective. *New Phytol.* **139,** 109–122 (1998).
5. Ainsworth, E. A., Yendrek, C. R., Sitch, S., Collins, W. J. & Emberson, L. D. The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annu. Rev. Plant Biol.* **63,** 637–661 (2012).
6. Wittig, V. E., Ainsworth, E. A., Naidu, S. L., Karnowky, D. F. & Long, S. P. Quantifying the impact of current and future tropospheric ozone on tree biomass, growth, physiology and biochemistry: a quantitative meta-analysis. *Glob. Change Biol.* **15,** 396–424 (2009).
7. McLaughlin, S. B., Nosal, M., Wullschleger, S. D. & Sun, G. Interactive effects of ozone and climate on tree growth and water use in a southern Appalachian forest in the USA. *New Phytol.* **174,** 109–124 (2007).
8. Lombardozzi, D., Levis, S., Bonan, G. & Sparks, J. P. Predicting photosynthesis and transpiration responses to ozone: decoupling modeled photosynthesis and stomatal conductance. *Biogeosciences* **9,** 3113–3130 (2012).
9. Cooper, O. R., Gao, R.-S., Tarasick, D., Leblanc, T. & Sweeney, C. Long-term ozone trends at rural ozone monitoring sites across the United States, 1990–2010. *J. Geophys. Res.* **117,** D22307 (2012).
10. Tørseth, K. *et al.* Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972–2009. *Atmos. Chem. Phys.* **12,** 5447–5481 (2012).
11. WHO. *Air Quality Guidelines for Europe* 2nd edn, 219–251, http://www.euro.who.int/__data/assets/pdf_file/0005/74732/E71922.pdf (World Health Organization Regional Office for Europe, 2000).
12. Karlsson, P. E. *et al.* New critical levels for ozone effects on young trees based on AOT40 and simulated cumulative leaf uptake of ozone. *Atmos. Environ.* **38,** 2283–2294 (2004).
13. Calatayud, V., Cerveró, J. & Sanz, M. J. Foliar, physiological and growth responses of four maple species exposed to ozone. *Wat. Air Soil Pollut.* **185,** 239–254 (2007).
14. Franks, P. J. *et al.* Sensitivity of plants to changing atmospheric CO_2 concentration: from the geological past to the next century. *New Phytol.* **197,** 1077–1094 (2013).
15. Hoshika, Y., Omasa, K. & Paoletti, E. Whole-tree water use efficiency is decreased by ambient ozone and not affected by O_3-induced stomatal sluggishness. *PLoS ONE* **7,** e39270 (2012).

Competing Financial Interests Declared none.

doi:10.1038/nature13113

Keenan *et al.* reply

REPLYING TO C. D. Holmes *Nature* **507,** http://dx.doi.org/10.1038/nature13113 (2014)

Forests have become more efficient at using water over the past two decades[1]. A series of hypotheses exist to explain this trend, but the only credible explanation to date is a response to rising atmospheric CO_2. Keenan *et al.*[1] show that the observed trend is physiologically plausible, but is much larger than expected from conventional theory and experimental evidence. This has led to suggestions that processes other than increased atmospheric CO_2 may have contributed to the observed trend[2]. One such process that has yet to be examined is the effect of tropospheric ozone on forest water-use efficiency (WUE). In the accompanying Comment[3], Holmes reports that ozone concentrations have declined in the northeastern and midwestern USA by about 50% from 1995 to 2010. Using empirical relationships, he estimates that this decline could explain roughly 15% of the reported increase in WUE over North America, and a significantly lower proportion of the trend in Europe.

As a preliminary test of the 'ozone hypothesis', we analyse 20 years of ozone concentration measurements at Harvard forest, in Massachusetts, USA, which were made concurrently with the carbon and water fluxes from which we derived WUE. In agreement with results presented by Holmes[3], extreme ozone concentrations have declined at this forest over the past two decades. Although the 50th percentile of ozone levels has stayed relatively constant, both the 95th percentile and the AOT40 metric show declining trends over the time period ($P = 0.09$ and 0.11, respectively; Kendall's τ test, Fig. 1a).

Despite the declining trend, we found no significant ($P = 0.46$, $r = -0.19$) relationship between annual means of WUE and the occurrence of high ozone concentrations (Fig. 1b). The observations at Harvard forest, therefore, do not support the claim that WUE is being affected by changes in ozone concentrations. That said, we acknowledge that it would be difficult to detect and attribute a change responsible for 15% of the trend we observe given the large influence of other factors.

As an additional test of the ozone hypothesis, we consider the trends of the component fluxes of WUE, published in ref. 1. Decreasing ozone concentrations are primarily expected to increase leaf photosynthesis, with stomatal conductance typically increasing to a lesser extent[3], although conductance responses vary[4]. We would therefore expect a similar response of photosynthesis and conductance in the data of ref. 1. In ref. 1, we report increasing photosynthesis at only 50% of the sites, whereas stomatal conductance showed large declines consistently across all sites.

It is worth noting that global ozone concentrations are increasing[5]. There is no significant trend in ozone concentrations in Europe[3], where half of the sites used in ref. 1 are located. Within the USA, trends vary greatly by region. Ozone in the western USA has increased over the past two decades owing to increased levels of precursors from Asia[5,6]. Globally, ozone concentrations show large declines only in the midwestern and eastern USA[3,7], where 8 of our 21 sites are based. We therefore agree that it is possible, following the calculations of Holmes[3], that changes in ozone concentrations contributed to a small proportion (roughly 15%) of the trend at those sites.

In conclusion, we can neither reject nor accept the ozone hypothesis of Holmes[3], although it is clear that the observed changes in WUE are not primarily driven by changes in ozone concentrations. His estimate[3] of an ozone-induced increase in WUE of 0.33% per year is probably an upper bound for the response at sites in the USA, but an overestimation of the response at European sites. That said, the changes in ozone

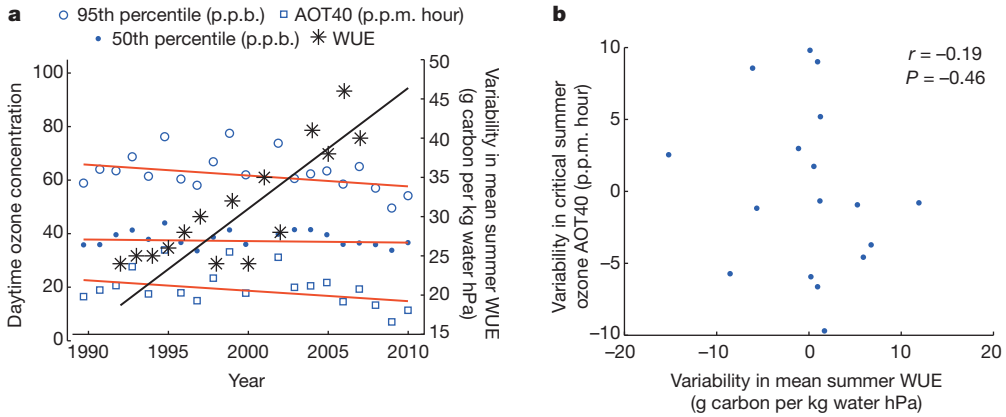


Figure 1 | Ozone and WUE at Harvard Forest. a, Long-term trends in ozone concentrations (p.p.b., parts per billion; p.p.m., parts per million) and WUE at Harvard forest, with trends (lines) estimated using the Sen-slope method. **b**, The relationships between variability in critical summer ozone concentrations (AOT40) and variability in mean summer WUE values at Harvard forest from 1992 to 2010.

concentrations reported by Holmes[3] are large in some heavily forested and productive regions, and may beneficially affect ecosystem function. We therefore agree with Holmes[3] that more work is needed to assess the impact of air quality on ecosystems globally, and the resulting change in ecosystem function.

Trevor F. Keenan[1], David Y. Hollinger[2], Gil Bohrer[3], Danilo Dragoni[4], J. William Munger[5], Hans Peter Schmid[6] & Andrew D. Richardson[7]
[1]Department of Biological Sciences, Macquarie University, North Ryde, New South Wales 2109, Australia.
email: trevor.keenan@mq.edu.au
[2]USDA Forest Service, Northern Research Station, Durham, New Hamphire 03824, USA.
[3]Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, Ohio 43210, USA.
[4]Department of Geography, Indiana University, Bloomington, Indiana 47405, USA.
[5]School of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.
[6]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen 82467, Germany.
[7]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

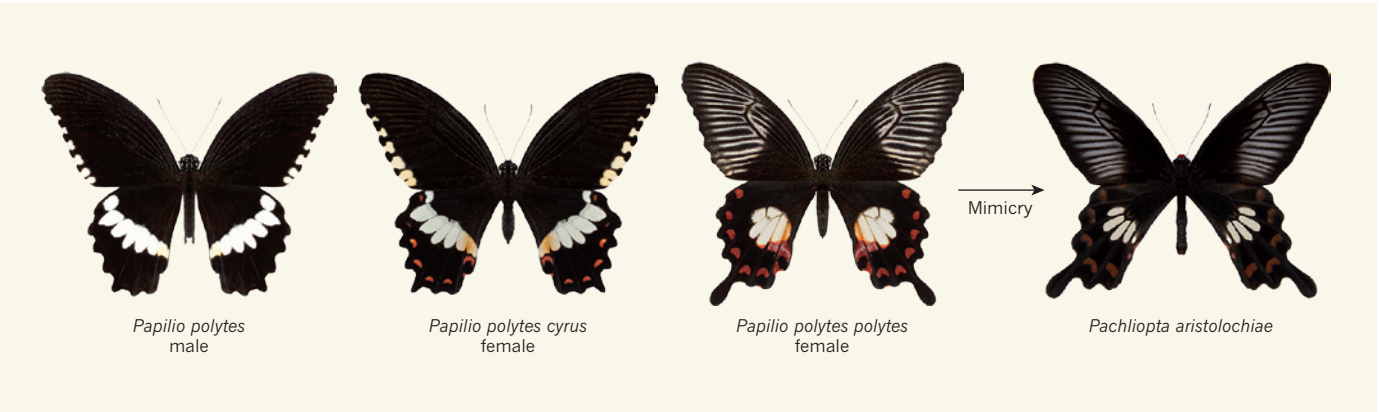
1. Keenan, T. F. *et al.* Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **499,** 324–327 (2013).
2. Medlyn, B. E. & De Kauwe, M. Carbon dioxide and water use in forest plants. *Nature* **499,** 287–289 (2013).
3. Holmes, C. D. Air pollution and forest water use. *Nature* **507,** http://dx.doi.org/10.1038/nature13113 (2014).
4. Lombardozzi, D., Sparks, J. P. & Bonan, G. Integrating O_3 influences on terrestrial processes: photosynthetic and stomatal response data available for regional and global modeling. *Biogeosciences* **10,** 6815–6831 (2013).
5. Parrish, D. D. *et al.* Long-term changes in lower tropospheric baseline ozone concentrations at northern mid-latitudes. *Atmos. Chem. Phys.* **12,** 11485–11504 (2012).
6. Cooper, O. R. *et al.* Increasing springtime ozone mixing ratios in the free troposphere over western North America. *Nature* **463,** 344–348 (2010).
7. Cooper, O. R., Gao, R.-S., Tarasick, D., Leblanc, T. & Sweeney, C. Long-term ozone trends at rural ozone monitoring sites across the United States, 1990–2010. *J. Geophys. Res.* **117,** D22307 (2012).

doi:10.1038/nature13114

EVOLUTIONARY BIOLOGY

Sex, lies and butterflies

Variation in an evolutionarily conserved sexual-differentiation gene, *doublesex*, has been found to explain how females of one species of butterfly mimic the colour patterns of several toxic species to avoid predation. SEE LETTER P.229



KRUSHNAMEGH KUNTE

Figure 1 | Female-specific mimicry. Males of the swallowtail butterfly species *Papilio polytes* exist in one form, but several female forms co-occur in the same population. Females with the *cyrus* form look like males, whereas other female forms mimic the colour patterns of distantly related toxic species, such as the *polytes* form, which resembles *Pachliopta aristolochiae*.

DAVID W. LOEHLIN & SEAN B. CARROLL

Biological mimicry, in which one species gains an advantage by closely resembling another, unrelated species, is one of the most spectacular phenomena in nature. Expert impostors such as the cuckoo, milk snake and bee orchid have long fascinated naturalists and have played an important part in the development of evolutionary theory. Indeed, in the years immediately after the publication of Charles Darwin's *On the Origin of Species*, Henry Walter Bates[1] and Alfred Russel Wallace[2] recognized that butterfly mimicry was an obvious adaptation that could be explained only by natural selection. For the next 150 years, however, the mechanisms that generate these striking patterns eluded biologists.

But no longer — there has been a burst of breakthroughs[3,4] in this long-standing mystery. On page 229 of this issue, Kunte *et al.*[5] reveal that the remarkable ability of females of a swallowtail butterfly species to closely match the colour patterns of several unrelated butterflies is due to variation at a single genetic region: the butterfly version of the well-studied *doublesex* regulatory gene.

In the classic case of 'Batesian' mimicry[1], the warning colours of unpalatable or toxic butterflies are co-opted by non-toxic free-rider species. Among some swallowtail butterflies (genus *Papilio*), Wallace described an intriguing twist in which mimicry is limited to females[2]. Further studies showed that several discrete female forms, each resembling the warning colour pattern of a different toxic butterfly, often co-occur in a population alongside non-mimetic females and males[6] (Fig. 1). Although it is still not known why one species can maintain several different mimetic and non-mimetic patterns, the inheritance of this variation is surprisingly simple[2]: female colour patterns stay intact in genetic crosses within, but not between, populations, with each pattern assorting as one of two possible variants from a single genetic locus[6].

Kunte *et al.* bring swallowtail Batesian mimicry into the molecular era by showing that the differences between female mimetic forms in *Papilio polytes* are tightly associated with genetic variation around the *doublesex* locus. This gene is a particularly satisfying explanation for the evolution of sex-specific colour patterns, because genes of the *Dmrt* family (which includes *doublesex*) control aspects of sexual differentiation in most animals[7]. The *doublesex* gene basically acts as a switch. Specifically, in the best-studied *doublesex* gene (that of the fruit fly *Drosophila melanogaster*), different RNA transcripts are produced in males and females by a process known as alternative splicing. The male and female transcripts encode distinct protein isoforms that are thought to activate or repress different sets of genes, leading to sex-specific differentiation[7,8].

How could intricate wing-pattern variation derive from this single genetic signal? In principle, different female wing-pattern gene variants could derive from mutations that alter *doublesex* transcription, splicing or protein structure. Kunte and colleagues report that the swallowtail *doublesex* transcripts are also alternatively spliced in different sexes, but they find no evidence for splicing differences between mimetic forms. Rather, they find several mutations in protein-coding sequences, and speculate that these could alter the structure and activity of the Doublesex protein.

However, the authors also make the intriguing observation that colour stripes in the forewings of mimetic females are accompanied by a striped pattern of Doublesex expression. This raises the strong possibility that changes to this pattern of *doublesex* expression are the cause of the different mimetic forms. This inference is grounded in insight into the mechanics of the *doublesex* gene in other insects. Specifically, rather than signalling 'male' or 'female' in every cell, *doublesex* is elaborately regulated and active in only certain cell populations, including those that make structures that differ between the sexes[8,9]. Indeed, evolutionary changes to regulatory sequences of the *doublesex* locus have reshaped the wings of male wasps[10], and shifts in *doublesex* expression have changed the position of sexually dimorphic structures in flies[9]. Therefore,

broadening of *doublesex* expression in the swallowtail to a different part of the wing might be sufficient to expand a pre-existing female-specific colour pattern and generate a new mimetic form that could then persist owing to the selective advantage it conveys.

The intricate patterns of *doublesex* expression also help to explain how such apparently complex morphological variation could map to a single genetic locus. The mimicry loci in *P. polytes* and other butterflies have been proposed to be 'supergenes' — linked clusters of trait-altering genes — because of the complexity of the colour pattern and the rare occurrence of individuals with mixed mimetic patterns[8]. Like other developmental regulatory genes, *doublesex* probably has multiple transcription-regulating elements (enhancers). In principle, different elements could control *doublesex* expression in different parts of the swallowtail wing, and genetic variation at two elements should occasionally separate when chromosomes cross over during gamete formation. It is possible that other supergenes are also well-known developmental regulatory genes that have accumulated multiple functional mutations in evolution.

By accomplishing the arduous task of gene mapping in a butterfly, Kunte *et al.* open the door to understanding the mechanics of how the insect's mimetic pattern is generated and how each wing variant is maintained in a population. Identifying the precise molecular mechanism behind this spectacular mimicry switch promises to be exciting, whether it is due to regulatory mutations, protein alterations or a combination of the two, especially in light of the central role of *Dmrt* genes in sexual dimorphism across the animal kingdom. ■

David W. Loehlin *and* **Sean B. Carroll**
are in the Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.
e-mails: loehlin@wisc.edu; sbcarrol@wisc.edu

1. Bates, H. W. *Trans. Linn. Soc. Lond.* **23,** 495–566 (1862).
2. Wallace, A. R. *Trans. Entomol. Soc. Lond.* **2,** 14–15 (1864).
3. Reed, R. D. *et al. Science* **333,** 1137–1141 (2011).
4. Martin, A. *et al. Proc. Natl Acad. Sci. USA* **109,** 12632–12637 (2012).
5. Kunte, K. *et al. Nature* **507,** 229–232 (2014).
6. Clarke, C. A. & Sheppard, P. M. *Phil. Trans. R. Soc. B* **263,** 431–458 (1972).
7. Kopp, A. *Trends Genet.* **28,** 175–184 (2012).
8. Robinett, C. C., Vaughan, A. G., Knapp, J.-M. & Baker, B. S. *PLoS Biol.* **8,** e1000365 (2010).
9. Tanaka, K., Barmina, O., Sanders, L. E., Arbeitman, M. N. & Kopp, A. *PLoS Biol.* **9,** e1001131 (2011).
10. Loehlin, D. W. *et al. PLoS Genet.* **6,** e1000821 (2010).

This article was published online on 5 March 2014.

ASTROPHYSICS

Cosmic lens reveals spinning black hole

The power of a cosmic lens to magnify and split the light from a distant, mass-accreting giant black hole into four components has allowed researchers to measure the black hole's spin. SEE LETTER P.207

GUIDO RISALITI

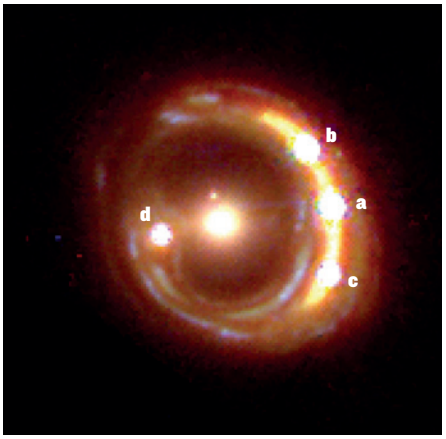
Quasars are the most powerful, continuously emitting sources of radiation in the Universe. They reside at the centre of a small fraction of galaxies, and are powered by supermassive black holes, which have masses millions to billions times greater than that of the Sun. Although giant black holes are present in most — possibly all — galaxies, not all of them are in an active state, in which they accrete gas from a surrounding disk. In fact, most of these objects are in a quiescent phase. It is the active type of supermassive black hole that drives quasars. The formation history of supermassive black holes is thought to be closely tied to that of their host galaxies, but how exactly they form and grow remains unclear. In this issue, Reis *et al.*[1] (page 207) describe how a cosmic lens has enabled them to find that a supermassive black hole powering a distant quasar has grown through coherent, rather than random, episodes of mass accretion.

Astronomers believe that supermassive black holes formed in the early Universe from small 'seeds' with masses of up to 10,000 solar masses. These seeds would have then grown to reach millions to billions of solar masses either through multiple mergers during galaxy collisions or through gas accretion from their host galaxies; this accretion would have consisted either of many short, unrelated accretion episodes or of fewer, longer and ordered accretion phases. Different models of galaxy evolution predict a different mix of these processes, so reconstructing the formation history of giant black holes would provide a way for us to understand how galaxies evolved.

Supermassive black holes are simple systems. They are characterized by just two quantities, their mass and their angular momentum (spin). Whereas the total amount of accretion and any mergers that a supermassive black hole undergoes are encoded in its mass, how this mass was assembled is encoded in its spin[2]. A few ordered accretion events or mergers of large black holes produce high spins, and short, random accretion processes produce low spins. Measuring these spins is therefore a major goal of extragalactic astronomy: the spins of supermassive black holes hold a key to understanding the evolution of their host galaxies.

But how can we measure the spins? According to Einstein's general theory of relativity, a black hole's gravitational field twists space-time around it. Such twisting depends on the black hole's spin, so measuring the twisting allows the spin to be estimated. The signature of space-time distortion is imprinted on the emission of radiation from regions close to the black hole's event horizon — the surface beyond which no radiation can escape. In quasars, the bulk of the huge, observed luminosity is emitted by the accretion disk at optical and ultraviolet wavelengths. However, this primary emission is nearly featureless, so, despite its vicinity to the event horizon, it does not provide an easy means to detect space-time distortions. The best way to perform such a measurement is to observe X-rays reflected by the disk.

The main source of X-ray emission in quasars is believed to be a compact cloud of hot electrons in the inner part of the black hole's accretion disk. Some of this radiation illuminates the accretion disk and is reflected towards the observer's line of sight. This reflected emission usually accounts for less than 1% of the total energy produced by quasars, but contains narrow spectral features — most notably, an iron spectral line at the object's



ACS & NICMOS/ESA/HST/STSCI/AURA/NASA/REF. 10

Figure 1 | A quadruple quasar. Reis and colleagues' analysis[1] of a distant quasar, whose light is magnified and split into four components (**a**–**d**) by the gravitational field of a foreground galaxy (central object), has enabled them to calculate the spin of the supermassive black hole that powers the quasar.

rest-frame energy of 6.4–7 kiloelectronvolts — the shape of which is strongly altered by the space-time warping around the black hole[3–5]. The shape of these features can be measured in high-quality X-ray spectra, providing a measurement of the spins of supermassive black holes[6–8].

This type of analysis is at the heart of Reis and colleagues' work. Until now, astronomers have struggled to obtain unambiguous spin measurements using this method. The X-ray spectra of active galactic nuclei are quite complex, and the reflection component, which contains the signatures of space-time distortions, is relatively weak. Moreover, certain absorption features can mimic the distortions[9]. As a result, only long observations of a few very bright sources in the local Universe made with the most powerful X-ray observatories — NASA's Chandra, Europe's XMM-Newton, Japan's Suzaku and, more recently, NASA's NuSTAR — have provided convincing results[6–8]. In their study, Reis *et al.* break new ground by obtaining a spin measurement of a quasar at a distance of more than 6 billion light years from Earth, from a time when the Universe was about half its current age.

This remarkable result was possible owing to the exceptional nature of the observed source — a quadruply imaged, gravitationally lensed quasar (Fig. 1). The light from the distant quasar is both magnified and split into four different images by the gravitational field of a foreground elliptical galaxy (the lens) that, by chance, is on the line of sight of the quasar. For this reason, the authors could analyse four 'copies' of the X-ray spectrum of the quasar, each with an intensity significantly magnified by the lens. The resulting X-ray spectra have a quality that matches the best that has been obtained for nearby sources, and allowed a robust measurement of the black hole's spin. As it turns out, the spin is large (close to the highest possible value that theory predicts), suggesting that the black hole acquired its mass through coherent phases of mass accretion.

Although X-ray spectra of a quality comparable to that obtained here cannot be currently obtained for standard, non-lensed sources at similar distances, Reis and colleagues have opened a window on what astronomers could observe with the next generation of X-ray telescopes, such as Europe's proposed ATHENA mission. ■

Guido Risaliti *is at the INAF-Osservatorio Astrofisico di Arcetri, 50125 Florence, Italy, and at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA.*
e-mail: risaliti@arcetri.astro.it

1. Reis, R. C., Reynolds, M. T., Miller, J. M. & Walton, D. J. *Nature* **507,** 207–209 (2014).
2. Volonteri, M., Madau, P., Quataert, E. & Rees, M. J. *Astrophys. J.* **620,** 69–77 (2005).
3. Fabian, A. C., Rees, M. J., Stella, L. & White, N. E. *Mon. Not. R. Astron. Soc.* **238,** 729–736 (1989).
4. Dauser, T., Wilms, J., Reynolds, C. S. & Brenneman, L. W. *Mon. Not. R. Astron. Soc.* **409,** 1534–1540 (2010).
5. Dovčiak, M., Karas, V. & Yaqoob, T. *Astrophys. J. Suppl.* **153,** 205–221 (2004).
6. Brenneman, L. W. & Reynolds, C. S. *Astrophys. J.* **652,** 1028–1043 (2006).
7. Walton, D. J. *et al. Mon. Not. R. Astron. Soc.* **428,** 2901–2920 (2013).
8. Risaliti, G. *et al. Nature* **452,** 449–451 (2013).
9. Miller, L., Turner, T. J. & Reeves, J. N. *Astron. Astrophys.* **483,** 437–452 (2008).
10. Claeskens, J.-F., Sluse, D., Riaud, P. & Surdej, J. *Astron. Astrophys.* **451,** 865–879 (2006).

This article was published online on 5 March 2014.

GEOLOGY

Earth's deep water reservoir

A tiny sample of a mineral included in a diamond confirms predictions from high-pressure laboratory experiments that a water reservoir comparable in size to all the oceans combined is hidden deep in Earth's mantle. **SEE LETTER P.221**

HANS KEPPLER

How well do we know what lies deep inside Earth, more than 500 kilometres below our feet? Surprisingly well, according to a paper by Pearson *et al.*[1] on page 221 of this issue. The authors describe the first sample of an unusual mineral from Earth's mantle transition zone, which is located between depths of 410 and 660 km. Their study suggests that the sample is rich in water, supporting the hypothesis that the transition zone is a hydrous region.

Observations of earthquakes show that the velocity of seismic waves changes abruptly at some discontinuities that separate the upper mantle from the transition zone and the lower mantle. Geologist Alfred E. Ringwood pioneered the idea that these discontinuities are due to phase changes of the mineral olivine — $(Mg,Fe)_2SiO_4$ — which makes up most of the upper mantle. Some of these changes result in high-pressure phases, including one with a crystal structure known as spinel structure[2]. When spinel-structure olivine was subsequently found in meteorites that had experienced high shock pressures during collisions in space, this polymorph was fittingly named ringwoodite[3].

Until now, nobody had ever seen ringwoodite from Earth's mantle, although geophysicists were sure that it must exist. Most people (including me) never expected to see such a sample. Samples from the transition zone and lower mantle are exceedingly rare and are only found in a few, unusual diamonds. But even inside a diamond, the decrease in ambient pressure as the diamond rises towards Earth's surface should normally cause ringwoodite to convert back into olivine. However, in the diamond studied by Pearson and colleagues, ringwoodite remained in its original structure. This is an amazing finding, and suggests that the transport of the diamond to the surface must have been extremely rapid — possibly caused by an explosive volcanic eruption that was fed directly by magma produced in the transition zone.

BRITISH LIBRARY/ROBANA VIA GETTY

Figure 1 | A view of the deep Earth. This illustration by Édouard Riou, from Jules Verne's novel *Journey to the Centre of the Earth* (1864), shows how the French novelist imagined an ocean in Earth's interior.

For a long time, it was believed that all the water that might once have been in Earth's interior was released from it by volcanic eruption over geological time and now resides in the oceans. This view changed after a study[4] pointed out that wadsleyite — another high-pressure polymorph of olivine first synthesized by Ringwood's group[5] — has an unusual crystal structure that should make it a potential phase for storing water. This prediction was later confirmed in many experimental studies, and one investigation[6] not only found plenty of

RICHARD SIEMENS/UNIV. ALBERTA

Figure 2 | A diamond with a ringwoodite inclusion. Pearson *et al.*[1] have discovered a microscopic sample of ringwoodite, a polymorph of the mineral olivine, in this diamond from Juína, Brazil. The diamond is 5 millimetres across in its longest dimension.

water in wadsleyite produced under high pressure in the laboratory, but also discovered up to 2 weight per cent water in ringwoodite. Water in ringwoodite? This sounded implausible because the mineral has a spinel structure, and naturally occurring spinels do not tolerate any water in the form of hydroxyl (OH) groups. However, subsequent studies confirmed that ringwoodite has a high capacity to store water, similar to wadsleyite. Negatively charged Mg^{2+} vacancies in the structure seem to be charge balanced by the attachment of protons (H^+ ions) to oxygen atoms. The OH groups created in this way represent 'water' that is chemically dissolved in the crystal structure.

The fact that plenty of water can be dissolved in ringwoodite in high-pressure experiments does not, of course, necessarily mean that ringwoodite in the mantle also contains water. But, because water solubility in ringwoodite is much higher than in other minerals, such as olivine, thermodynamics predicts that the transition zone, where ringwoodite is stable, should be greatly enriched in water compared with the upper mantle[7]. This idea of a hydrous transition zone has excited Earth scientists over many years, and various attempts[8–10] have been made to infer water content from observations such as remote sensing of electrical conductivities in the transition zone and the precise depth of seismic discontinuities. But the results have been mixed, partly because any property of the mantle that can be measured will depend not only on the water content of the minerals present, but also on other parameters — most notably, temperature. It would be nice to have a sample of a mineral from the transition zone for which the water content could be measured in the laboratory.

Pearson and colleagues' study provides just such a sample. The authors find that the infrared spectra of their sample in the wavelength region where the OH groups absorb this radiation are strikingly similar to those of synthetic samples with water content of about 1 wt%, near the 2 wt% upper limit discussed earlier. If this sample were representative of the entire depth range of the lower part of the transition zone, between 520 and 660 km, where ringwoodite is stable, it would translate to a total of 1.4×10^{21} kg of water — about the same as the mass of all the world's oceans combined. In some ways, it is an ocean in Earth's interior, as visualized by Jules Verne in his 1864 novel *Journey to the Centre of the Earth* (Fig. 1), although not in the form of liquid water, but as OH groups in an unusual mineral.

Although Pearson and co-workers confirmed the occurrence, and the high water content, of ringwoodite in a diamond (Fig. 2) from the transition zone, it is not entirely clear how representative that sample is of the whole transition zone. Diamonds are usually brought to the surface by 'kimberlite' magmas, which feed extremely explosive volcanic eruptions[11]. No such kimberlite eruption has ever been recorded, but there is evidence that kimberlite magma is extremely rich in volatile components such as water and carbon dioxide, and probably taps an unusually water-rich part of the upper mantle somewhere above the transition zone. If the source of the kimberlite magma is an unusual mantle reservoir, there is the possibility that, at other places in the transition zone, ringwoodite contains less water than the sample found by Pearson and colleagues. However, in light of this sample, models with anhydrous or water-poor transition zones seem rather unlikely. ■

Hans Keppler *is at Bayerisches Geoinstitut, University of Bayreuth, 95440 Bayreuth, Germany.*
e-mail: hans.keppler@uni-bayreuth.de

1. Pearson, D. G. *et al. Nature* **507,** 221–224 (2014).
2. Ringwood, A. E. *Am. Mineral.* **44,** 659–661 (1959).
3. Binns, R. A., Davis, R. J. & Reed, S. J. B. *Nature* **221,** 943–944 (1969).
4. Smyth, J. R. *Am. Mineral.* **72,** 1051–1055 (1987).
5. Ringwood, A. E. & Major, A. *Earth Planet. Sci. Lett.* **1,** 241–245 (1966).
6. Kohlstedt, D. L., Keppler, H. & Rubie, D. C. *Contrib. Mineral. Petrol.* **123,** 345–357 (1996).
7. Bolfan-Casanova, N., Keppler, H. & Rubie, D. C. *Earth Planet. Sci. Lett.* **182,** 209–221 (2000).
8. Huang, X. G., Xu, Y. S. & Karato, S. I. *Nature* **434,** 746–749 (2005).
9. Yoshino, T., Manthilake, G., Matsuzaki, T. & Katsura, T. *Nature* **451,** 326–329 (2008).
10. Frost, D. J. & Dolejs, D. *Earth Planet. Sci. Lett.* **256,** 182–195 (2007).
11. Kelley, S. P. & Wartho, J. A. *Science* **289,** 609–611 (2000).

NEURODEGENERATIVE DISEASES

G-quadruplex poses quadruple threat

Multiplication of repetitive DNA sequences is often the cause of neurodegenerative diseases. A four-stranded structure has been found to form in one such expansion in the gene *C9orf72*, altering gene function in four ways. SEE ARTICLE P.195

J. PAUL TAYLOR

The most common cause of familial and sporadic forms of the neurodegenerative diseases amyotrophic lateral sclerosis and frontotemporal dementia is abnormal expansion of a repeated six-nucleotide DNA sequence[1,2]. The repeated sequence is located in a non-protein-coding region of the gene *C9orf72*, in which two nucleotides, guanine (G) and cytosine (C), recur in GGGGCC repeating units. But how expansion of the repeat leads to disease is unknown. On page 195 of this issue, Haeusler *et al.*[3] show that DNA and RNA molecules comprised of GGGGCC repeats adopt a peculiar secondary structure that could account for several of the pathological features found in *C9orf72*-related diseases.

In addition to the double helix described by James Watson and Francis Crick (in which two DNA strands align, G pairing with C and adenine pairing with thymine), a nucleic-acid strand that is rich in G can fold in on itself to form a four-stranded secondary structure

called a G-quadruplex[4]. To adopt this topology, four G bases associate through atypical pairing, forming a square planar structure called a G-quartet, and two or more G-quartets stack on top of one another to form a G-quadruplex (Fig. 1). Haeusler and colleagues used several approaches to characterize the secondary structures formed by the expanded GGGGCC repeats in *C9orf72*–related amyotrophic lateral sclerosis (ALS) and frontotemporal dementia. In addition to G-quadruplex formation by GGGGCC-repeating RNA, which has already been shown for these diseases[5,6], they found stable G-quadruplex formation in DNA with this repeat. More importantly, they found that enhanced G-quadruplex formation in RNA and DNA has consequences for normal expression of *C9orf72*.

Sequences that can form G-quadruplexes have been conserved throughout evolution, and are enriched in functional locations such as transcriptional start sites, at which RNA polymerase, the enzyme responsible for gene transcription, attaches and begins to transcribe DNA. This positioning suggests a functional role for these structures *in vivo*. A flurry of data indicates that G-quadruplex assembly influences diverse cellular processes, including transcription, translation and RNA localization[4]. In the context of the *C9orf72* GGGGCC expansion, Haeusler and co-workers found that G-quadruplex assembly in DNA causes increased pauses in transcription in the expanded repeat region, and that this pausing could prevent normal elongation of transcripts to full length.

Furthermore, Haeusler *et al.* determined that G-quadruplexes in *C9orf72* DNA promote the formation of stable R-loops — triple-stranded nucleic-acid structures that assemble when a newly formed RNA transcript exiting the RNA polymerase invades the DNA double helix and binds to one DNA strand, displacing the other in the process. R-loops often form as a normal by-product of transcription, but protective mechanisms are in place to remove them. If R-loops are not resolved, however, they can halt transcriptional elongation, and pose a threat to efficient gene expression[7].

Haeusler and co-workers have therefore identified two distinct mechanisms, increased pausing and R-loop formation, whereby G-quadruplex assembly in *C9orf72* could cause abortive transcription, producing short, repeat-containing transcripts but reducing overall production of the C9orf72 protein. Consistent with these findings, the authors observed that abortive GGGGCC-containing RNAs accumulated in the spinal cord and the motor cortex region of the brain of patients with *C9orf72*-expanded ALS. These results potentially resolve the paradox of how GGGGCC-containing RNA foci — a hallmark of neurodegeneration, consisting of aggregates of RNA that form secondary structures and bind cellular proteins — could accumulate in patients with GGGGCC-repeat expansion, despite a reported reduction in overall levels of *C9orf72* messenger RNA[1,8,9]. The findings might also partly explain why, in studies[10,11] that relied on various methods to monitor different regions of the transcript, reduced *C9orf72* mRNA levels were inconsistently found in cell lines derived from patients with *C9orf72*-related disease.

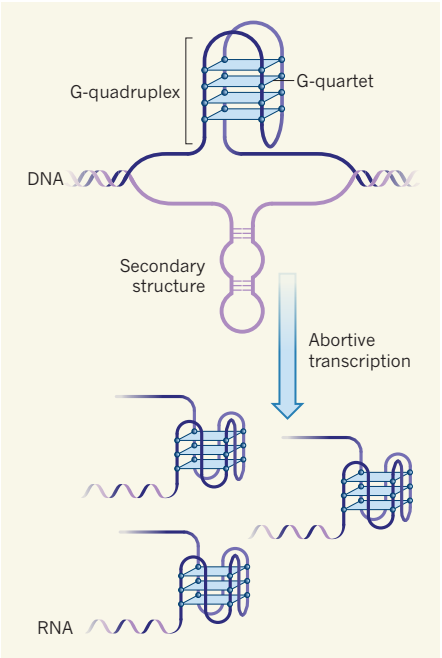


Figure 1 | G-quadruplex formation in nucleic-acid strands. G-quaduplexes form in guanine (G)-rich regions of DNA or RNA. G bases bind through atypical pairing to form square planar structures called G-quartets, which can stack to form a G-quadruplex. In DNA, secondary structures also form in the complementary strand. Haeusler *et al.*[3] report G-quadruplex formation in DNA when a repeated six-nucleotide sequence expands in the gene *C9orf72*, as observed in patients with certain neurodegenerative diseases. This causes abortive transcription, resulting in accumulation of short, repeat-containing RNA transcripts that also form G-quadruplexes.

There has been tremendous interest in identifying the RNA-binding proteins that associate with *C9orf72* transcripts containing expanded GGGGCC repeats. Interaction between the RNA G-quadruplex and key RNA-binding proteins might sequester the proteins, preventing them from functioning normally. This could contribute to pathogenesis, analogous to the functional depletion of splicing factors (RNA-binding proteins that excise non-protein-coding segments of immature RNA following transcription) in the multisystemic disease myotonic dystrophy[12]. So far, a few dozen proteins that bind GGGGCC repeats *in vitro* have been identified, a subset of which co-localize in RNA foci in cells from patients with *C9orf72*-related diseases[10,11,13–15]. Haeusler and colleagues have taken a leap forward not only by performing a comprehensive, quantitative assessment of proteins that bind *C9orf72* transcripts, but also by determining the distinct RNA topologies favoured by these proteins. This enormously valuable catalogue will propel efforts to determine the role of individual RNA-binding proteins in disease.

A highlight of Haeusler and co-workers' analyses is the finding that the protein nucleolin interacts with RNA G-quadruplexes formed by *C9orf72*, consistent with a previous report that nucleolin can bind G-quadruplexes formed in other genes[16]. Nucleolin is a multifunctional RNA-binding protein found in the nucleolus, a structure in the cell nucleus. Among other roles, nucleolin is required for the nucleolus' main function: assembling the ribosome[16], the cellular machinery responsible for translation. The authors observed that nucleolin is mislocalized to RNA foci in neurons of the motor cortex of patients with *C9orf72*-related disease. In concert with this abnormal location of nucleolin, the authors found signs that the nucleolus was unable to produce mature ribosomes normally, culminating in the build-up of untranslated mRNA in the cytoplasm — one possible reason for pathology.

C9orf72-related neurodegeneration could result from two modes of toxicity: loss of function of *C9orf72* or gain of function mediated by its product (or both). It will be crucial to elucidate the relative contribution of each of these two modes, because this will guide strategies for therapeutic intervention. In particular, it will be imperative to determine the role of reduced *C9orf72* expression in disease, as well as the consequences of further decreases in levels of the C9orf72 protein (efforts are under way to silence this gene, for which no function is known) to prevent the accumulation of RNA foci[10,11,17]. Relevant to this, two groups recently reported that reduction or loss of *C9orf72* in zebrafish resulted in the degeneration of motor neurons[18,19]. Although depletion of *C9orf72* in the central nervous system of mice[17] or in cultured human motor neurons[10,11,17] has not been associated with toxicity, it will be vital to assess the consequences of complete *C9orf72* loss in mammals.

G-quadruplex assembly by *C9orf72* might also cause toxic gain of function, not only by generating abortive GGGGCC-containing transcripts, but also because of the specific properties of G-quadruplex RNA. For example, G-quadruplexes can form either within a single RNA molecule or between two different RNA molecules, and this latter conformation might drive assembly of disease-related RNA foci. Moreover, the secondary structures of the RNAs that accrue in these foci could potentially influence the spectrum of sequestered proteins, shaping the resultant pathology. Finally, G-quadruplex formation could, through an unknown mechanism, promote the production of toxic dipeptides from GGGGCC-containing RNA transcripts[6]. Thus, G-quadruplexes of *C9orf72* in both DNA and

RNA might simultaneously promote loss of normal function and gain of toxic function, posing a quadruple threat. ■

J. Paul Taylor *is at St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.*
e-mail: jpaul.taylor@stjude.org

1. DeJesus-Hernandez, M. *et al. Neuron* **72,** 245–256 (2011).
2. Renton, A. E. *et al. Neuron* **72,** 257–268 (2011).
3. Haeusler, A. R. *et al. Nature* **507,** 195–200 (2014).
4. Bochman, M. L., Paeschke, K. & Zakian, V. A. *Nature Rev. Genet.* **13,** 770–780 (2012).
5. Fratta, P. *et al. Sci. Rep.* **2,** 1016 (2012).
6. Reddy, K., Zamiri, B., Stanley, S. Y. R., Macgregor, R. B. Jr & Pearson, C. E. *J. Biol. Chem.* **288,** 9860–9866 (2013).
7. Aguilera, A. & García-Muse, T. *Mol. Cell* **46,** 115–124 (2012).
8. Gijselinck, I. *et al. Lancet Neurol.* **11,** 54–65 (2012).
9. Xi, R. *et al. Am. J. Hum. Genet.* **92,** 981–989 (2013).
10. Sareen, D. *et al. Sci. Transl. Med.* **5,** 208ra149 (2013).
11. Donnelly, C. J. *et al. Neuron* **80,** 415–428 (2013).
12. Echeverria, G. V. & Cooper, T. A. *Brain Res.* **1462,** 100–111 (2012).
13. Mori, K. *et al. Acta Neuropathol. (Berl.)* **125,** 413–423 (2013).
14. Lee, Y.-B. *et al. Cell Rep.* **5,** 1178–1186 (2013).
15. Xu, Z. *et al. Proc. Natl Acad. Sci. USA* **110,** 7778–7783 (2013).
16. Abdelmohsen, K. & Gorospe, M. *RNA Biol.* **9,** 799–808 (2012).
17. Lagier-Tourenne, C. *et al. Proc. Natl Acad. Sci. USA* **110,** E4530–E4539 (2013).
18. Therrien, M., Rouleau, G. A., Dion, P. A. & Parker, A. *PLoS ONE* **8,** e83450 (2013).
19. Ciura, S. *et al. Ann. Neurol.* **74,** 180–187 (2013).

This article was published online on 5 March 2014.

NEUROSCIENCE

Ordered randomness in fly love songs

A systematic and painstaking analysis reveals that much of the complexity and variability of the courtship song of male fruit flies can be accounted for by simple rules that relate sensory experience to motor output. SEE LETTER P.233

BENCE P. ÖLVECZKY

Well-crafted love songs can be the ticket to fun times and reproductive success, whether you are a member of the Beatles or one of the many animals that woo their mates by singing. Although some troubadours serve up monotonous repetitions of stereotyped songs, most animals, including birds, mammals and insects, like to jazz things up by varying their song patterns. But how the brain generates such variability, and improvisation more generally, remains largely a mystery. On page 233, Coen *et al.*[1] shed light on this issue by showing that much of the variability in the love songs of fruit flies can be predicted from the singers' movements.

Neuroscientists' fascination with the sex life of the fruit fly *Drosophila melanogaster* began more than 35 years ago with the discovery of *fruitless,* a gene essential for the male courtship ritual[2]. This unique handle on a complex social behaviour, in an organism amenable to genetic modification, paved the way for an exceedingly detailed anatomical mapping of the underlying neural circuitry[3,4]. Deciphering the details of what these circuits do and determining what they can teach us about brain function more broadly are major challenges that would be greatly helped by having a comprehensive description of the computations that the circuits perform and the behaviours they implement.

One thing that we know these circuits do is transform their male owners into mini Casanovas. On encountering a receptive virgin female, a male fly will gently tap her rear end, serenade her with a 'song' by vibrating one of his wings, and lick her genitalia[5]. Although these behaviours are part of any self-respecting fly's lovemaking repertoire, the duration and ordering of the different courtship elements can be highly unpredictable.

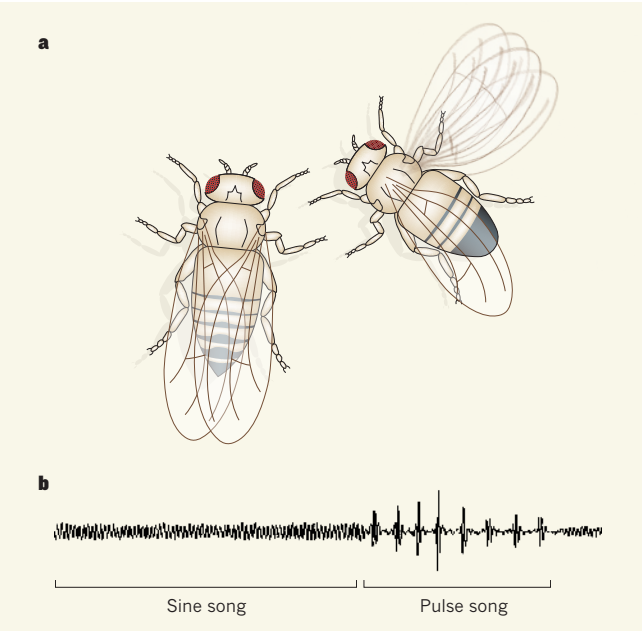


Figure 1 | The male fruit fly's serenade. **a**, Male fruit flies attract females by vibrating one of their wings. **b**, The fly has two distinct song types — the humming sine song and the purring pulse song — and switches between them to generate variable song sequences. Coen *et al.*[1] found that these switches can be predicted by the fly's movements. (Data depicted in **b** taken from Fig. 6 of ref. 11.)

What gives rise to such seemingly random behaviour? Is the variability due to stochastic fluctuations in the underlying neural networks (neural noise)[6,7], or the result of a dynamic sensory experience?

To address these questions, Coen and colleagues focused on the male fly's song, itself a variable sequence of distinct elements[1,8]. Just as the Beatles made a career of mixing 'love', 'you', 'me', 'she' and 'baby' in different ways, so male fruit flies switch between 'sine' and 'pulse' songs to impress their audience (Fig. 1). By eavesdropping on more than 100,000 love songs while carefully monitoring the whereabouts of the courting couple, the authors suggest that a logic and order exist in the apparent musical randomness.

Coen *et al.* performed a statistical analysis of their high-resolution behavioural data, and found that transitions between sine and pulse songs can be predicted from the courted female's movements. The authors further discovered that the male's visual experience of the female shapes his song through neural circuits that control locomotion. In fact, the best predictor of song structure is not the female's movements, but the singer's own. Even blind flies, who are induced to sing by the scent of virgin females, show a bias in their song transitions that can be predicted from their movements. The picture that emerges from all of this is one in which the male fly executes a tightly integrated song-and-dance number, inspired by (if he can see her) his partner's movements.

As impressive as that may be, the extent to which the female cares about the details of her lover's intricate performance remains unclear. Does she use information embedded in his song pattern to determine his desirability? Does his ability to couple changes in his song to body movements — his or hers — correlate with other qualities that she would want in a mate? In other words, is song patterning an example of a carefully tuned signalling system, or does it reflect a coupling between leg and wing movements that evolved for unrelated reasons?

50 Years Ago

Recent investigations have shown that the fluoride content of Greek teeth from the cities of Athens and Salonika was considerably high. This may explain, at least in part, the low prevalence of dental caries observed in Greece … With the exception of sea salt, however, the fluoride content of other foods commonly produced and consumed in Greece is not known … The analyses showed that the fluoride content of olive oil from the Island of Crete was 0.36 p.p.m. and that from the area of Kalamai 0.63 p.p.m … it appears that the inclusion of olive oil in the daily Greek diet does not make any significant contribution to the amount of ingested fluoride. Thus, at present, sea salt remains an important source of dietary fluoride in Greece for protection against dental caries. This may well be the case in other countries, such as Taiwan, Ceylon and Lebanon, where because of local food customs the amount of sea salt consumed has been estimated to be considerable: about 16–20 g per person per day.
From *Nature* 14 March 1964

100 Years Ago

Think of the Niagaras of speech pouring silently through the New York telephone exchanges where they are sorted out, given a new direction, and delivered audibly perhaps a thousand miles away. New York has 450,000 instruments — twice the number of those in London. Los Angeles has a telephone to every four inhabitants … Our whole social structure has been reorganised. We have been brought together in a single parlour for conversation and to conduct affairs, because the American Telephone and Telegraph company spends annually for research … a sum greater than the total income of many universities.
From *Nature* 12 March 1914

Initial experiments to address these questions have failed to provide clear answers. Coen *et al.* show that song transitions are similar whether or not the singer is ultimately successful in mating. Yet pheromone-insensitive males, who sing for normal durations but have altered song patterning[8], tend to be slower and less successful in convincing females to mate[1,8]. Whether these flies are handicapped in the courting game because of a defect in how they vary their songs, or because of unrelated effects, remains to be seen. But whether song patterning matters to females or not, we now know that its variability, and probably the variability of many other 'fixed' behaviours, is not simply the consequence of noise in nervous-system function[6,7]. Rather, a sizeable fraction of that variability is likely to reflect computations performed by reliable and predictable brains on an ever-changing sensory environment.

Importantly, this insight was made possible by simultaneously observing, at high temporal resolution, the sensory environment and behavioural output of a genetically tractable organism during a complex social interaction. Such detailed analysis applied to natural behaviours has the power, as Coen *et al.* aptly demonstrate, to distil seemingly complex and unpredictable behavioural patterns into simple rules and sensorimotor transformations[9,10]. With such an approach, rather than being the fog that prevents us from understanding nervous-system function, behavioural variability and complexity can be the searchlight that helps us to identify the computational problems that brains evolved to solve. ■

Bence P. Ölveczky *is in the Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.*
e-mail: olveczky@fas.harvard.edu

1. Coen, P. *et al. Nature* **507,** 233–237 (2014).
2. Hall, J. C. *Behav. Genet.* **8,** 125–141 (1978).
3. Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirián, L. & Dickson, B. J. *Cell* **121,** 795–807 (2005).
4. Yu, J. Y., Kanai, M. I., Demir, E., Jefferis, G. S. X. E. & Dickson, B. J. *Curr. Biol.* **20,** 1602–1614 (2010).
5. Spieth, H. T. *Annu. Rev. Entomol.* **19,** 385–405 (1974).
6. Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. *Nature Rev. Neurosci.* **9,** 292–303 (2008).
7. Destexhe, A. & Rudolph-Lilith, M. *Neuronal Noise* (Springer, 2012).
8. Trott, A. R., Donelson, N. C., Griffith, L. C. & Ejima, A. *PLoS ONE* **7,** e46025 (2012).
9. Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C. & Couzin, I. D. *Proc. Natl Acad. Sci. USA* **108,** 18720–18725 (2011).
10. Censi, A., Straw, A. D., Sayaman, R. W., Murray, R. M. & Dickinson, M. H. *PLoS Comput. Biol.* **9,** e1002891 (2013).
11. Rideout, E. J., Dornan, A. J., Neville, M. C., Eadie, S. & Goodwin, S. F. *Nature Neurosci.* **13,** 458–466 (2010).

This article was published online on 5 March 2014.

EVOLUTIONARY BIOLOGY

Speciation undone

Hybridization can cause two species to fuse into a single population. New observations suggest that two species of Darwin's finches are hybridizing on a Galapagos island, and that a third one has disappeared through interbreeding.

PETER R. GRANT & B. ROSEMARY GRANT

The process of speciation, in which one species splits into two, is vulnerable to collapse in its early stages through interbreeding and the exchange of genes, a process referred to as introgression. As explained by the evolutionary biologist Theodosius Dobzhansky[1], "Introgressive hybridization may, then, be a passing stage in the process of species formation. On the other hand, the adaptive value of hybrids may be as high as that of their parent; introgressive hybridization may lead to obliteration of the differences between the incipient species and their fusion into a single variable one, thus undoing the result of the previous divergent development." Writing in *American Naturalist*, Kleindorfer *et al.*[2] offer a possible example of this process, in a study suggesting that one population of Darwin's finches has become extinct through interbreeding with another.

Until Kleindorfer and colleagues' report, three species of tree finch were known to occur together in the highlands of Floreana Island in the Galapagos (Fig. 1). They differ in body size and in the size and shape of the beak, but, unlike many birds elsewhere, not in plumage. The medium tree finch (*Camarhynchus pauper*) is present only on Floreana, whereas the small tree finch (*Camarhynchus parvulus*) and large tree finch (*Camarhynchus psittacula*) also occur together on several other islands. The pattern of distribution and size differences led evolutionary biologist David Lack to suggest[3] that speciation had occurred on Floreana through the invasion of large tree finches from Isabela Island, followed by evolutionary reduction in average size. The resulting medium tree finches did not interbreed with the large tree finches that arrived later, apparently from Santa Cruz Island.

Kleindorfer and colleagues now report that this pattern no longer exists: the large tree finch

A–C: P. R. GRANT & B. R. GRANT

has disappeared from Floreana! By comparing the morphological features of present-day Floreana finches (studied in 2005 and 2010) with historical data, and conducting a genetic study of current populations using DNA-sequence markers (microsatellites), the authors show that there are currently only two distinct populations on the island, corresponding to the small and medium tree finches. The analyses also revealed that individuals that do not fit into either population show intermediate characteristics, suggesting that they are hybrids. Consistent with the hypothesis of ongoing hybridization on the island, the authors observed females of the morphologically larger group (the medium tree finch) pairing with males of the smaller group, and they identified 15% of yearling males in 2010 as hybrids.

The authors suggest that hybridization may have been responsible for the disappearance of the large tree finch from Floreana, and that it may now be causing the remaining two species to fuse into one: speciation in reverse[4]. What has brought this about? The most likely answer is anthropogenic change to the habitat. A human settlement was established on the island just before Darwin's visit in 1835. The natural vegetation subsequently became rapidly degraded, and by the end of the nineteenth century two species of finch and a species of mockingbird had become extinct[5]. The large tree finch was rare: only 4 male and 13 female specimens were collected for museums between 1852 and 1906. The birds may have experienced difficulty in finding mates of their own species, hybridized with medium ground finches and become absorbed into the population[2,5].

Alternatively, the large tree finch may have become extinct through changes in the food supply alone, without any interbreeding. One way of distinguishing between the two hypotheses might be to use molecular markers to search for evidence of past introgression. If markers could be identified in the genomes of large tree finches on Santa Cruz (and the museum specimens from Floreana) but not on Isabela, and also found in the medium tree finches, they could be the smoking gun of introgression.

To identify hybrids between the remaining medium and small tree finches, Kleindorfer *et al.* relied on a clustering technique with (acknowledged) low statistical power. But two other identifying clues are at hand. The first is beak shape, which is known to be a marker of species identity in these finches[3,5]. The second is song. Different species of Darwin's finches sing different songs, which are acquired through learning by nestlings and fledglings and used by adults to identify mates[6,7]. A song of a large tree finch was tape-recorded on Floreana in January 1962[6], but we failed to see or hear any of these birds on five visits during 1979–2004[8]. The song of a large tree finch coming from the mouth of a medium tree finch could be a wail from the ghost of an interbreeding past.

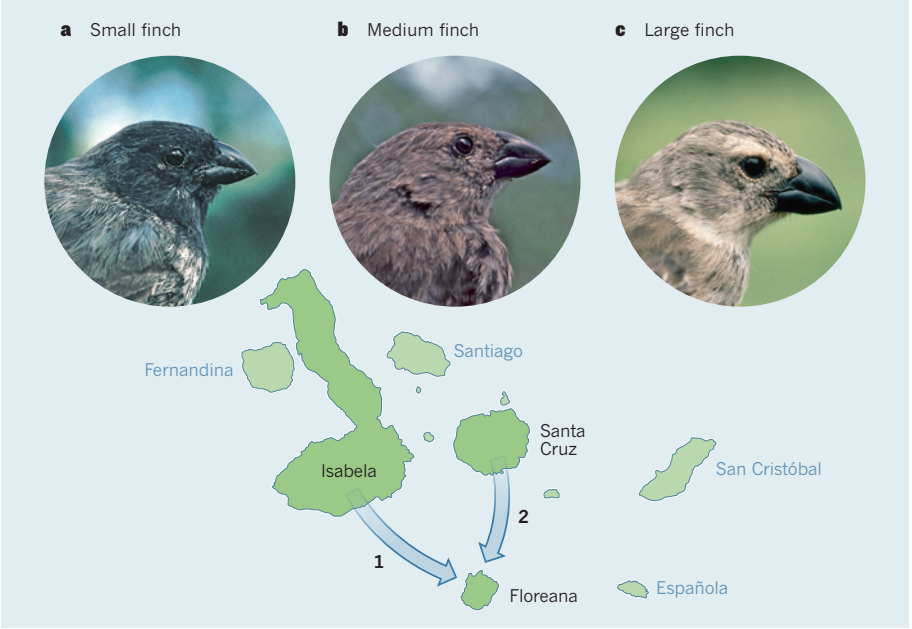


Figure 1 | Invasion, evolution and loss. Three species of Darwin's finches, the small tree finch *Camarhynchus parvulus* (**a**), the medium tree finch *Camarhynchus pauper* (**b**) and the large tree finch *Camarhynchus psittacula* (**c**), have been known to inhabit the Galapagos island of Floreana. The medium finch occurs nowhere else in the archipelago, and its morphological distinctiveness was interpreted by evolutionary biologist David Lack to be the result of invasion of a small form of *C. psittacula* from Isabela (**1**), followed by an evolutionary reduction in size and change in beak shape[3,5]. Later, a larger form of *C. psittacula* invaded from Santa Cruz (**2**), and remained unchanged. However, Kleindorfer *et al.*[2] now report that this large species is no longer found on Floreana.

Although there is some uncertainty about hybrid identification in this study, the disappearance of a species through hybridization is certainly plausible. On the small uninhabited island of Daphne Major, two species of ground finch (genus *Geospiza*) have been converging morphologically and genetically for more than 30 years as a result of persistent (although rare) introgressive hybridization following a natural change in the food supply. If introgression continues at the same rate, the two species will fuse into one in approximately 40 years[7]. The finches of Daphne Major are also a reminder that, under special circumstances, hybridization can lead to the opposite outcome — the formation of a new species. A new genetic lineage has become established on Daphne by an immigrant hybrid from Santa Cruz, and is now behaving as a separate species (see ref. 7 for further details).

Kleindorfer and colleagues' findings suggest that the small and medium tree finches on Floreana may also be fusing into one species. The authors raise the intriguing possibility that hybrids between these populations have an immunological advantage over the parental species in the face of attack by a parasitic fly, *Philornis downsi*, whose larvae eat and kill finch nestlings. The fly was introduced to the archipelago 50 years ago[9]. Their study is important because it adds weight to a growing concern that we humans are causing loss of biodiversity by altering habitats, in some cases by bringing separate species into proximity and causing their extinction through interbreeding[10,11]. Rapid radiations of fishes[11,12] and finches[5] are especially at risk because their morphological evolution is not accompanied by strong barriers to gene exchange. Uniquely valuable in showing how speciation is done[7,13], such species deserve special protection from being artificially undone. ■

Peter R. Grant *and* **B. Rosemary Grant**
are in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08540, USA.
e-mails: prgrant@princeton.edu; rgrant@princeton.edu

1. Dobzhansky, T. *Genetics and the Origin of Species* 350 (Columbia Univ. Press, 1941).
2. Kleindorfer, S. *et al. Am. Nat.* **183,** 325–341 (2014).
3. Lack, D. *Darwin's Finches* (Cambridge Univ. Press, 1947).
4. McKinnon, J. S. & Taylor, E. B. *Nature* **482,** 313–314 (2012).
5. Grant, P. R. *Ecology and Evolution of Darwin's Finches* (Princeton Univ. Press, 1999).
6. Bowman, R. I. in *Patterns of Evolution in Galapagos Organisms* (eds Bowman, R. I., Berson, M. & Leviton, A. E.) 237–537 (Am. Assoc. Adv. Sci., Pacif. Div., 1983).
7. Grant, P. R. & Grant, B. R. *40 Years of Evolution* (Princeton Univ. Press, in the press).
8. Grant, P. R., Grant, B. R., Petren, K. & Keller, L. F. *Biol. Conserv.* **122,** 499–503 (2005).
9. Causton, C. E. *et al. Ann. Entomol. Soc. Am.* **99,** 121–143 (2006).
10. Seehausen, O., Takimoto, G., Roy, D. & Jokela, J. *Mol. Ecol.* **17,** 30–44 (2008).
11. Vonlanthen, P. D. *et al. Nature* **482,** 357–362 (2012).
12. Seehausen, O. *Proc. R. Soc. B* **273,** 1987–1998 (2006).
13. Grant, P. R. & Grant, B. R. *How and Why Species Multiply* (Princeton Univ. Press, 2008).

REVIEW

doi:10.1038/nature13118

The present and future role of microfluidics in biomedical research

Eric K. Sackmann[1], Anna L. Fulton[2] & David J. Beebe[3]

Microfluidics, a technology characterized by the engineered manipulation of fluids at the submillimetre scale, has shown considerable promise for improving diagnostics and biology research. Certain properties of microfluidic technologies, such as rapid sample processing and the precise control of fluids in an assay, have made them attractive candidates to replace traditional experimental approaches. Here we analyse the progress made by lab-on-a-chip microtechnologies in recent years, and discuss the clinical and research areas in which they have made the greatest impact. We also suggest directions that biologists, engineers and clinicians can take to help this technology live up to its potential.

More than a decade ago, we wrote that "microfluidics has the potential to significantly change the way modern biology is performed"[1]. Indeed, we were part of a chorus of researchers that recognised the possibility of new microfluidic tools making substantial contributions to biology and medical research[2–5]. The optimism surrounding microfluidics was well warranted, given the compelling advantages that microfluidic approaches could possibly have over traditional assays used in cell biology. Conceptually, the idea of microfluidics is that fluids can be precisely manipulated using a microscale device built with technologies first developed by the semiconductor industry and later expanded by the micro-electromechanical systems (MEMS) field. These devices, commonly referred to as miniaturized total analysis systems (μTASs)[6,7] or lab-on-a-chip (LoC) technologies, could be applied to biology research to streamline complex assay protocols; to reduce the sample volume substantially; to reduce the cost of reagents and maximize information gleaned from precious samples; to provide gains in scalability for screening applications and batch sample processing analogous to multi-well plates; and to provide the investigator with substantially more control and predictability of the spatio-temporal dynamics of the cell microenvironment.

The field of microfluidics is characterized by the study and manipulation of fluids at the submillimetre length scale. The fluid phenomena that dominate liquids at this length scale are measurably different from those that dominate at the macroscale (Box 1). For example, the relative effect of the force produced by gravity at microscale dimensions is greatly reduced compared to its dominance at the macroscale. Conversely, surface tension and capillary forces are more dominant at the microscale; these forces can be used for a variety of tasks, such as passively pumping fluids in microchannels[8]; precisely patterning surfaces with user-defined substrates[9]; filtering various analytes[10]; and forming monodisperse droplets[11] in multiphase fluid streams for a variety of applications. These examples represent only a fraction of the myriad problems that microfluidic technologies have attempted to address.

The development of comprehensive microfluidic solutions to address problems in biology and clinical research has been embraced by engineers. However, despite material advances in microfluidics as a technology platform, the adoption of novel μTAS techniques in mainstream biology research has not matched the initial enthusiasm surrounding the field[12]. Some argue the technology is still in search of a 'killer application', where the sample-to-answer concept provides a solution that greatly outperforms current methods[13,14]. In this perspective, we will examine the impact of microfluidic technologies on cell biology and medical research within the past decade. We discuss some of the barriers to adoption of microfluidic technologies in mainstream biomedical research, and use a case study to illustrate and highlight these challenges. We focus our attention on recent developments in the field that are facilitating the application of microfluidic technologies to solving problems in diagnostics and biology research. In this area, we highlight the innovative use of different materials that are more optimally suited to performing a given task; and we examine how researchers are taking advantage of μTAS methods to enable scientific inquiry in ways that were not possible using traditional methods. Finally, we will discuss positive trends in the field and infer lessons that can be applied to future microfluidic technology development.

The impact of microfluidics on biomedical research

A primary goal for much of the microfluidics community is to develop technologies that enhance the capabilities of investigators in biology and medical research. Many microfluidic studies describe methods that aim to replace traditional macroscale assays, and usually perform proof-of-concept (PoC) experiments that attempt to demonstrate the efficacy of the new approach. These novel microfluidic methods are usually published in journals that might be characterized as 'engineering' journals, or publications whose readership comprises largely engineers and other members of the physical sciences (for example, chemists and physicists). If publishing PoC studies in engineering journals represents the development phase for a novel biology assay, then the implementation of the technique can be characterized as when the technology is used and published in a biology or medical journal. After all, the stated goal of virtually all PoC studies is to demonstrate new technologies that enable biologists in their everyday research.

We measured the use of microfluidic technologies in mainstream biomedical research over the past decade to assess their impact beyond the engineering community (Fig. 1). In order to identify broad trends of what journals have published papers that use microfluidics (search terms "microfluidic*" and "nanofluidic*"; see Fig. 1 legend), we defined three categories: (1) 'engineering' journals (for example, *Lab on a Chip*, *Small*, *Analytical Chemistry*); (2) 'biology and medicine' journals (for example, *Blood*, *Cell*, *Journal of Clinical Investigation*); and (3) 'multidisciplinary' journals (for example, *Nature*, *Science*, *Proceedings of the National Academy*

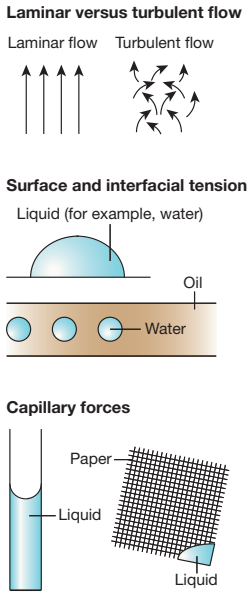
[1]Materials Science Program, Department of Biomedical Engineering, Wisconsin Institutes for Medical Research, University of Wisconsin-Madison, 1111 Highland Avenue, Madison, Wisconsin 53705-2275, USA. [2]Wendt Commons Library, University of Wisconsin-Madison, 215 North Randall Avenue, Madison, Wisconsin 53706, USA. [3]Department of Biomedical Engineering, Wisconsin Institutes for Medical Research, University of Wisconsin-Madison, 1111 Highland Avenue, Room 6009, Madison, Wisconsin 53705-2275, USA.

BOX 1
Useful microfluidics concepts

Laminar versus turbulent flow. The Reynolds number (Re) is a dimensionless quantity that describes the ratio of inertial to viscous forces in a fluid. Re is proportional to the characteristic velocity of the fluid and the length scale of the system; it is inversely proportional to the fluid viscosity. High-Re (~2,000) fluids have flow profiles that increasingly mix stochastically (turbulent flow; Box 1 Figure below). For microfluidic systems, Re is almost always in the laminar flow regime, allowing for highly predictable fluid dynamics. Molecular transport also changes dramatically at this scale because convective mixing does not occur, enabling predictable diffusion kinetics.

Surface and interfacial tension. Surface tension describes the tendency of a fluid in a surface to reduce its free energy by contracting at the surface–air interface. Interfacial tension is a similar phenomenon, but is generally applied to two immiscible fluids (for example, oil and water). These forces play more dominant roles on the microscale (Box 1 Figure below) compared to gravity, which is much more dominant on the macroscale. Researchers have used these phenomena to conduct protein and cell sorting, perform nanoreactions for protein crystallization, and passively drive fluids through microchannels.

Capillary forces. Capillary action describes the movement of a fluid through a narrow constriction, such as a narrow tube or porous material (Box 1 Figure below). At the microscale, capillary action is a more dominant force, allowing fluids to advance in opposition to gravity. Capillary forces have been used to manipulate fluids in many applications, the most famous examples perhaps being the at-home pregnancy test and portable glucometers to monitor blood glucose levels.



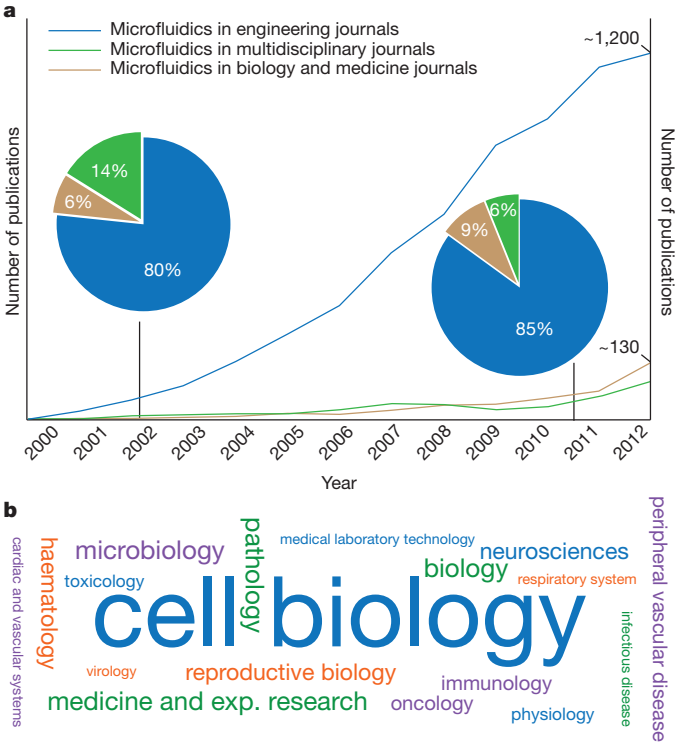


Figure 1 | Microfluidic publications in engineering, multidisciplinary, and biology and medicine journals from 2000 to 2012. **a**, In 2012, there were roughly 10 times more microfluidic publications in engineering journals compared to biology and medicine (biomedical) journals (left-hand pie chart inset). However, the share of microfluidics papers being published in multidisciplinary journals decreased as publication share in biomedical and engineering journals increased (right-hand pie chart). **b**, Word cloud illustrating what fields most frequently used microfluidics. The size of the font is proportional to the cumulative number of publications in the Web of Science (WoS) category (2000–12), with the exception of 'cell biology', which would need to be ~5 times larger. Methodology of the searches was as follows. A literature search was performed using WoS (provided by Thomson Reuters) to determine the number of microfluidics publications in various disciplines. The search was performed for the terms "microfluidic*" and "nanofluidic*". The number of publications were obtained from the WoS analytics reporting system for each search term, and then summed before being presented above. Three categories were characterized by the WoS search that capture the relevant journals for the years 2000–12. The analysis shown here as "Microfluidics in engineering journals" reports the number of microfluidic publications in the 'Nanoscience and nanotechnology' WoS category. The analysis shown here as 'Microfluidics in multidisciplinary journals' corresponds to the 'Multidisciplinary' WoS category. The analysis shown here as 'Microfluidics in biology and medicine journals' reports publications from WoS categories shown in Fig. 1b. The search explicitly excluded 'reviews', 'book chapters', 'book reviews', 'meeting abstracts', 'meeting summaries', and included 'articles'. The data shown reflects the most recent literature search, performed on 21 March 2013. The following search general string was used: Topic = (microfluidic*) AND Year Published = (2000-2012) AND Document Types = (Article) NOT Document Types = (Book OR Book Chapter OR Book Review OR Meeting Abstract OR Meeting Summary OR Proceedings Paper OR Review). This string yielded the total 'microfluidic*' publications in all WoS categories ('*' allows for permutations of the keyword). The search was then refined by the WoS categories shown above (for example, Web of Science Categories = (MULTIDISCIPLINARY). Importantly, the nominal results of this search would probably vary if other search tools such as SCOPUS, Google Scholar and PubMed were used[98,99]. For example, the gross number of publications would probably increase if SCOPUS were used for the search, as this tool indexes a higher number of journals than WoS[99].

of Sciences). The results reveal, unsurprisingly, that the overwhelming number of microfluidics papers are still being published in engineering journals (Fig. 1a). These engineering journals have facilitated the technological development and growth of microfluidics over the past decade. It is important to note that some of these 'engineering' studies may have been designed for non-biomedical purposes, but this does illustrate where the majority of microfluidic activity and exposure has occurred. Today the majority of microfluidics publications still appear in engineering journals (85%) as the microfluidics community has grown substantially, and 'biology and medicine' journals have taken some publication share from interdisciplinary journals (9% and 6%, respectively).

Last, we analysed what fields within the biomedical research community are using microfluidic technologies the most (Fig. 1b). 'Cell biology' and 'Biology' encompass most of the microfluidics publications, possibly because these categories are somewhat generic and incorporate several subcategories. Following these, the most use of microfluidics is seen in 'Haematology', 'Medicine and experimental research' and 'Immunology'. Most of these publications are for diagnostic applications (in the case of Medicine and experimental research) and the manipulation of blood samples for biology research (Haematology and Immunology)—applications where microfluidics has compelling advantages over traditional methods. However, despite these few examples, the evidence suggests that a 'killer application' that propels microfluidics into the mainstream has yet to emerge.

A case study in chemotaxis assays

The state of the art for most conventional assays used in cell biology research is evolving and improving over time. Biologists understand better than anyone the deficiencies of the techniques they use, and individual groups occasionally make modifications to traditional assays that are adopted more broadly by other biology researchers. An example of this technological evolution can be observed in visual chemotaxis assays—techniques that measure the directional migration of a cell in response to a source of chemotactic factors that change concentration in space and time.

Chemotaxis assays have improved substantially since their initial introduction in the 1960s (Fig. 2). The most widely used chemotaxis assay is known as the 'Boyden chamber' or 'Transwell' assay, developed in 1962 by Boyden[15]. The Transwell assay works by creating a concentration gradient of chemoattractant compounds between two wells that are separated by a microporous membrane. Chemotactic cells located in the upper well sense the gradient in concentration and migrate across the membrane towards the solution in the lower well where the cells are counted. Its simplicity and ease of use (no special instrumentation is required) has contributed to its widespread use over the past 50 years. Investigators have used the method to identify chemotactic factors for various cell types, despite the fact that the technique disallows observation of the cell migration path or cell morphology. This experimental limitation (along with others) led to the development of visual chemotaxis assays such as the Zigmond chamber[16]. In this system, cells can be observed as they undergo chemotaxis on a coverslip across a narrow constriction (tens of micrometres) towards a source chemoattractant. It is worth noting that the Zigmond chamber is a microfluidic device developed by biologists at least a decade before the emergence of the microfluidic/μTAS field as we know it. Importantly, this technique allows for clear imaging of cell migration and morphology. Modifications to this design, called the Dunn[17] and Insall[18] chambers, were subsequently developed, and these advances substantially improved the high-resolution, long-term imaging capabilities of visual chemotaxis assays (Fig. 2). The Insall chamber represents the most recent of a long evolution of direct-viewing chemotaxis chambers that have been developed over the course of three decades.

Microfluidics has offered many solutions for next-generation chemotaxis assays (reviewed in refs 19 and 20); however, none of these methods have seen widespread adoption at the level of the aforementioned traditional assays. Additionally, efforts to commercialize microfluidic chemotaxis assays—notable products include μ-Slide Chemotaxis (ibidi), Iuvo Chemotaxis Assay Plate (BellBrook Labs), and EZ-TAXIScan (Effector Cell Institute)—have had limited success in the marketplace. The generation of chemical gradient profiles is an area where microfluidic technologies are uniquely qualified because of the highly predictable[21], diffusion-dominant characteristics of the fluid flow at this scale (Box 1). Yet traditional assays are still predominantly used for chemotaxis studies in cell biology research. The low adoption rate of microfluidic chemotaxis assays may be due to the fluid handling expertise and infrastructure required in early designs[22,23], which may have acted as a barrier to entry for biologists[24]. Recently published microfluidic chemotaxis techniques are beginning to take usability requirements into consideration, and demonstrate simpler chemotaxis assay designs that do not require active pumping systems[25–27]. Another possibility is that biologists are more comfortable with using the existing direct-viewing chemotaxis assays that have been developed and vetted over nearly 40 years (Fig. 2). Notably, each iterative improvement on the Zigmond chamber design was published by investigators with appointments in biology (Zigmond); experimental pathology (Boyden and Dunn); and cancer research (Insall)—none of the designs were produced from 'engineering' disciplines. These technical advances were made by biologists to address unmet needs in their own research. And in the case of visual chemotaxis, the methods were, in fact, microfluidic by any reasonable definition, yet they are not typically included within the microfluidic vernacular. In the case of chemotaxis assays, engineers have sometimes erred by imposing technological complexity and functionality where it was not necessarily needed or wanted. This case study illustrates the continuing need for engineers and biologists to work closely during assay development to create usable and robust solutions that build on biologically validated approaches, while adding functionality that allows new avenues of biological inquiry.

Materials tailored for specific applications

Unlike the semiconductor industry where silicon is the backbone material on which the technology has been built[28,29], the materials used for developing microfluidic devices have undergone a large transition over the years. Early μTAS devices were fabricated from silicon[30] and glass[31] using clean-room techniques that were translated to microfluidic device fabrication. This was largely a choice of convenience (because the techniques

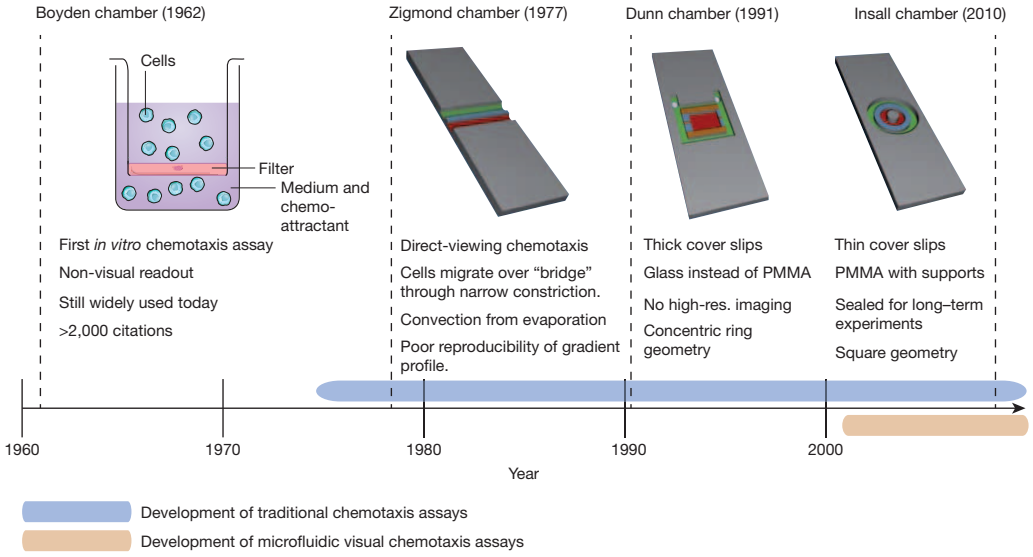


Figure 2 | The development of visual chemotaxis assays over time. The Boyden chamber assay was the first popular *in vitro* chemotaxis technique. The Zigmond chamber design has undergone several evolutionary changes (Dunn and Insall chambers) to address problems with previous versions of the assay. Note the relatively short time for which microfluidics techniques have been available in comparison to the classical visual chemotaxis assays (bars on the timeline compare the development of visual chemotaxis assays). Chamber images adapted with permission from ref. 100, Nature Publishing Group (Boyden chamber) and from ref. 18.

and facilities were already in place) and necessity (early microfluidics focused largely on electrophoretic phenomena where glass is a preferred material), but not a long-term solution for cell biology research. Silicon is opaque to visible and ultraviolet light, making this material incompatible with popular microscopy methods. Glass and silicon are both brittle materials, they have non-trivial bonding protocols for closing microchannels, and in general they require expensive, inaccessible fabrication methods. These materials were well suited for some applications (for example, electrophoresis), but were ultimately limited in their growth potential. Cheaper, more accessible materials and fabrication methods were needed to fuel the growth of microfluidic technology development and adoption.

Elastomeric micromoulding techniques were developed by Bell Labs in the 1970s[32], and first applied to microfluidics and cell biology in the 1980s[33]. In 1998, Whitesides used polydimethylsiloxane (PDMS)—an optically transparent, gas- and vapour-permeable elastomer— for the fabrication of more complex microfluidic devices[34] and helped 'soft lithography' become the most widely adopted method for fabricating microfluidic devices. It would be hard to exaggerate how important and enabling PDMS has been for microfluidics, contributing to the growth of the field in both technological development and number of publications[35]. Adoption of the material can be attributed to several key factors, including (1) the relatively cheap and easy set-up for fabricating small numbers of devices using PDMS in a university setting; (2) the ability to tune the hydrophobic surface properties to become more hydrophilic[36,37]; (3) the ability to reversibly and (in some cases) irreversibly bond PDMS to glass, plastic, PDMS itself, and other materials; and (4) the elasticity of PDMS, which allows for easy removal from delicate silicon moulds for feature replication. In addition to the practical fabrication considerations of using an elastomer, there are also useful functional advantages. Researchers have used the elasticity of PDMS to create micropillar arrays that assay the mechanobiology of various cell types[38,39]. However, perhaps most importantly, the elasticity of PMDS allows for valving and actuation[40,41], which has led to a plethora of microfluidic designs and publications. Fluidigm—the largest commercial μTAS technology company currently in the market—build their microfluidic systems using deformable elastomers (NanoFlex valves).

Despite all the beneficial properties of PDMS that enabled its rapid adoption amongst university engineers, there are several limitations to implementing the material in biomedical research. For example, PDMS has been found to leach uncrosslinked oligomers from the curing process into solution[42], requiring additional device preparation to mitigate this potentially harmful effect[43]. Additionally, PDMS has been shown to absorb small molecules[42,44], which can affect critical cell signalling dynamics. Furthermore, the vapour permeability of PDMS means that evaporation can occur in an experiment[45], which can be detrimental for cell microenvironments at micro- and nanolitre fluid volumes[46,47]. Strategies such as parylene coating the microchannel surface[48] and other techniques[49,50] have been developed to mitigate these problems, but these processes are consequences of deploying a non-ideal material for cell biology applications—the often cited 'biocompatability' of PDMS appears to be something of a misnomer. Last, the manufacture and distribution of PDMS devices to collaborators is not easily scalable, because high-throughput methods such as injection moulding, rolling and embossing cannot be used for PDMS devices. Thus, making PDMS prototypes for iterating on a new design concept is relatively easy, but making many of these devices and packaging them for collaborators or commercialization is non-trivial[51]. Given these limitations, clearly PDMS is not a one-size-fits-all material for all microfluidic applications, and particularly for cell biology research[52].

The limitations of PDMS have prompted researchers to explore alternative materials in recent years (Fig. 3). In the microfluidics community, there has been a push towards the use of thermoplastics such as polystyrene and cyclic olefin copolymer[53] for microfluidic devices (Fig. 3A), although some research laboratories have always used these materials in lieu of PDMS[54,55]. Thermoplastic materials such as polymethyl methacrylate and polycarbonate[56,57] were popular for the fabrication of μTAS

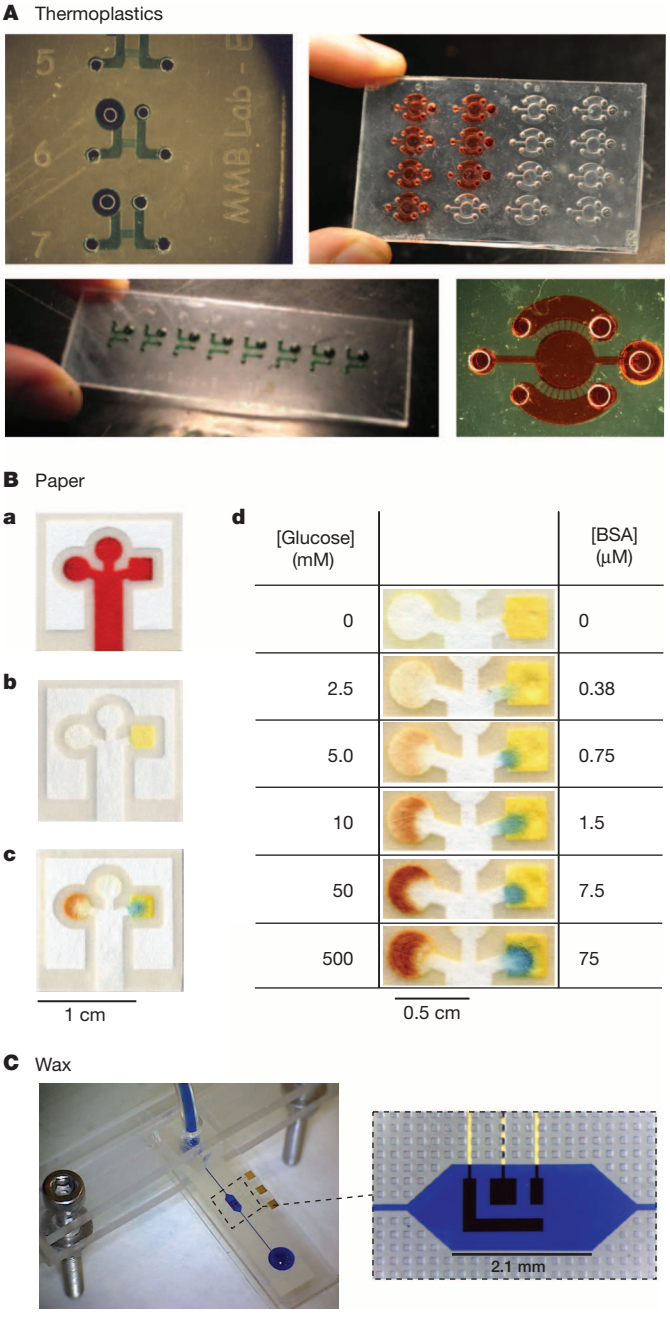


Figure 3 | Materials other than PMDS are being used for microfluidic device design. **A**, Several research groups have demonstrated accessible methods of thermoplastic microfluidic device fabrication. Examples of various microfluidic designs fabricated in polystyrene are shown. **B**, **C**, Paper (**B**), and to a lesser extent, wax (**C**) are being used in the developing world for diagnostic applications owing to benefits in device cost, operation and destructibility with limited waste infrastructure. **B**, An example of a paper-based microfluidic device for detecting glucose and protein. The integrity of the hydrophilic patterning is shown with a red dye (**a**); the detection zones for glucose (circular region on left) and protein (square region on right) are also shown (**b**); and representative tests detecting a single concentration (**c**) and multiple concentrations (**d**) of protein and glucose from an artificial urine sample are also shown. **C**, An example of a wax microfluidic device (zoomed view in inset) that can perform an enzymatic immunoassay. Figure sources, used with permission: **A**, ref. 59; **B**, ref. 64; **C**, ref. 101. **B** is adapted with permission from Martinez, A. W., Phillips, S. T., Whitesides, G. M. & Carrilho, E. Diagnostics for the developing world: microfluidic paper-based analytical devices. *Analytical Chemistry* **82,** 3–10 (2010). Copyright (2010) American Chemical Society.

devices in the 90s, but lost favour with researchers because the fabrication methods were more difficult and expensive than those of PDMS for the typical academic laboratory. However, the microfluidics community has addressed this issue by developing more accessible fabrication methods for thermoplastic μTAS devices[58–60], although these techniques are not without limitations[35,61]. We have recently argued that polystyrene should be preferred over PDMS for many cell biology applications, particularly because biologists have a long history of using polystyrene for cell culture[35]. Furthermore, the use of polystyrene mitigates or eliminates many material property issues associated with PDMS, including the bulk absorption of small molecules and evaporation through the device, and polystyrene makes handling and packaging easier for use in collaborations.

In addition to thermoplastic materials, there has been substantial progress in using destructible, cheap materials such as paper (Fig. 3B), wax (Fig. 3C) and cloth[62] for point-of-care applications in low-resource settings. These materials have the benefit of being cheap and easily incinerated[63], making them ideal choices for settings where safe disposal of biological samples is challenging[3,64]. Currently there is increasing activity in developing microfluidic paper-based analytical devices (μPADs). These μPAD devices are expansions on tried-and-tested lateral flow assays (for example, pregnancy strip test) and operate by passively wicking biological samples through patterned hydrophilic regions using capillary forces; they often use colorimetric readouts. The hydrophobic channel patterning can be accomplished using a variety of methods, such as wax printing[65], photolithographic patterning of photoresist[66], inkjet printing of PMDS[67], and flexographically printed polystyrene[68]. μPAD devices are becoming increasingly sophisticated[69,70], with a recent study demonstrating a single-step enzyme-linked immunosorbent assay (ELISA) for the detection of human chorionic gonadotropin[71].

The movement beyond PDMS with the use of thermoplastics and other materials is a positive development for the microfluidics community. Rather than solely relying on PDMS for device fabrication regardless of its limitations, researchers are beginning to consider new materials that more suitably meet the requirements of biological assays and are amendable to high-throughput manufacturing. The shift to materials beyond PDMS enables researchers to more effectively export technologies in scale, and allows for new solutions to problems in performing cell biology and diagnostic assays. However, different materials often require a re-thinking of component design. For example, it is difficult to implement the displacement valves and pumps so ubiquitous in PDMS devices in other non-elastic materials. Therefore, technological progress using alternative materials will require creative new approaches from engineers that design powerful and user-friendly μTAS devices.

When μTAS technologies are the only solution

Most of the microfluidic technologies that were developed for cell biology applications in the early 2000s sought to improve on existing macroscale assays. Many of these technologies delivered on the promised performance improvements, yet were never adopted by mainstream biology researchers. Another possible reason for this lack of adoption, beyond those we have previously discussed, is that these technologies are improvements on established techniques. Although microfluidic methods may in some cases be technologically superior, they are often only iterative improvements on methods that already exist. Someone interested in performing protein analysis might conduct a western blot or ELISA. To study cell chemotaxis, a researcher might perform a Transwell assay. To investigate tissue regeneration after a wound, an investigator might scratch some cells with a micropipette tip and see what happens. Microfluidic techniques exist that perform many of these assays with equivalent or improved performance[24], but they have not offered fundamentally new capabilities compared to the current state-of-the-art.

Within the past several years there have been a growing number of microfluidic technologies that solve problems that have not yet been addressed by macroscale approaches. Two recognizable examples that embody this distinction can be found in the glucometer and the pregnancy test (or more broadly, lateral flow assays). Each test passively wicks bodily fluids into porous materials, either blood (glucometer) or urine (pregnancy test), and performs a previously complex biochemical assay in a single step to provide an immediate measurement. Although there were benchtop assays that could perform these tasks, the portability and rapid feedback these assays provided was transformative for the end user. There are currently applications like these for which microfluidic methods have demonstrable advantages over traditional methods. These various applications share overlapping qualities that make them potentially useful techniques. However, for the purpose of this discussion, we will break them into three categories: diagnostic devices for low-resource settings; the rapid processing of biofluids for research and clinical applications; and more physiologically relevant *in vitro* models for drug discovery, diagnostics and research applications.

Diagnostics for low-resource settings

The western model of centralized laboratories processing clinical samples with expensive equipment does not translate well to the developing world. Many low-resource settings do not have the means or infrastructure to perform these tests and analyses, necessitating creative alternative solutions to meet this largely unsolved problem. Microfluidic methods are being developed to perform a variety of diagnostic tests with built-in analysis capabilities that are compatible with the infrastructure in the developing world (Fig. 4). As discussed earlier, new material systems such as paper, wax and others are being explored in this area[53,64,72]. Common themes with these devices include being ultra-simple to operate and the provision of some qualitative or quantitative output that can be measured with low-cost and ubiquitous equipment (for example, a mobile-phone camera or scanner). Also, ideally, the materials used to make the devices are easily destructible to avoid unsafe contamination, and are cheap and scalable to manufacture (preferably locally). In a recent study, Chin *et al.*[73] aimed to meet these requirements in a microfluidic chip that performs an ELISA-like assay within ~20 minutes using volumes of blood that can be obtained from a lancet puncture (Fig. 4). Importantly, the assay did not require external pumping systems; it emphasized straightforward operation; and it used cheap photodetectors for the rapid optical readout. The authors analysed more than 70 blood samples obtained from a hospital in Rwanda and successfully diagnosed human immunodeficiency virus (HIV) in all but one patient, achieving sensitivity and specificity values that rival a laboratory-based ELISA test. This study and others are promising indications that μTAS technologies could make meaningful contributions to healthcare in the developing world.

Low cost is arguably the most important feature when aiming to increase access to diagnostics in the developing world, but it is also an increasingly important factor in the developed world. If we can achieve appropriate performance/cost combinations for the developing world, many believe that these technologies will play an important role in transforming the way medicine is delivered in the developed world by enabling in-home testing and treatment. However, traditional lateral flow assays achieve a low-cost/high-performance benchmark, and thus represent a high standard against which new approaches are compared.

Rapidly assaying biofluids with microfluidics

Engineers have made use of properties unique to the microscale to enable studies that would be difficult or impossible using macroscale approaches (Fig. 5). These methods have found clinical applications, because they use ultra-low volumes of biofluids for the sample processing and can usually be accomplished rapidly and easily. To some degree these assays mimic what macroscale assays accomplish, but the methods offer new approaches that enable fundamentally new applications. For example, the rapid purification and analysis of neutrophils—the phagocytotic cells that are first responders for the innate immune system—have been demonstrated in several studies in recent years for clinical and research applications[26,74,75]. Importantly, these techniques reduce blood processing times from roughly an hour (using millilitres of blood from a venipuncture[76]) to a few minutes (using only microlitres of blood from a finger prick). Thus, the methods can be applied to measure neutrophil

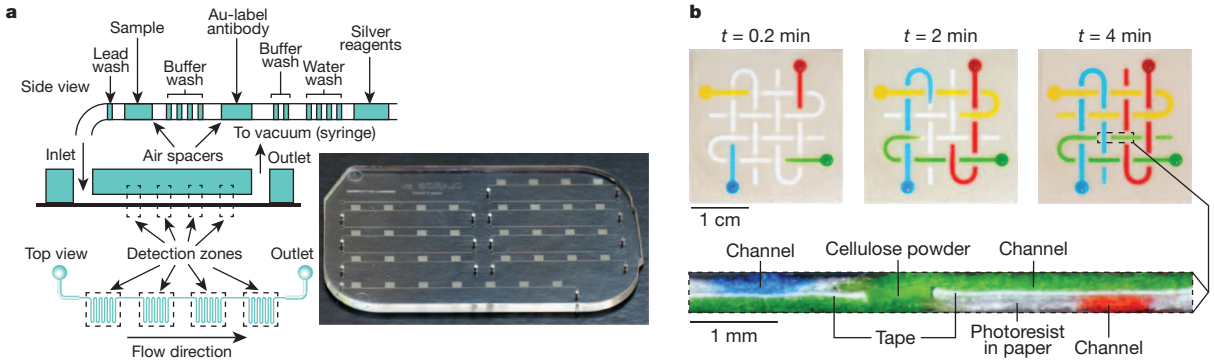


Figure 4 | Diagnostics in the developing world. These are excellent examples of exploiting the benefits of μTAS technologies where classical (Western) diagnostic paradigms fail. **a**, A user-friendly cartridge to perform enzyme-linked immunosorbent assays (ELISAs) for the diagnosis of HIV and other diseases. A schematic showing the functional steps of the assay is shown on the left and the microfluidic device is shown on the right. **b**, 3D μPAD showing complex fluid handling operations that occur passively in a paper device for diagnostics in resource-limited settings. The example shows the flow of several coloured dyes in a patterned μPAD device, with a cross-section of the 3D structure also shown. Figure sources, used with permission: **a**, ref. 73; **b**, ref. 64. **b** is reprinted with permission from Martinez, A. W., Phillips, S. T., Whitesides, G. M. & Carrilho, E. Diagnostics for the developing world: microfluidic paper-based analytical devices. *Analytical Chemistry* **82,** 3–10 (2010). Copyright (2010) American Chemical Society.

function for diagnostic and research purposes, enabling a new class of studies that have previously been beyond the capabilities of macroscale methods[26]. Other purification schemes have been developed that take advantage of the increased dominance of surface tension at the microscale to sort target analytes in biofluids across multiphase barriers (for example, oil and water; see Box 1) using fast and simple procedures[77,78]. Not only is this purification scheme simpler and faster than most macroscale methods, but improved sensitivities for protein and genetic purifications may be achievable owing to a reduction in the number of wash cycles required to carry out an experiment. These applications are only some of the examples where microscale benefits are being used to perform experiments that are not reasonably achievable using macroscale techniques.

More physiologically relevant *in vitro* models

The pharmaceutical industry is currently faced with unsustainable research and development (R&D) costs[79,80] that require it to change how the development and approval of new drugs are pursued[81–83]. The industry faces multiple headwinds, such as the exclusivity on blockbuster drugs soon expiring for several companies, and dramatically fewer new drugs being approved by the Food and Drug Administration (FDA) in recent years. These circumstances necessitate new strategies for drug development that increase R&D productivity in order to avoid a potential drought in effective new drugs coming to market.

Microfluidics researchers are taking aim at this problem by developing potentially transformative technologies to mitigate the cost of new drug development. A new class of microfluidic devices seeks to replicate *in vivo* organ function on a microchip (Fig. 6). This new class of so called 'organ-on-a-chip' technologies integrates several well-understood microfluidic components into a single *in vitro* device, allowing researchers to more closely recapitulate *in vivo* function (both normal and disease states). This ambitious effort is still in its infancy, though several promising studies have developed examples of these biomimetic systems. Examples of organ (or disease)-on-a-chip technologies include gut-on-a-chip[84], lung-on-a-chip[85], blood vessel-on-a-chip[86–88], cancer-on-a-chip[89–91] and kidney-on-a-chip[92]. Furthermore, these modular systems could theoretically be combined into a complete 'human-on-a-chip' model that mimics *in vivo* function of these organs working in concert[93]. The result would be a class

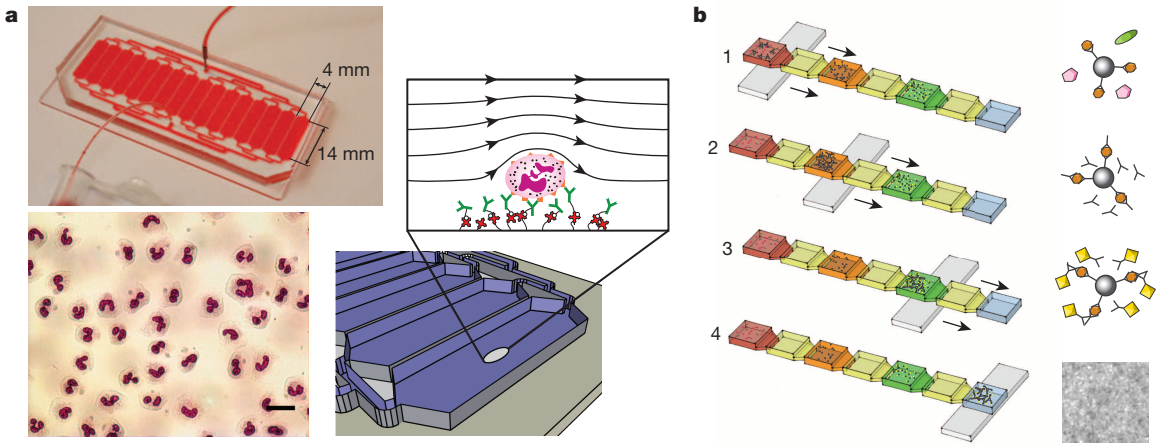


Figure 5 | Rapid purification microfluidic systems. **a**, A microfluidic device to purify neutrophils within minutes using antibody-based capture for subsequent diagnostic or research analysis. The microfluidic device is shown at the upper left with stained neutrophils that have been sorted from whole blood below (scale bar = 20 μm); an illustration of the neutrophils captured within the microchannels by antibodies (zoomed view in inset) is also shown. **b**, A technique to purify target analytes such as RNA, cells and proteins by simply sliding a magnet across an immiscible aqueous–oil interface. An example shown here illustrates four steps to purify protein from a sample (zoomed view to the right shows detail) by (1) removing the analyte bound to paramagnetic particles across the first aqueous-oil barrier, (2) binding a primary antibody to the analyte and dragging it across another aqueous-oil barrier, (3) binding a fluorescently labelled secondary antibody to the complex and bringing it across another aqueous-oil barrier into the imaging well (4), where the fluorescence is measured to detect the amount of analyte (white and grey fluorescent image). Figure sources, used with permission: **a**, ref. 74; **b**, ref. 10. **b** is adapted with permission from Berry, S. M., Maccoux, L. J. & Beebe, D. J. Streamlining immunoassays with immiscible filtrations assisted by surface tension. *Analytical Chemistry* **84,** 5518–5523 (2012). Copyright 2012 American Chemical Society.

of sophisticated *in vitro* assays with which drugs could be tested, in the hope of increasing the predictability of a new drug (that is, hit rate) before animal testing (possibly even replacing animal trials) and human clinical trials. In a tangential application, blood vessel-on-a-chip devices have already been used for the diagnosis of sickle cell disease in the clinic[86,94]. For example, Tsai *et al.*[86] described a microfluidic chip that recapitulated *in vivo* conditions of a blood vessel—such as blood flow rate, endothelial cell shear stress and biochemical activation states—in order to reliably detect vascular occlusions due to sickle-cell disease. This system highlights how certain properties of microfluidic systems, such as high-resolution micropatterning and precise control of the haemodynamic and shear profiles in the microchip, enabled the measurement of biophysical abnormalities in a clinical setting. Much more work is still required before organ-on-a-chip methods can be adopted in mainstream drug R&D, although early developments in this area are promising. Indeed, AstraZeneca—a multinational pharmaceutical company—has recently announced a collaboration with Harvard's Wyss Institute to research the integration of organ-on-a-chip technologies into their drug development.

Where we go from here

The question of how to increase the adoption of microfluidic technologies in mainstream biomedical research remains largely unanswered, and we argue there are no guaranteed routes to achieve adoption. We have shown that microfluidic technologies are being used for some studies in biology research and diagnostic applications; however, the large majority of microfluidics publications are still in technical journals specific to the field (Fig. 1). Adoption of new technologies that supplant or even complement existing methods is often a slow process. For evidence of this, we consider the computer mouse, which took 20 years to appear in the Macintosh computer after its invention by Engelbart in the 1960s. But this does not mean microfluidics engineers should become disillusioned or discouraged. Researchers in the field must develop deliberate and thoughtful strategies that will best push the technology forward. We now have several decades of experience to draw on, and there are some useful lessons we can apply.

Fostering mutually beneficial collaborations

During the early years of microfluidics, the field did not have a successful strategy for transferring technological developments to non-engineering users. Perhaps the idea was that researchers from the biology community would rush to work out how to make use of these new technologies. Clearly this formula of engineers and biologists leading separate academic lives does not benefit either community. Fortunately researchers have acknowledged that a divide between the developers of the technology and the end-users is counterproductive. Most of the recent microfluidics papers published in 'Biology and medicine' journals are co-authored by engineers, biologists and clinicians. This evidence of increasing collaboration is a promising development for everyone involved. In order to sustain this trend, microfluidic researchers should court collaborators from biology and clinical laboratories (and vice versa). Direct interaction and feedback from the end-user is tremendously beneficial during technology development. Furthermore, new applications and ideas can be generated from biology collaborators that engineers—being non-experts in cell biology or clinical research—would never have considered.

The simplest solution is almost always best

All the signs indicate that there is no simple solution for accelerating the adoption process; however, there are design choices engineers can make in order to lower the barrier to entry for biologists. How the end-user interacts with a new technology is a critical aspect of whether the method is adopted. Microfluidics engineers have been attempting to simplify fluid handling challenges in their designs with passive pumping approaches that only require a micropipette to operate[8,25–27,77,78,95]. Additionally, some have explored the use of centrifugal forces to perform complex assays using a 'lab-on-a-CD' design[96]. Many microfluidic applications require the use of external pumps and pneumatic fluid handling systems; examples include most organ-on-a-chip devices and techniques that require continuous flow to generate specific shear profiles (for example, biomimetic blood vessel models). However, engineers should limit the use of these external systems whenever possible. Creating a simpler approach often requires more creative solutions, but this can greatly improve the

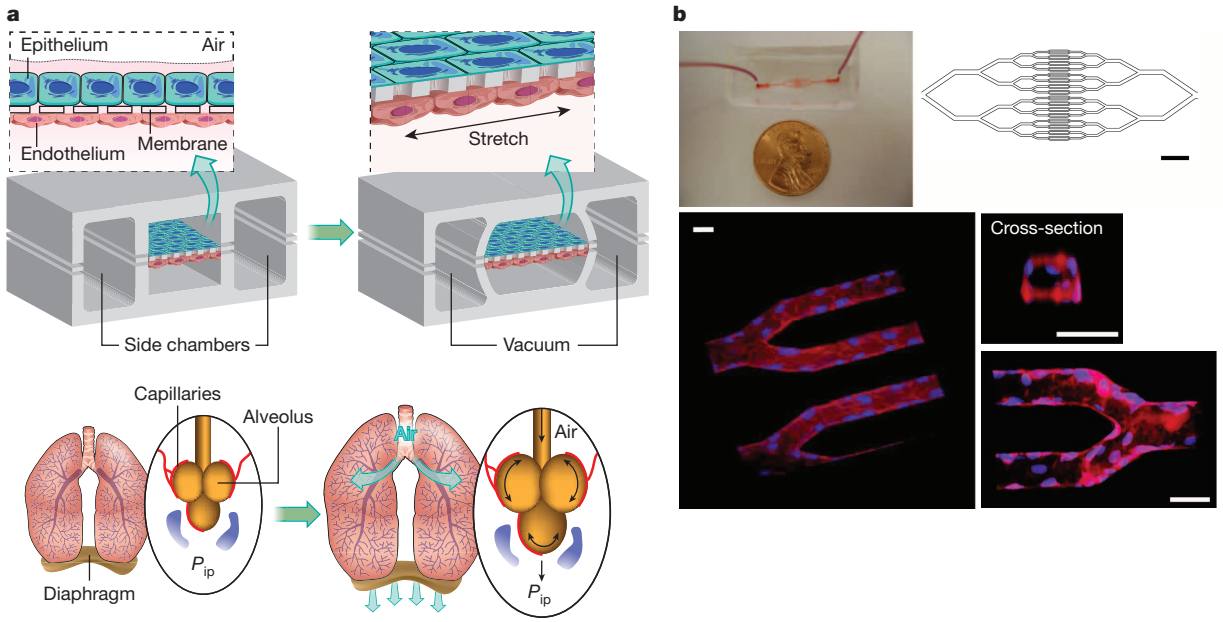


Figure 6 | Organ-on-a-chip assays for drug development and specialized diagnostic applications. **a**, Complex microsystems can be developed to recreate an organ's physiology, such as the physiology of the lung, directly on a microfluidic device. The diagram illustrates a biomimetic microfluidic design that actuates stretching of tissue in a breathing-like manner by using vacuum in side chambers to strain the cell-coated PDMS membrane. This process mimics the reduction in intrapleural pressure (P_{ip}) in the lungs during breathing. **b**, Biomimetic blood vessel and capillary networks can also be recreated *in vitro* to diagnosis SCD and other diseases involving blood vessel–whole blood interactions. An image of the microfluidic device is shown (top) next to a penny for scale, with a diagram at right showing the increasingly narrow capillary network; confocal microscopy images of the endothelial cell-lined lumens within the device are also shown (bottom) with the cell nucleus (blue) and cell membrane (red) visible. Scale bars: 600 μm (black), 30 μm (white). Figure sources, reprinted with permission: **a**, ref. 85; **b**, ref. 86.

experience for the end-user. Paper diagnostic assays are an excellent example of single-step, automated and user-friendly μTAS solutions where the technology is not visible and the user can focus on interpreting the results[64]. We have recently developed a similarly straightforward, automated approach for general cell biology applications that does not require external pumping equipment or even a micropipette to perform complex assay protocols[97]. General problems of packaging and distributing microfluidic technologies to collaborators will also need to be addressed until microfluidic assays become more commercially viable in the academic research market. These problems should be viewed through the lens of user-friendly assay design.

Finding the right problems to solve

The case study we have used (chemotaxis assays) helps to illustrate how competing technology platforms continue to improve over time as microfluidic technologies develop. Some of the touted advantages of microfluidic systems that existed 20 years ago are not as stark today because technological improvements have been made to more traditional and widely accepted assays, often narrowing the initially perceived performance advantage of microfluidic solutions. This evolution in the technology landscape highlights the need for finding the right problems in biology and medicine to solve with microfluidic approaches. For example, microfluidic solutions have advantages over many technologies for diagnostics in the developing world. However commercializing these technologies is challenging because, by definition, the desired diagnostic devices will not generate much revenue or profit. So the breadth and depth of impact may be great for this particular application, but a disconnect exists between development and commercialization. Likewise, there may be niche biological questions that can be addressed using microfluidic methods, but for which broad commercial markets do not exist. A key consideration in the development of new microfluidic methods in academic research should be whether the use of microfluidics introduces truly enabling functionality compared to current methods. When a potential application passes this test, the chances of contributing useful technology to the field are substantially higher.

Received 2 August 2013; accepted 31 January 2014.

1. Beebe, D. J., Mensing, G. A. & Walker, G. M. Physics and application of microfluidics in biology. *Annu. Rev. Biomed. Eng.* **4,** 261–286 (2002).
2. Hansen, C. & Quake, S. R. Microfluidics in structural biology: smaller, faster…better. *Curr. Opin. Struct. Biol.* **13,** 538–544 (2003).
3. Yager, P. *et al.* Microfluidic diagnostic technologies for global public health. *Nature* **442,** 412–418 (2006).
4. El-Ali, J., Sorger, P. & Jensen, K. Cells on chips. *Nature* **442,** 403–411 (2006).
5. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442,** 368–373 (2006).
6. Manz, A., Graber, N. & Widmer, H. M. Miniaturized total chemical-analysis systems — a novel concept for chemical sensing. *Sens. Actuators B* **1,** 244–248 (1990).
7. Reyes, D. R., Iossifidis, D., Auroux, P.-A. & Manz, A. Micro total analysis systems. 1. Introduction, theory, and technology. *Anal. Chem.* **74,** 2623–2636 (2002).
This pioneering publication described the concept of a μTAS device.
8. Walker, G. A passive pumping method for microfluidic devices. *Lab Chip* **2,** 131–134 (2002).
Describes a method to passively pump fluids within microchannels using only a micropipette.
9. Lee, S. H. *et al.* Capillary based patterning of cellular communities in laterally open channels. *Anal. Chem.* **82,** 2900–2906 (2010).
10. Berry, S. M., Maccoux, L. J. & Beebe, D. J. Streamlining immunoassays with immiscible filtrations assisted by surface tension. *Anal. Chem.* **84,** 5518–5523 (2012).
11. Anna, S. L., Bontoux, N. & Stone, H. A. Formation of dispersions using 'flow focusing' in microchannels. *Appl. Phys. Lett.* **82,** 364–366 (2003).
12. Whitesides, G. M. Cool, or simple and cheap? Why not both? *Lab Chip* **13,** 11–13 (2012).
13. Blow, N. Microfluidics: in search of a killer application. *Nature Methods* **4,** 665–670 (2007).
14. Becker, H. Hype, hope and hubris: the quest for the killer application in microfluidics. *Lab Chip* **9,** 2119–2122 (2009).
15. Boyden, S. The chemotactic effect of mixtures of antibody and antigen on polymorphonuclear leucocytes. *J. Exp. Med.* **115,** 453–466 (1962).
16. Zigmond, S. H. Ability of polymorphonuclear leukocytes to orient in gradients of chemotactic factors. *J. Cell Biol.* **75,** 606–616 (1977).
17. Zicha, D., Dunn, G. A. & Brown, A. F. A new direct-viewing chemotaxis chamber. *J. Cell Sci.* **99,** 769–775 (1991).
18. Muinonen-Martin, A. J. A., Veltman, D. M. D., Kalna, G. G. & Insall, R. H. R. An improved chamber for direct visualisation of chemotaxis. *PLoS ONE* **5,** e15309 (2010).
19. Keenan, T. M. & Folch, A. Biomolecular gradients in cell culture systems. *Lab Chip* **8,** 34–57 (2008).
A detailed review of microfluidic chemical gradient generators.
20. Irimia, D. Microfluidic technologies for temporal perturbations of chemotaxis. *Annu. Rev. Biomed. Eng.* **12,** 259–284 (2010).
21. Kamholz, A. E. & Yager, P. Theoretical analysis of molecular diffusion in pressure-driven laminar flow in microfluidic channels. *Biophys. J.* **80,** 155–160 (2001).
22. Jeon, N. L. *et al.* Generation of solution and surface gradients using microfluidic systems. *Langmuir* **16,** 8311–8316 (2000).
23. Irimia, D., Geba, D. A. & Toner, M. Universal microfluidic gradient generator. *Anal. Chem.* **78,** 3472–3477 (2006).
24. Paguirigan, A. L. & Beebe, D. J. Microfluidics meet cell biology: bridging the gap by validation and application of microscale techniques for cell biological assays. *Bioessays* **30,** 811–821 (2008).
25. Butler, K. L. *et al.* Burn injury reduces neutrophil directional migration speed in microfluidic devices. *PLoS ONE* **5,** e11921 (2010).
26. Sackmann, E. K. *et al.* Microfluidic kit-on-a-lid: a versatile platform for neutrophil chemotaxis assays. *Blood* **120,** e45–e53 (2012).
27. Jowhar, D., Wright, G., Samson, P. C., Wikswo, J. P. & Janetopoulos, C. Open access microfluidic device for the study of cell migration during chemotaxis. *Integr. Biol.* **2,** 648–658 (2010).
28. Kilby, J. S. Miniaturized electronic circuits. US Patent 3,138,743 (issued 23 June 1964).
29. Noyce, R. N. Semiconductor device-and-lead structure. US Patent 2,981,877 (issued 25 April 1961).
30. Van Lintel, H., Vandepol, F. & Bouwstra, S. A piezoelectric micropump based on micromachining of silicon. *Sens. Actuators* **15,** 153–167 (1988).
31. Harrison, D. J., Manz, A., Fan, Z. H., Ludi, H. & Widmer, H. M. Capillary electrophoresis and sample injection systems integrated on a planar glass chip. *Anal. Chem.* **64,** 1926–1932 (1992).
32. Aumiller, G. D., Chandross, E. A., Tomlinson, W. J. & Weber, H. P. Submicrometer resolution replication of relief patterns for integrated optics. *J. Appl. Phys.* **45,** 4557–4562 (1974).
An early example of replicating microfluidic structures with elastomer materials.
33. Masuda, M., Masao, W. & Nanba, T. Novel method of cell fusion in field constriction area in fluid integrated circuit. *IEEE Trans. Ind. Appl.* **25,** 732–737 (1989).
34. Duffy, D. C. D., McDonald, J. C. J., Schueller, O. J. O. & Whitesides, G. M. G. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Anal. Chem.* **70,** 4974–4984 (1998).
35. Berthier, E., Young, E. W. K. & Beebe, D. Engineers are from PDMS-land, Biologists are from Polystyrenia. *Lab Chip* **12,** 1224–1237 (2012).
36. Chaudhury, M. K. & Whitesides, G. M. Direct measurement of interfacial interactions between semispherical lenses and flat sheets of poly(dimethylsiloxane) and their chemical derivatives. *Langmuir* **7,** 1013–1025 (1991).
37. Morra, M. *et al.* On the aging of oxygen plasma-treated polydimethylsiloxane surfaces. *J. Colloid Interface Sci.* **137,** 11–24 (1990).
38. Yang, M. T., Fu, J., Wang, Y.-K., Desai, R. A. & Chen, C. S. Assaying stem cell mechanobiology on microfabricated elastomeric substrates with geometrically modulated rigidity. *Nature Protocols* **6,** 187–213 (2011).
39. Choi, C. K., Breckenridge, M. T. & Chen, C. S. Engineered materials and the cellular microenvironment: a strengthening interface between cell biology and bioengineering. *Trends Cell Biol.* **20,** 705–714 (2010).
40. Quake, S. R. From micro- to nanofabrication with soft materials. *Science* **290,** 1536–1540 (2000).
41. Unger, M. A. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288,** 113–116 (2000).
42. Regehr, K. J. *et al.* Biological implications of polydimethylsiloxane-based microfluidic cell culture. *Lab Chip* **9,** 2132–2139 (2009).
A study that details the biological implications of using PDMS in cell biology research.
43. Lee, J. N., Park, C. & Whitesides, G. M. Solvent compatibility of poly(dimethylsiloxane)-based microfluidic devices. *Anal. Chem.* **75,** 6544–6554 (2003).
44. Toepke, M. W. & Beebe, D. J. PDMS absorption of small molecules and consequences in microfluidic applications. *Lab Chip* **6,** 1484–1486 (2006).
45. Berthier, E., Warrick, J., Yu, H. & Beebe, D. J. Managing evaporation for more robust microscale assays. Part 1. Volume loss in high throughput assays. *Lab Chip* **8,** 852–859 (2008).
46. Wu, M.-H. M., Dimopoulos, G. G., Mantalaris, A. A. & Varley, J. J. The effect of hyperosmotic pressure on antibody production and gene expression in the GS-NS0 cell line. *Biotechnol. Appl. Biochem.* **40,** 41–46 (2004).
47. deZengotita, V., Kimura, R. & Miller, W. M. Effects of CO_2 and osmolality on hybridoma cells: growth, metabolism and monoclonal antibody production. *Cytotechnology* **28,** 213–227 (1998).
48. Heo, Y. S. *et al.* Characterization and resolution of evaporation-mediated osmolality shifts that constrain microfluidic cell culture in poly(dimethylsiloxane) devices. *Anal. Chem.* **79,** 1126–1134 (2007).
49. Futai, N., Gu, W., Song, J. W. & Takayama, S. Handheld recirculation system and customized media for microfluidic cell culture. *Lab Chip* **6,** 149–154 (2006).

50. Chang, W.-J., Akin, D., Sedlak, M., Ladisch, M. R. & Bashir, R. Poly (dimethylsiloxane)(PDMS) and silicon hybrid biochip for bacterial culture. *Biomed. Microdevices* **5,** 281–290 (2003).
51. Becker, H. It's the economy… *Lab Chip* **9,** 2759–2762 (2009).
52. Mukhopadhyay, R. When PDMS isn't the best. *Anal. Chem.* **79,** 3248–3253 (2007).
53. Chin, C. D. *et al.* Microfluidics-based diagnostics of infectious diseases in the developing world. *Nature Med.* **17,** 1015–1019 (2011).
A study that diagnosed HIV from blood samples in Rwanda using a simple microfluidic chip.
54. Henry, A. C. *et al.* Surface modification of poly(methyl methacrylate) used in the fabrication of microanalytical devices. *Anal. Chem.* **72,** 5331–5337 (2000).
55. Browne, A. W., Rust, M. J., Jung, W., Lee, S. H. & Ahn, C. H. A rapid prototyping method for polymer microfluidics with fixed aspect ratio and 3D tapered channels. *Lab Chip* **9,** 2941–2946 (2009).
56. Becker, H. & Heim, U. Hot embossing as a method for the fabrication of polymer high aspect ratio structures. *Sens. Actuators A* **83,** 130–135 (2000).
57. Martynova, L. *et al.* Fabrication of plastic microfluid channels by imprinting methods. *Anal. Chem.* **69,** 4783–4789 (1997).
58. Goral, V. N., Hsieh, Y.-C., Petzold, O. N., Faris, R. A. & Yuen, P. K. Hot embossing of plastic microfluidic devices using poly(dimethylsiloxane) molds. *J. Micromech. Microeng.* **21,** 017002 (2011).
59. Young, E. W. K. *et al.* Rapid prototyping of arrayed microfluidic systems in polystyrene for cell-based assays. *Anal. Chem.* **83,** 1408–1417 (2011).
60. Wang, Y. *et al.* Benchtop micromolding of polystyrene by soft lithography. *Lab Chip* **11,** 3089–3097 (2011).
61. Young, E. W. K., Berthier, E. & Beebe, D. J. Assessment of enhanced autofluorescence and impact on cell microscopy for microfabricated thermoplastic devices. *Anal. Chem.* **85,** 44–49 (2013).
62. Nilghaz, A. *et al.* Flexible microfluidic cloth-based analytical devices using a low-cost wax patterning technique. *Lab Chip* **12,** 209–218 (2011).
63. von Lode, P. Point-of-care immunotesting: approaching the analytical performance of central laboratory methods. *Clin. Biochem.* **38,** 591–606 (2005).
64. Martinez, A. W., Phillips, S. T., Whitesides, G. M. & Carrilho, E. Diagnostics for the developing world: microfluidic paper-based analytical devices. *Anal. Chem.* **82,** 3–10 (2010).
65. Carrilho, E., Martinez, A. W. & Whitesides, G. M. Understanding wax printing: a simple micropatterning process for paper-based microfluidics. *Anal. Chem.* **81,** 7091–7095 (2009).
66. Martinez, A. W., Phillips, S. T., Wiley, B. J., Gupta, M. & Whitesides, G. M. FLASH: A rapid method for prototyping paper-based microfluidic devices. *Lab Chip* **8,** 2146–2150 (2008).
67. Abe, K., Suzuki, K. & Citterio, D. Inkjet-printed microfluidic multianalyte chemical sensing paper. *Anal. Chem.* **80,** 6928–6934 (2008).
68. Olkkonen, J., Lehtinen, K. & Erho, T. Flexographically printed fluidic structures in paper. *Anal. Chem.* **82,** 10246–10250 (2010).
69. Fu, E. *et al.* Two-dimensional paper network format that enables simple multistep assays for use in low-resource settings in the context of malaria antigen detection. *Anal. Chem.* **84,** 4574–4579 (2012).
70. Martinez, A. W., Phillips, S. T. & Whitesides, G. M. Three-dimensional microfluidic devices fabricated in layered paper and tape. *Proc. Natl Acad. Sci. USA* **105,** 19606–19611 (2008).
A review of advancements in μPAD devices for diagnostics in developing regions.
71. Apilux, A., Ukita, Y., Chikae, M., Chailapakul, O. & Takamura, Y. Development of automated paper-based devices for sequential multistep sandwich enzyme-linked immunosorbent assays using inkjet printing. *Lab Chip* **13,** 126–135 (2012).
72. Watkins, N. N. *et al.* Microfluidic $CD4^+$ and $CD8^+$ T lymphocyte counters for point-of-care HIV diagnostics using whole blood. *Sci. Transl. Med.* **5,** 214ra170 (2013).
73. Chin, C. D. *et al.* Microfluidics-based diagnostics of infectious diseases in the developing world. *Nature Med.* **17,** 1015–1019 (2011).
74. Kotz, K. T. *et al.* Clinical microfluidics for neutrophil genomics and proteomics. *Nature Med.* **16,** 1042–1047 (2010).
This study investigated the relationship between protein/genetic information and the clinical condition of burn patients using a simple microfluidic device.
75. Warner, E. A. *et al.* Microfluidics-based capture of human neutrophils for expression analysis in blood and bronchoalveolar lavage. *Lab. Invest.* **91,** 1787–1795 (2011).
76. Bach, M. K. & Brashler, J. R. Single-step separation of red blood cells, granulocytes and mononuclear leukocytes on discontinuous density gradients of Ficoll-Hypaque. *Exp. Cell Res.* **61,** 387–396 (1970).
77. Berry, S. M., Alarid, E. T. & Beebe, D. J. One-step purification of nucleic acid for gene expression analysis via Immiscible Filtration Assisted by Surface Tension (IFAST). *Lab Chip* **11,** 1747–1753 (2011).
78. Berry, S. M., Strotman, L. N., Kueck, J. D., Alarid, E. T. & Beebe, D. J. Purification of cell subpopulations via immiscible filtration assisted by surface tension (IFAST). *Biomed. Microdevices* **13,** 1033–1042 (2011).
79. Cressey, D. Pfizer slashes R&D. *Nature* **470,** 154 (2011).
80. DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **22,** 151–185 (2003).
81. Cressey, D. Traditional drug-discovery model ripe for reform. *Nature* **471,** 17–18 (2011).
82. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Rev. Drug Discov.* **9,** 203–214 (2010).
83. Ljunggren, H. G. Academia and big pharma united. *Sci. Transl. Med.* **6,** 217ed1 (2014).
84. Kimura, H., Yamamoto, T., Sakai, H., Sakai, Y. & Fujii, T. An integrated microfluidic system for long-term perfusion culture and on-line monitoring of intestinal tissue models. *Lab Chip* **8,** 741–746 (2008).
85. Huh, D. *et al.* Reconstituting organ-level lung functions on a chip. *Science* **328,** 1662–1668 (2010).
86. Tsai, M. *et al. In vitro* modeling of the microvascular occlusion and thrombosis that occur in hematologic diseases using microfluidic technology. *J. Clin. Invest.* **122,** 408–418 (2012).
This study utilized precisely patterned microvessels to diagnose vasco-occlusions in patient samples.
87. Bischel, L. L., Young, E. W. K., Mader, B. R. & Beebe, D. J. Tubeless microfluidic angiogenesis assay with three-dimensional endothelial-lined microvessels. *Biomaterials* **34,** 1471–1477 (2013).
88. Song, J. W. & Munn, L. L. Fluid forces control endothelial sprouting. *Proc. Natl Acad. Sci. USA* **108,** 15342–15347 (2011).
89. Walsh, C. L. *et al.* A multipurpose microfluidic device designed to mimic microenvironment gradients and develop targeted cancer therapeutics. *Lab Chip* **9,** 545–554 (2009).
90. Zervantonakis, I. K. *et al.* Three-dimensional microfluidic model for tumor cell intravasation and endothelial barrier function. *Proc. Natl Acad. Sci. USA* **109,** 13515–13520 (2012).
This study describes a system that more closely mimics tumour cell intravasation *in vitro* compared to standard cell biology techniques such as modified Transwell assays.
91. Sung, K. E. *et al.* Transition to invasion in breast cancer: a microfluidic *in vitro* model enables examination of spatial and temporal effects. *Integr. Biol.* **3,** 439–450 (2011).
92. Jang, K.-J. & Suh, K.-Y. A multi-layer microfluidic device for efficient culture and analysis of renal tubular cells. *Lab Chip* **10,** 36–42 (2010).
93. Huh, D., Hamilton, G. A. & Ingber, D. E. From 3D cell culture to organs-on-chips. *Trends Cell Biol.* **21,** 745–754 (2011).
94. Wood, D. K., Soriano, A., Mahadevan, L., Higgins, J. M. & Bhatia, S. N. A biophysical indicator of vaso-occlusive risk in sickle cell disease. *Sci. Transl. Med.* **4,** 123ra26 (2012).
95. Berthier, E., Warrick, J. & Casavant, B. Pipette-friendly laminar flow patterning for cell-based assays. *Lab Chip* **11,** 2060–2065 (2011).
96. Gorkin, R. *et al.* Centrifugal microfluidics for biomedical applications. *Lab Chip* **10,** 1758–1773 (2010).
97. Berthier, E. *et al.* Kit-On-A-Lid-Assays for accessible self-contained cell assays. *Lab Chip* **13,** 424–431 (2013).
98. Bakkalbasi, N., Bauer, K., Glover, J. & Wang, L. Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomed. Digit. Libr.* doi:10.1186/1742-5581-3-7 (2006).
99. Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. & Pappas, G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J.* **22,** 338–342 (2008).
100. Reymond, N., Borda d'Água, B. & Ridley, A. J. *et al.* Crossing the endothelial barrier during metastasis. *Nature Rev. Cancer* **13,** 858–870 (2013).
101. Díaz-González, M. & Baldi, A. Fabrication of biofunctionalized microfluidic structures by low-temperature wax bonding. *Anal. Chem.* **84,** 7838–7844 (2012).

Acknowledgements We thank S. Berry, B. Casavant, P. Thomas and L. Strotman for discussions during the preparation of this manuscript.

Author Contributions E.K.S. and D.J.B. wrote the manuscript. A.L.F. contributed to the design and execution of the literature searches that measured the quantity of microfluidic publications in various categories.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to D.J.B. (djbeebe@wisc.edu).

ARTICLE

doi:10.1038/nature12930

Alveolar progenitor and stem cells in lung development, renewal and cancer

Tushar J. Desai[1,2], Douglas G. Brownfield[1] & Mark A. Krasnow[1]

Alveoli are gas-exchange sacs lined by squamous alveolar type (AT) 1 cells and cuboidal, surfactant-secreting AT2 cells. Classical studies suggested that AT1 arise from AT2 cells, but recent studies propose other sources. Here we use molecular markers, lineage tracing and clonal analysis to map alveolar progenitors throughout the mouse lifespan. We show that, during development, AT1 and AT2 cells arise directly from a bipotent progenitor, whereas after birth new AT1 cells derive from rare, self-renewing, long-lived, mature AT2 cells that produce slowly expanding clonal foci of alveolar renewal. This stem-cell function is broadly activated by AT1 injury, and AT2 self-renewal is selectively induced by EGFR (epidermal growth factor receptor) ligands *in vitro* and oncogenic *Kras(G12D) in vivo*, efficiently generating multifocal, clonal adenomas. Thus, there is a switch after birth, when AT2 cells function as stem cells that contribute to alveolar renewal, repair and cancer. We propose that local signals regulate AT2 stem-cell activity: a signal transduced by EGFR–KRAS controls self-renewal and is hijacked during oncogenesis, whereas another signal controls reprogramming to AT1 fate.

Pulmonary gas exchange occurs in delicate alveolar sacs lined by two epithelial cell types[1] (Extended Data Fig. 1). Squamous alveolar type (AT) 1 cells mediate gas exchange, whereas cuboidal AT2 cells secrete surfactant that prevents alveolar collapse; AT2 cells are one of the medically most important cells in the neonate and, as described below, one of the most dangerous in adults. Serious diseases including respiratory distress syndrome and idiopathic pulmonary fibrosis involve a failure to establish or maintain AT1 and AT2 cells, and alveoli are a major site of lung cancer, the leading cause of cancer death[2].

Despite their importance, the identity of alveolar progenitor and stem cells is controversial and their activity throughout life uncharted[3,4]. Classical morphologic and autoradiographic studies in rodents suggested that progenitors mature into AT2 cells during development, some of which differentiate into AT1 cells[5]. Maintenance is difficult to study because of slow turnover[6,7], but lineage tag expression in isolated AT1 cells is observed following bulk labelling of AT2 cell populations[8,9]. To circumvent slow turnover, lung injury models have been used and provide evidence that AT2 cells can contribute to alveolar repair[8,10,11]. Currently, six different cell populations have been proposed as alveolar stem cells on the basis of their capacity for clonal propagation and multilineage differentiation in culture[9,12–14]. Transplantation assays have also been used, but because many cells are implanted they cannot assess whether individual cells self-renew and undergo multilineage differentiation[4,15]. When one of these putative stem cell populations was fate-mapped *in vivo*, it failed to demonstrate the multilineage differentiation achieved in culture[16]; similar disparity between *ex vivo* and *in vivo* behaviour of putative stem cells has been found for other organs[17]. Here we use a battery of alveolar markers, lineage tracing and clonal analysis in mice to identify alveolar progenitor and stem cells *in vivo* and map their locations and activity during lung development, maintenance and cancer.

AT1 and AT2 cells arise from a bipotent progenitor

Mature AT1 and AT2 cells appear about 1 day before birth, when distal tubules begin to dilate ('sacculation', Fig. 1a–c)[18,19]. We mapped progression of sacculation in three dimensions by analysing finely staged whole-mount lungs immunostained for E-cadherin (Cdh1) to visualize individual cells (Fig. 1a–c and Extended Data Fig. 1e–f). Dilation begins at the bronchoalveolar junction then progresses distally towards the airway tip (Fig. 1a–c).

The classical model proposing that progenitors in development are pre-AT2 cells is difficult to reconcile with the finding that some AT1 cell markers are expressed up to 5 days before sacculation[20]. To molecularly classify progenitors, we validated 15 extant AT1 and AT2 markers (Supplementary Table 1) then analysed the transition in labelling between distal (progenitors) and proximal (nascent AT1 and AT2 cells) positions in a sacculating airway (Fig. 1d) to infer dynamic expression changes during differentiation (Fig. 1e–p). Markers fell into six expression classes (Extended Data Table 1), distinguishing seven stages in alveolar development (Fig. 1e–p). However, instead of a progenitor to AT2 to AT1 progression, our data support a model in which bipotent progenitors (P) expressing a subset of AT1 (1) and AT2 (2) markers ($P1^E$, $P1^L$, $P2^E$ and $P2^L$) give rise to either AT1 or AT2 cells by shutting off inappropriate cell type markers early (E) or late (L) in differentiation, then turning on cell type-specific late (L) markers ($A1^L$, $A2^L$) as they complete maturation (Fig. 1q). Co-expression of AT1 and AT2 markers by progenitors indicates that these specialized cell types may have evolved from a primordial pneumocyte with features of both, similar to those in less derived vertebrates such as lungfishes[21].

Three additional lines of evidence support the bipotent progenitor model. First, clonal analysis of individual distal airway epithelial tip cells[22] labelled on embryonic day (E) 15 using an inducible Cre recombinase (linked with the oestrogen receptor (ER), Shh–Cre–ER) demonstrated localized alveolar lineage clusters with marked AT1 and AT2 cells (Extended Data Fig. 2a,b), confirming that individual cells are bipotent. Second, ultrastructural analysis of early sacculation revealed three classes of distal epithelial cells (Fig. 1r–u): cuboidal cells with glycogen vacuoles but no lamellar bodies (bipotent progenitors), cuboidal cells with vacuoles and lamellar bodies (early AT2 cells), and partially flattened cells with vacuoles (early AT1 cells). We never observed partially flattened cells with lamellar bodies, the presumed AT2-to-AT1 intermediate predicted by the classical model (Extended Data Fig. 1g). Third, lineage

[1]Department of Biochemistry and Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305-5307, USA. [2]Department of Internal Medicine, Division of Pulmonary and Critical Care, Stanford University School of Medicine, Stanford, California 94305-5307, USA.

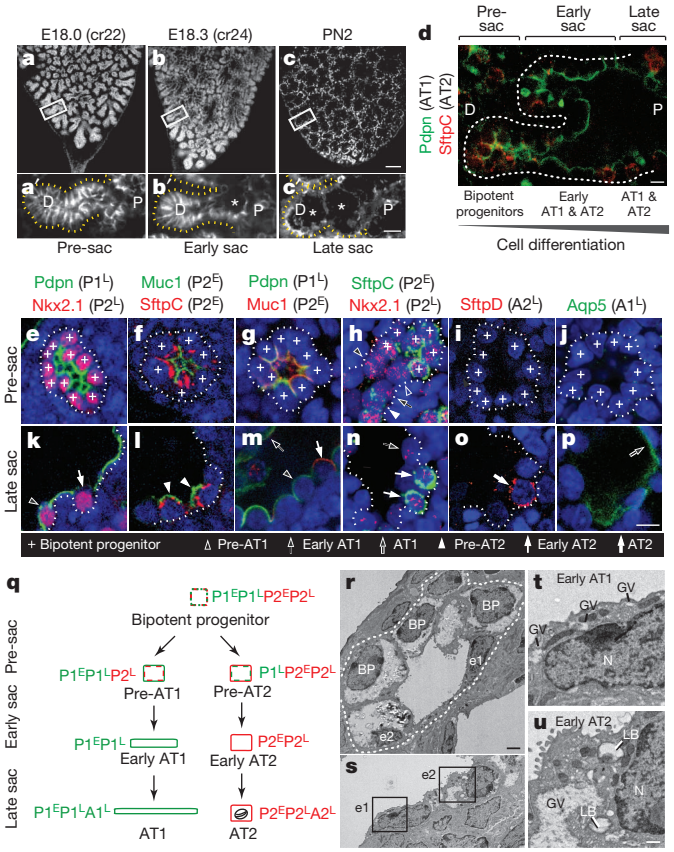


Figure 1 | Development of alveolar type 1 (AT1) and AT2 cells from bipotent progenitors. **a**–**c**, Mouse lung lobe tips stained for E-cadherin (E, embryonic day; cr, crown-rump length in mm; PN, postnatal day). **a'**–**c'**, Sacculation (sac; asterisks) proceeds proximally (P) to distally (D) along the airway. Scale bars, 100 μm (**a**–**c**), 20 μm (**a'**–**c'**). **d**, Staining for AT1 (Pdpn) and AT2 (SftpC) markers at E18.3 shows coexpression in pre-sacculation zone and restriction in late sacculation zone. Scale bar, 10 μm. **e**–**p**, Tips imaged in pre-sacculation (**e**–**j**) and late sacculation (**k**–**p**) zones identify six classes of marker expression profiles ($P1^E$,$P1^L$,$A1^L$,$P2^E$,$P2^L$,$A2^L$; Extended Data Table 1). Scale bar, 20 μm. **q**, Inferred differentiation pathways showing changes in AT1 (green) and AT2 (red) marker classes. Oval, lamellar body. **r**, **s**, Electron micrographs at E18.3 showing early (**r**) and late (**s**) sacculation zones. BP, bipotent progenitors; e1, early AT1 cells; e2, early AT2 cells. Scale bar, 2 μm. **t**, **u**, Boxed regions in S showing (e1) an early AT1 (squamous, glycogen vacuoles (GV) without lamellar bodies (LB)) and (e2) an early AT2 (cuboidal, GV, LB). N, nucleus; bar, 0.5 μm.

tracing of newly differentiated AT2 cells using a Cre recombinase knock-in (LysM–Cre) at the lysozyme M (*Lyz2*) locus, a late AT2 marker gene ($A2^L$), labelled many AT2 cells in the embryo but no marked AT1 cells were seen by 2 weeks postnatally (Extended Data Fig. 2c, e).

Little epithelial proliferation was detected during sacculation and for several weeks afterward (Extended Data Fig. 3), indicating that maturation of bipotent progenitors generates most or perhaps all AT1 and AT2 cells in development. Flow cytometry at early sacculation (E18.1) showed bipotent progenitors or pre-AT2 cells ($Muc1^+$ $Pdpn^+$) made up 8% of distal cells, whereas early or mature AT2 cells ($Muc1^+$ $Pdpn^-$) made up 5%. Bipotent progenitors seem to be fully exhausted by postnatal day 4, when sacculation has completed throughout the lung. In adult lungs few if any bipotent progenitors were found by flow cytometry (0.3% ± 0.2% $Muc1^+$ $Pdpn^+$ distal epithelial cells) or labelling with Shh–Cre–ER, suggesting a transition to other sources.

Alveolar renewal foci by rare activation of AT2 cells

Classical autoradiographic studies suggested alveolar epithelium is a slowly renewing population maintained by diffuse proliferation, presumably of AT2 cells[23]. We used *in vivo* lineage tracing to investigate the role of AT2 cells in maintenance, marking them in two complementary ways, with similar results. In one, a Cre–ERT2–rtTA knock-in at surfactant protein C (*SftpC*) locus (SftpC–Cre–ER) was used with a membrane-localized fluorescent reporter (mTmG) to pulse-label AT2 cells by tamoxifen induction at postnatal (PN) 18 days; mice were analysed 13 or 192 days later. Although SftpC is often used as a mature AT2 marker, the results above (Extended Data Table 1) show that as a class $P2^E$ marker it is also expressed by the bipotent progenitor. Hence, we also labelled AT2 cells using LysM–Cre described above, a class $A2^L$ marker that initiates expression only in mature AT2 cells (Extended Data Fig. 4a, b, d, m). Both lines robustly labelled AT2 cells throughout the lung, and at 2 months after marking (SftpC–Cre–ER) or postnatally (LysM–Cre) showed scattered AT1 cells labelled with the AT2 lineage tag (Extended Data Fig. 2d, f and Extended Data Fig. 5). AT2 cells did not significantly contribute to maintenance of bronchiolar lineages including Clara, ciliated and neuroendocrine cells (Extended Data Fig. 4c, e–m), and although rare cells near bronchoalveolar junctions that co-expressed SftpC and Clara cell marker CCSP/Scgb1a1 were also tagged ('bronchoalveolar stem cells' or BASCs[12]), they proliferated little if at all (Extended Data Fig. 4f–h, j–m).

To investigate the frequency and spatial localization of AT1 renewal by AT2 cells, we labelled lungs as above using LysM–Cre and analysed them at 1, 2, 4, 8 and 16 months of age. Less than 1% of AT1 cells expressed the AT2 lineage tag at 1 month, 3.9% at 4 months and 7.5% at 16 months (Fig. 2a–c). AT1 cell replacement occurred preferentially in alveoli abutting arterioles and in the lung periphery (Fig. 3a, b). Wherever new AT1 cells arose, the alveolus was replaced essentially in half or its entirety, indicating alveoli include just one or two AT1 cells. With ageing, adjacent alveolar units were also renewed, evidenced by slowly enlarging clusters of AT1 cells marked with the AT2 lineage tag. These patches of AT1 replacement were indistinguishable from surrounding areas, except for AT1 labelling with the AT2 lineage tag; similar results were obtained with SftpC–Cre–ER (Extended Data Fig. 5). We conclude that AT1 replacement by AT2 cells occurs intermittently and in a spatially patchy distribution in 'renewal foci' that slowly enlarge over time, with perivascular and peripheral regions serving as relative 'hot spots.'

Renewal foci derive from single, differentiated AT2 cells

Each renewal focus could derive from a single AT2 cell, in which case the cells would be clonally related, or from multiple AT2 cells. To distinguish these possibilities, we used a Confetti Cre-dependent reporter that stochastically expresses one of four fluorescent proteins in each cell that undergoes recombination. We analysed the membrane CFP lineage marker and found that foci were similar in size and progressive enlargement to those observed using the single-colour reporter, indicating that renewal foci derive from a single 'founder' AT2 cell. Because of fortuitously inefficient recombination of the Confetti reporter using LysM–Cre (~2% of AT2 cells expressed the membrane-targeted CFP lineage tag at 16 months), we were able to distinguish marked AT2 cells within renewal foci. Most contained one or two AT2 cells and multiple AT1 cell progeny colonizing up to six contiguous alveoli (Fig. 2d–f). We also observed small clonal foci comprising only AT2 cells (Extended Data Fig. 6b, c), indicating they are capable of dividing without forming AT1 cells (self-duplication). Founder AT2 cells expressed mature AT2 markers including LAMP-1, a lysosome-associated protein, indicating they continue to produce surfactant (Extended Data Fig. 6a), and they did not express Clara, ciliated and neuroendocrine markers. Their only distinguishing feature was their nucleus, which was often slightly larger and stained more intensely for Nkx2.1 than nearby AT2 cells (see also Fig. 2d, e). We conclude that each renewal focus derives from a single, self-renewing AT2 cell that can generate multiple AT1 and/or AT2 cells, is long-lived, and remains closely associated with its progeny, as evidenced by its persistence in large foci at advanced ages (Fig. 2e). Because founder cells appear to be mature (LysM–Cre lineage positive), functional (SftpC–Cre–ER lineage and LAMP-1 positive) AT2 cells that share the essential properties of conventional stem cells (multipotent,

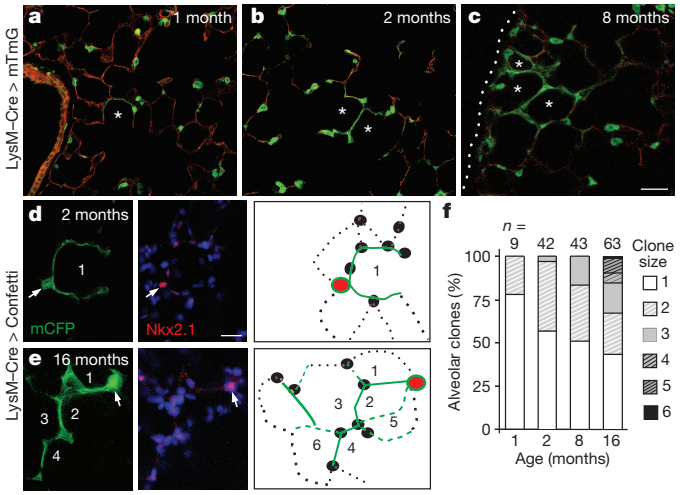


Figure 2 | Mature AT2 cells renew AT1 cells in clonal foci. **a**–**c**, AT2 lineage-labelled AT1 cell foci (green) enlarge with ageing, incorporating adjacent alveoli (asterisks). Dotted line, mesothelium. Scale bar, 150 μm. **d**, **e**, AT2 cells were sparsely marked with Confetti reporter, and mCFP-labelled foci (green) are shown co-stained for AT2 marker Nkx2.1 (red) at age 2 (**d**) and 16 (**e**) months. Note 'founder' AT2 cell (arrow) and its labelled AT1 progeny (green). Numbers, incorporated alveoli. Entire clone schematized in right panel. Red, founder AT2; green lines, AT1 daughters visible (solid) or outside (dashed) focal plane; black, unlabelled AT1 (dotted lines) and AT2 (ovals) cells. Scale bar, 20 μm (**d**, **e**). **f**, Clone size increases with ageing ($P = 0.03$, Kruskal–Wallis test). *n*, clones scored; clone size, number of incorporated alveoli.

self-renewing, persist for life), they are 'bi-functional' stem cells, executing both differentiated and regenerative functions.

AT2 cell activation by acute AT1 cell injury

Elevated oxygen tension is toxic to AT1 cells, but not to AT2 cells[24]. Exposure of 2 month old mice carrying the AT2 lineage tag to 88% oxygen for 120 h tripled the number of renewed AT1 cells (Fig. 3c–e). This finding shows that AT2 cells become activated following hyperoxic injury, and supports the idea that although normally only a rare subset of AT2 cells executes a stem cell function, others can be recruited to repair alveolar damage. Whether every AT2 cell can be activated this way could not be determined because more severe hyperoxia was lethal.

Oncogenic Kras selectively activates AT2 self-renewal

Adenocarcinoma, the major form of lung cancer, is associated with activating mutations in *Kras* or *Egfr*[25]. Typically located in peripheral lung regions, nearly all tumour cells express SftpC, leading to long-standing speculation that they originate from AT2 cells or their progenitors. However, the identity of the tumour-initiating cell(s) remains controversial[12,26–28]. To test the effect of oncogenic *Kras(G12D)* on mature AT2 cells, a conditional $Kras^{LSL\text{-}G12D}$ allele ($Kras^{LSL\text{-}G12D}$ is a knock-in at the *Kras* locus in which the wild-type *Kras* coding sequence is replaced by a lox-STOP-lox-*Kras(G12D)*) was activated using LysM-Cre along with the mTmG Cre-dependent reporter. Tumour nodules grew rapidly throughout the lungs (Fig. 4a–d), with dense replacement of virtually the entire alveolar region by 1 month after induction and death shortly thereafter (Fig. 4e). When lungs were examined in the first few days following induction, we found with a Rainbow multi-colour reporter that nearly every epithelial cell expressing the AT2 lineage tag proliferated, demonstrating highly efficient AT2 transformation by *Kras(G12D)* (Fig. 4f–i). The biggest tumours were found in peripheral and perivascular regions, sites where physiological AT1 renewal by AT2 cells was commonly observed (compare Figs 3a, b and 4o, p). Similar results were obtained using SftpC–Cre–ER to activate *Kras(G12D)* in adult mice (Extended Data Fig. 7a, c). By contrast, when we used CCSP–Cre–ER, most Clara cells were unaffected or divided minimally, whereas at bronchoalveolar junctions some formed small clonal adenomas (Extended Data Fig. 7b). We also used ubiquitously expressed ROSA–Cre–ER to activate the *Kras(G12D)* allele at random, resulting in many singlets and minimally affected cells throughout the lung, even 18 days after induction (Fig. 4j). The transformed AT2 cells comprising the adenomas continued to express AT2 markers (Nkx2.1, SftpD) and did not turn on a Clara marker (CCSP) or, with rare exceptions, AT1 markers (Pdpn, LEL) (Extended Data Fig. 8). Thus, oncogenic *Kras(G12D)* seems to selectively and permanently induce AT2 self-renewal, without deprogramming the cells to the bipotent progenitor or causing reprogramming to AT1 or Clara cell fates.

By examining lineage-tagged *Kras(G12D)* mutant lungs at progressive stages, we could infer the cellular mechanism of adenoma formation (Fig. 4k–m). Proliferation of the activated AT2 cell generates daughter cells that spread laterally yet maintain a monolayer ('lepidic' expansion), the first histological sign of the tumour (Fig. 4l). Later, cells heap up and form a nodule ('hilical' expansion) that obliterates the lumen and begins compressing and invading adjacent alveoli (Fig. 4m). Infrequent

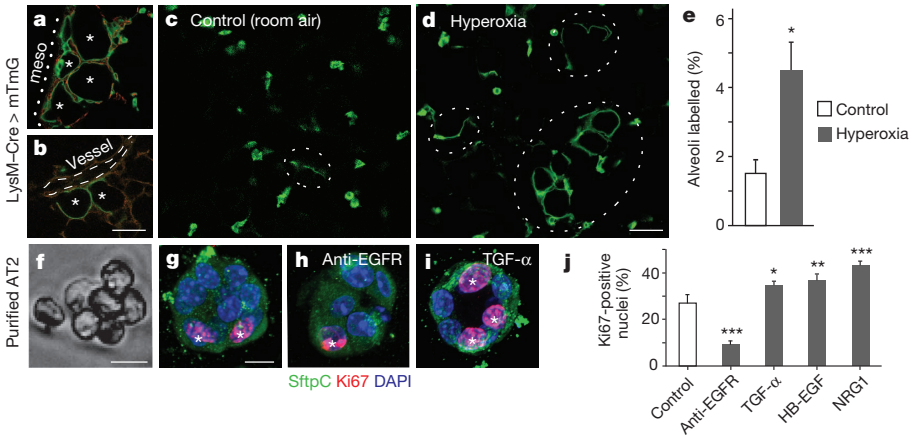


Figure 3 | Activation of AT2 stem cell function *in vivo* and proliferation *in vitro*. **a**, **b**, Renewal foci (AT2 lineage label, green; other cells, red) commonly involve alveoli (asterisks) in peripheral (**a**, mesothelium, dots) and perivascular (**b**, dashes) domains. Scale bar, 50 μm. **c**–**e**, Alveolar regions under room air (**c**) or after hyperoxia (88% O_2, 5 days) to injure AT1 cells (**d**). Note increased foci (dashed ovals) after injury. Scale bar, 50 μm. Quantification (**e**) shows increased alveolar surface ($n = 2$; mean ± s.e.m.) from AT2-lineage labelled cells after hyperoxia. $P < 0.05$ (Mann–Whitney U test). **f**, **g**, Freshly isolated AT2 cells (**f**, phase contrast) cultured 4 days in Matrigel (**g**) proliferate, shown by Ki67 staining (red, asterisks), but maintain AT2 marker expression (SftpC, green). **h**–**j**, Images (**h**, **i**) and quantification (500 cells per biological replicate, $n = 4, 3, 3, 3, 4$) (**j**) of proliferation with EGFR blocking antibody (anti-EGFR, 2.5 μg ml^{-1}) and EGF ligands indicated (4 μM). Scale bar, 10 μm (**f**–**i**). Mean ± s.e.m.; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$ (Tukey's multiple comparisons test).

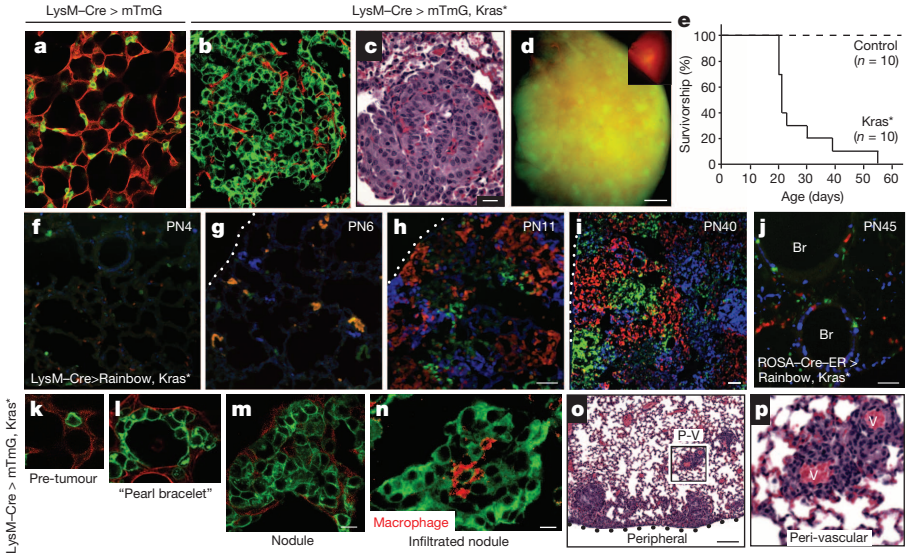


Figure 4 | Transformation of mature AT2 cells by activated Kras. **a**–**c**, LysM-Cre > mTmG control (**a**) and LysM-Cre > mTmG,Kras$^{LSL-G12D/+}$ (abbreviated Kras*) (**b**, **c**) lungs at age 7 weeks, showing proliferated AT2 cells (**b**, green cells) compressing surrounding cells (red) and forming adenomas (**c**, haematoxylin and eosin stain). Scale bar, 20 μm. **d**, Lung lobe as in **b** showing widespread infiltration by tumour (green). Inset, control lobe. Scale bar, 1 mm. **e**, Survival curves. **f**–**i**, LysM-Cre > Rainbow,Kras* lungs at indicated ages (PN, postnatal day). Note rapid clonal (single colour) expansion of labelled AT2 cells with minimal cell mixing. Scale bar, 50 μm (**f**–**h**), 100 μm (**i**). Dotted line, mesothelium. **j**, PN45 ROSA–Cre–ERT2> Rainbow,Kras* lung 18 days after induction. Note minimal expansion. Br, bronchi; scale bar, 50 μm. **k**–**n**, A recombined (green) AT2 cell (**k**) generates progeny that spread laterally, giving 'pearl bracelet' appearance (**l**). Cells later 'heap up' into nodules (**m**), some infiltrated by macrophages (**n**; red, anti-F4/80). Scale bar, 10 μm. **o**, **p**, Haematoxylin and eosin stain shows robust tumour growth peripherally and perivascularly (P-V). Boxed area (**o**, close-up in **p**) shows proliferated cells around blood vessels (v). Scale bar, 200 μm.

clones developed into large, well-formed papillary structures, although most formed simple adenomas[29]. At advanced stages, a notable pattern of large, closely packed, single-colour tumours was evident, showing minimal cellular exchange between neighbouring tumour foci (Fig. 4i). Some advanced tumours became infiltrated by macrophage 'nests' (Fig. 4n), a poor prognostic sign[30].

EGFR activity controls AT2 self-renewal in culture

We profiled the transcriptome of bipotent progenitors and AT2 cells. Both express *Egfr*, two other EGFR family members (*Erbb2* and *Erbb3*) (Supplementary Table 2), and receptors for many other signals (Supplementary Tables 2 and 3). Purified AT2 cells (Fig. 3f) retain a robust ability to both proliferate (>25% Ki67-positive, Fig. 3g, j) and to differentiate into AT1-like cells (>95% thin, flat morphology and aquaporin 5-positive, $n = 240$ cells) (Extended Data Fig. 6d). Addition of purified EGF ligands (transforming growth factor α (TGF-α), heparin-binding EGF (HB-EGF), NRG1) stimulated AT2 proliferation, whereas EGFR-blocking antibody inhibited proliferation (Fig. 3h–j). None of these treatments induced differentiation into AT1 cells. EGF signalling is therefore a critical and selective regulator of AT2 proliferation, at least under these conditions.

Discussion

Our characterization of alveolar progenitor and stem cells throughout the lifespan supports a model in which AT1 and AT2 cells arise independently during development from a bipotent progenitor (Fig. 5a). Several weeks after birth, when alveolar development is complete, there is a switch and mature AT2 cells become a renewable source of AT1 and AT2 cells. But these postnatal renewal events are rare and occur in monoclonal foci ('alveolar renewal focus', Fig. 5b) that slowly expand over months or years. Only a small fraction (~1%) of mature AT2 cells normally express this stem cell function and they divide intermittently (~40-day doubling time) and supply only local domains, giving an overall renewal rate of just 7% of alveoli per year. This stem cell function is more broadly induced by AT1 injury, indicating that other AT2 cells may have similar potential ('alveolar repair focus', Fig. 5b) that can be activated by dying AT1 cells. Classical pneumonectomy experiments support the idea that most AT2 cells possess latent regenerative capacity[31], and AT2 ablation induces self-duplication of surviving AT2 cells[9] ('AT2 replacement focus'). Our data do not exclude alternative sources of new alveolar cells, especially following severe injuries[14] that deplete AT2 cells regionally.

If many or all AT2 cells can serve as stem cells, why does only a minority execute this function for maintenance, producing large monoclonal foci? Perhaps alveolar turnover is coupled with a stem cell hierarchy whereby an initially activated AT2 cell suppresses nearby AT2 cells and becomes the dominant stem cell. It is also unclear why renewal foci slowly enlarge over time, because there is no obvious recurrent injury. Perhaps foci are programmed anatomical domains of alveolar renewal, with new cells moving out from a specialized niche akin to intestinal crypts, albeit with much slower turnover. Whatever the explanation, there are clearly regional influences as both renewal and tumour foci are preferentially located in perivascular and peripheral lung domains, possible sources of stem cell signals[32].

EGFR signalling selectively stimulates proliferation of AT2 cells *in vitro*, and oncogenic *Kras(G12D)* permanently and selectively activates proliferation *in vivo*, efficiently transforming AT2 cells into rapidly growing monoclonal tumours ('lung tumour focus', Fig. 5b). By virtue of their large numbers, class susceptibility and robust adenomatous response to *Kras(G12D)*, AT2 cells may constitute the major cell type responsible for human lung adenocarcinoma, making them among the most dangerous cells in the body.

We propose that the oncogenic potential of AT2 cells is a direct consequence of their stem cell function: AT2 cells are poised to function as alveolar stem cells and EGFR/KRAS signalling regulates the self-renewal part of the stem cell program (and the related process of self-duplication); another signalling pathway (Supplementary Table 2) must control AT2 reprogramming to AT1 fate (Fig. 5a). This model predicts that dying AT1 cells secrete an EGF that initiates self-renewal, plus another signal for fate reprogramming. Dying AT2 cells presumably produce only the former. It will be important to identify these signals and the events they control. This could suggest new strategies for early

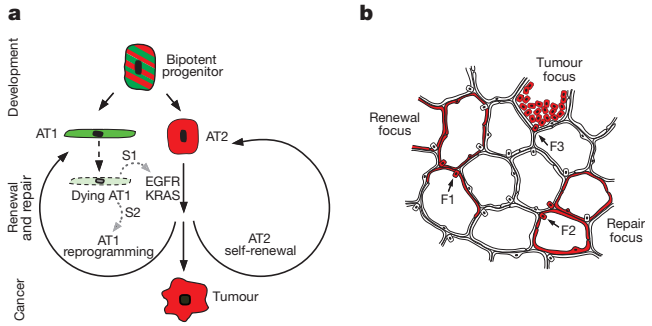


Figure 5 | Model of alveolar progenitors and stem cells in development, maintenance, and cancer. **a**, Bipotent progenitors expressing some AT1 (green) and AT2 (red) markers differentiate into AT1 or AT2 cells. Mature AT2 cells function as stem cells intermittently activated for alveolar renewal and repair. Dying AT1 cells are proposed to produce a signal (S1) transduced by EGFR-KRAS that activates division of a nearby AT2 cell (self-renewal); another signal (S2) reprograms a daughter into an AT1 cell. Activating mutations of *Egfr* or *Kras* in AT2 cells drive constitutive self-duplication, forming tumour of AT2-like cells. **b**, Rare AT2 cells function as stem cells, giving rise to clonal renewal foci (red, left) that slowly enlarge, with persistence of founder AT2 cell (F1). With injury, additional AT2 cells (F2) are recruited to generate repair foci (red, right). Activating *Kras* mutation in AT2 cells (F3) initiates tumour focus (red, top).

detection and treatment of tumours and for replenishing diseased alveoli.

METHODS SUMMARY

Cre recombinase was used to activate Rosa26-reporters[33] and a $Kras^{LSL\text{-}G12D}$ knock-in[34]. For immunostaining, lungs were agarose-inflated, then fixed in para-formaldehyde (PFA) and cryo-embedded or vibratome-sliced and fixed in methanol:dimethylsulphoxide or PFA. Secondary antibodies were Alexa Fluor-conjugated, or horseradish peroxidase-conjugated with tyramide amplification. Specimens were imaged by confocal (slices) or epifluorescence (cryosections) microscopy. For cell purification, lungs were dissociated then sorted by fluorescent markers or surface antigens using FACS or MACS. AT2 cells were cultured on Matrigel or on glass with serum. Expression profiling used Affymetrix platform.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 December 2012; accepted 3 December 2013.
Published online 5 February 2014.

1. Bertalanffy, F. D. & Leblond, C. P. Structure of respiratory tissue. *Lancet* **266,** 1365–1368 (1955).
2. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA Cancer J. Clin.* **63,** 11–30 (2013).
3. Rock, J. R. & Hogan, B. L. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu. Rev. Cell Dev. Biol.* **27,** 493–512 (2011).
4. Chapman, H. A. *et al.* Integrin α6β4 identifies an adult distal lung epithelial population with regenerative potential in mice. *J. Clin. Invest.* **121,** 2855–2862 (2011).
5. Adamson, I. Y. & Bowden, D. H. Derivation of type 1 epithelium from type 2 cells in the developing rat lung. *Lab. Invest.* **32,** 736–745 (1975).
6. Spencer, H. & Shorter, R. G. Cell turnover in pulmonary tissues. *Nature* **194,** 880 (1962).
7. Evans, M. J. & Bils, R. F. Identification of cells labeled with tritiated thymidine in the pulmonary alveolar walls of the mouse. *Am. Rev. Respir. Dis.* **100,** 372–378 (1969).
8. Rock, J. R. *et al.* Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc. Natl Acad. Sci. USA* **108,** E1475–E1483 (2011).
9. Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* **123,** 3025–3036 (2013).
10. Evans, M. J., Cabral, L. J., Stephens, R. J. & Freeman, G. Renewal of alveolar epithelium in the rat following exposure to NO_2. *Am. J. Pathol.* **70,** 175–198 (1973).
11. Adamson, I. Y. & Bowden, D. H. The type 2 cell as progenitor of alveolar epithelial regeneration. A cytodynamic study in mice after exposure to oxygen. *Lab. Invest.* **30,** 35–42 (1974).
12. Kim, C. F. *et al.* Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121,** 823–835 (2005).
13. McQualter, J. L., Yuen, K., Williams, B. & Bertoncello, I. Evidence of an epithelial stem/progenitor cell hierarchy in the adult mouse lung. *Proc. Natl Acad. Sci. USA* **107,** 1414–1419 (2010).
14. Kumar, P. A. *et al.* Distal airway stem cells yield alveoli *in vitro* and during lung regeneration following H1N1 influenza infection. *Cell* **147,** 525–538 (2011).
15. Kajstura, J. *et al.* Evidence for human lung stem cells. *N. Engl. J. Med.* **364,** 1795–1806 (2011).
16. Rawlins, E. L. *et al.* The role of $Scgb1a1^+$ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* **4,** 525–534 (2009).
17. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479,** 189–193 (2011).
18. Burri, P. H. & Moschopulos, M. Structural analysis of fetal rat lung development. *Anat. Rec.* **234,** 399–418 (1992).
19. Buckingham, S., McNary, W. F., Jr & Sommers, S. C. Pulmonary alveolar cell inclusions: their development in the rat. *Science* **145,** 1192–1193 (1964).
20. Williams, M. C. & Dobbs, L. G. Expression of cell-specific markers for alveolar epithelium in fetal rat lung. *Am. J. Respir. Cell Mol. Biol.* **2,** 533–542 (1990).
21. Hughes, G. M. Ultrastructure of the lung of *Neoceratodus* and *Lepidosiren* in relation to the lung of other vertebrates. *Folia Morphol. (Praha)* **21,** 155–161 (1973).
22. Miller, L. A., Wert, S. E. & Whitsett, J. A. Immunolocalization of sonic hedgehog (Shh) in developing mouse lung. *J. Histochem. Cytochem.* **49,** 1593–1603 (2001).
23. Messier, B. & Leblond, C. P. Cell proliferation and migration as revealed by radioautography after injection of thymidine-H^3 into male rats and mice. *Am. J. Anat.* **106,** 247–285 (1960).
24. Bowden, D. H., Adamson, I. Y. & Wyatt, J. P. Reaction of the lung cells to a high concentration of oxygen. *Arch. Pathol.* **86,** 671–675 (1968).
25. Herbst, R. S., Heymach, J. V. & Lippman, S. M. Lung cancer. *N. Engl. J. Med.* **359,** 1367–1380 (2008).
26. Sutherland, K. D. & Berns, A. Cell of origin of lung cancer. *Mol. Oncol.* **4,** 397–403 (2010).
27. Xu, X. *et al.* Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. *Proc. Natl Acad. Sci. USA* **109,** 4910–4915 (2012).
28. Lin, C. *et al.* Alveolar type II cells possess the capability of initiating lung tumor development. *PLoS ONE* **7,** e53817 (2012).
29. Nikitin, A. Y. *et al.* Classification of proliferative pulmonary lesions of the mouse: recommendations of the mouse models of human cancers consortium. *Cancer Res.* **64,** 2307–2316 (2004).
30. Takanami, I., Takeuchi, K. & Kodaira, S. Tumor-associated macrophage infiltration in pulmonary adenocarcinoma: association with angiogenesis and poor prognosis. *Oncology* **57,** 138–142 (1999).
31. Brody, J. S., Burki, R. & Kaplan, N. Deoxyribonucleic acid synthesis in lung cells during compensatory lung growth after pneumonectomy. *Am. Rev. Respir. Dis.* **117,** 307–316 (1978).
32. Ding, B. S. *et al.* Endothelial-derived angiocrine signals induce and sustain regenerative lung alveolarization. *Cell* **147,** 539–553 (2011).
33. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* **148,** 33–45 (2012).
34. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15,** 3243–3248 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Andalon for technical assistance; H. Chapman (SftpC–Cre–ER-rtTA), B. Hogan (CCSP–Cre–ER), H. Ueno and I. Weissman (Rainbow), H. Clevers (Confetti), L. Luo (mTmG), and J. Sage ($Kras^{LSL\text{-}G12D}$) for strains; B. Stripp for goat anti-CCSP antibody; F. H. Espinoza for annotated gene lists; R. Metzger, H. Chapman, and members of the Krasnow laboratory for discussions and comments on the manuscript; and Maria Petersen for help preparing figures and the manuscript.

Author Contributions T.J.D. conducted the experiments except the gene expression profiling and AT2 cell cultures, which were done by D.G.B.; T.J.D., D.G.B. and M.A.K. conceived the experiments, analysed the data and wrote the manuscript. This work was supported by a Parker B. Francis Foundation Fellowship and NIH 5KO8HL084095 Award (T.J.D.), NIH T32HD007249 (D.G.B.), and an NHLBI U01HL099995 Progenitor Cell Biology Consortium grant (M.A.K.). M.A.K. is an investigator of the Howard Hughes Medical Institute.

Author Information Microarray datasets were deposited at Gene Expression Omnibus (accession code GSE49346) and GEXC (https://gexc.stanford.edu/population/detail/998 and https://gexc.stanford.edu/population/detail/999). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.K. (krasnow@stanford.edu) or T.J.D. (tdesai@stanford.edu).

ARTICLE

doi:10.1038/nature13124

C9orf72 nucleotide repeat structures initiate molecular cascades of disease

Aaron R. Haeusler[1,2], Christopher J. Donnelly[3,4], Goran Periz[1,2], Eric A. J. Simko[1,2], Patrick G. Shaw[5], Min-Sik Kim[5], Nicholas J. Maragakis[3], Juan C. Troncoso[6], Akhilesh Pandey[5], Rita Sattler[3,4], Jeffrey D. Rothstein[2,3,4] & Jiou Wang[1,2]

A hexanucleotide repeat expansion (HRE), $(GGGGCC)_n$, in *C9orf72* is the most common genetic cause of the neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Here we identify a molecular mechanism by which structural polymorphism of the HRE leads to ALS/FTD pathology and defects. The HRE forms DNA and RNA G-quadruplexes with distinct structures and promotes RNA•DNA hybrids (R-loops). The structural polymorphism causes a repeat-length-dependent accumulation of transcripts aborted in the HRE region. These transcribed repeats bind to ribonucleoproteins in a conformation-dependent manner. Specifically, nucleolin, an essential nucleolar protein, preferentially binds the HRE G-quadruplex, and patient cells show evidence of nucleolar stress. Our results demonstrate that distinct *C9orf72* HRE structural polymorphism at both DNA and RNA levels initiates molecular cascades leading to ALS/FTD pathologies, and provide the basis for a mechanistic model for repeat-associated neurodegenerative diseases.

Nucleotide repeat elements, including microsatellites and short tandem repeats, are common in eukaryotic genomes[1]. Recently, a hexanucleotide repeat expansion (HRE), $(GGGGCC)_n$, in a noncoding region of *C9orf72* was linked to the neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[2,3]. ALS is characterized by a loss of motor neurons, with 90% of ALS cases being sporadic and the other ~10% having a family history[4]; the *C9orf72* HRE represents the most common genetic cause of both familial and sporadic ALS[5]. FTD is characterized by degeneration of the frontal and temporal lobes of the brain and is the second most common type of dementia in people younger than 65[6]. Again, the *C9orf72* HRE is one of the most common genetic causes of FTD[6]. Increasing evidence suggests that ALS and FTD are two related diseases in a continuous clinical spectrum[7], and there is a possibility that the *C9orf72* HRE also contributes to Alzheimer's and Huntington's diseases[8–11].

Normal human *C9orf72* alleles have 2 to 25 intronic GGGGCC repeats, with the majority having fewer than eight repeats and more than half having two repeats[12]. The expanded repeats associated with ALS/FTD are thought to have variable lengths, ranging from tens to thousands of hexanucleotide repeats[2,3], but correlations between the repeat lengths and clinical onset or progression have yet to be established. Although a molecular understanding of *C9orf72* HRE pathological phenotypes has begun to emerge, the mechanisms by which the GGGGCC repeat expansion causes ALS/FTD pathology are unknown and our understanding of nucleotide repeats in the context of human disease is still in its infancy[13–23].

Here we report that the *C9orf72* HRE DNA/RNA sequence is structurally polymorphic; it can fold into stable G-quadruplex secondary structures and form transcriptionally induced RNA•DNA hybrids known as R-loops. The DNA of the *C9orf72* HRE forms both antiparallel- and parallel-stranded G-quadruplexes, whereas the RNA adopts only parallel-stranded G-quadruplex conformations. These structural features of the *C9orf72* HRE lead to truncated HRE-containing abortive transcripts. We identified ribonucleoproteins bound to the repeat RNA in an RNA conformation-dependent manner. Nucleolin (NCL), an essential nucleolar protein that binds specifically to *C9orf72* HRE G-quadruplexes, is mislocalized in patient cells carrying the mutation. Accordingly, nucleolar function is impaired in patient cells. Furthermore, this nucleolar pathology can be recapitulated by introducing *C9orf72* HRE-containing abortive transcripts into wild-type cells. These results point to the structural conformations of both the DNA and RNA hexanucleotide repeats as fundamental determinants of the pathogenic mechanisms of *C9orf72* HRE-linked ALS/FTD.

G-quadruplexes formed by HRE DNA/RNA

To understand how expansion of the GGGGCC repeat in the *C9orf72* gene impedes cellular functions and leads to the associated diseases, we began by examining the repeat sequence for unique structural characteristics. The *C9orf72* HRE GGGGCC DNA repeat sequence has properties that would allow it to form G-quadruplexes, which are stacks of planar tetramers consisting of four guanines connected by Hoogsteen hydrogen bonds[24]. To look for such quadruplexes we first examined the secondary structures formed by the GGGGCC hexanucleotide repeats using circular dichroism (CD) absorptivity. The $(GGGGCC)_4$ DNA shows a characteristic spectrum for antiparallel G-quadruplexes in 100 mM KCl[25], with a maximum absorbance at 295 nm and a minimum at 260 nm (Fig. 1a). This conformational change at physiologically relevant levels of KCl (100 mM) is also apparent in the decreased mobility of the DNA in a native polyacrylamide gel electrophoresis (PAGE) shift assay (Fig. 1b).

To provide further conformational insight into the DNA G-quadruplex formed by $(GGGGCC)_4$, we used dimethyl sulphate (DMS) footprinting following classical Maxam and Gilbert sequencing methods[26]. On the sequencing gel (Fig. 1c), the band pattern resulting from guanine depurination is markedly reduced in the presence of KCl, indicating that the guanine N7 positions are protected from DMS methylation and subsequent base cleavage as a result of hydrogen bonding in a G-quartet conformation. These results provide evidence that the $(GGGGCC)_4$ DNA forms a four-stack antiparallel G-quadruplex motif, with the guanines on the exterior of the stacks being more accessible to chemical modifications than the guanines buried in the interior (Fig. 1d). Through

[1]Department of Biochemistry and Molecular Biology, Johns Hopkins University Baltimore, Maryland 21205, USA. [2]Department of Neuroscience, Johns Hopkins University Baltimore, Maryland 21205, USA. [3]Department of Neurology, Johns Hopkins University Baltimore, Maryland 21205, USA. [4]The Brain Science Institute, Johns Hopkins University Baltimore, Maryland 21205, USA. [5]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University Baltimore, Maryland 21205, USA. [6]Department of Pathology, Johns Hopkins University Baltimore, Maryland, 21205, USA.

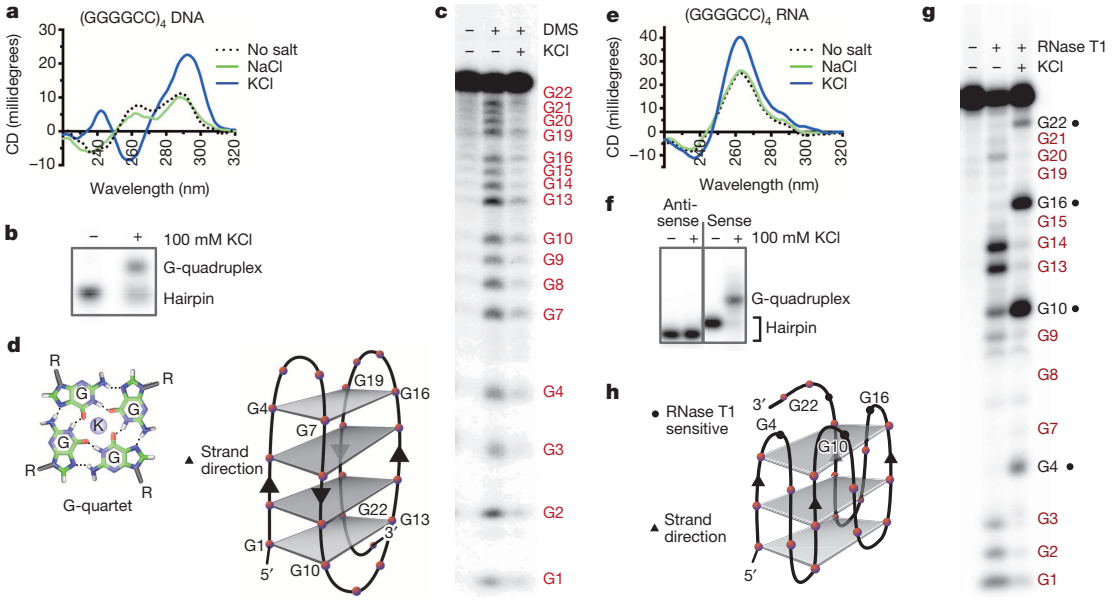


Figure 1 | DNA and RNA of the *C9orf72* HRE form G-quadruplexes. **a**, The CD absorptivity shows characteristic K^+-dependent spectra for antiparallel DNA G-quadruplexes with $(GGGGCC)_4$. **b**, The presence of K^+ during annealing induces a conformational change that decreases the mobility of $(GGGGCC)_4$ DNA in a gel mobility shift assay. **c**, A DMS footprinting assay on $(GGGGCC)_4$ DNA shows protection of the N7 positions on all the guanines when a G-quadruplex is formed. **d**, The proposed topology for the antiparallel DNA G-quadruplex formed by $(GGGGCC)_4$. Each grey plane represents a G-quartet, as shown in the lower corner. Four separate G-quartets are stacked 5′ to 3′ with two cytosines forming each loop region. **e**, The CD spectra identified the formation of a parallel G-quadruplex for the RNA $(GGGGCC)_4$ in the presence of 100 mM KCl. **f**, $(GGGGCC)_4$ RNA demonstrates a slower mobility in the presence of KCl compared to $(CCCCGG)_4$ RNA. **g**, The RNase T1 protection assay identifies single-stranded guanine residues (denoted in black) not involved in the formation of the G-quadruplex. **h**, The proposed parallel G-quadruplex topology formed by the RNA of the *C9orf72* HRE. Three stacks form the parallel G-quadruplex, with the RNase T1-sensitive guanines shown as black dots.

a series of spectroscopic characterizations, we observed that the GGGGCC sequence can adopt intermolecular, intramolecular, and also parallel G-quadruplex structures when the number of repeats is varied (Extended Data Fig. 1a–f and Supplementary Results). However, the antiparallel G-quadruplex seems to be the most dominant and the most stable conformation at physiologically relevant conditions.

We next investigated the structural conformation and stability of the *C9orf72* RNA HRE sequences. Through a series of experiments measuring CD absorptivity, the $(GGGGCC)_4$ RNA shows a signature spectrum for parallel G-quadruplexes[27], with a minimum absorbance at 240 nm and a maximum at 260 nm (Fig. 1e), which is consistent with recent results for similar RNA HRE sequences[15,16]. When the $(GGGGCC)_4$ RNA was examined by native PAGE, its mobility is decreased in the presence of 100 mM KCl, in agreement with a K^+-dependent formation of G-quadruplexes (Fig. 1f).

To provide further conformational insight, we performed an RNase protection assay. The RNA digestion pattern for endonuclease RNase T1, which cleaves single-stranded RNAs at the 3′ end of guanine residues, shows clear structural differences in the presence and absence of KCl. Without KCl, the $(GGGGCC)_4$ RNA forms a secondary structure consistent with single-stranded bulges and hairpin regions as indicated by the RNase protection pattern (Fig. 1g). In contrast, in the presence of 100 mM KCl the RNA adopts a G-quadruplex structure, in which every fourth guanine is highly sensitive to single-stranded cleavage. This provides evidence for a topology of a three-stacked parallel-stranded G-quadruplex with a guanine and two cytosines in the single-stranded loop region (Fig. 1h). Together, these data indicate that the *C9orf72* HRE RNA preferentially adopts a parallel G-quadruplex topology and is stable at physiological KCl concentrations, pH and temperature (Extended Data Fig. 1g).

Impeded transcription and abortive RNAs

To understand the functional consequence of structural polymorphism at the *C9orf72* HRE, we examined transcription of the repeat region. First, we established an *in vitro* transcription assay by creating GGGGCC hexanucleotide repeats of varying lengths (3–70 repeats) *de novo* via repeat-primed PCR, and placing the repeats downstream of a T7 promoter as transcription templates. When *in vitro* transcriptional products are resolved by denaturing PAGE periodic abortive transcripts are observed, with longer repeats producing increasing amounts of truncated transcripts that correspond to the template hexanucleotide repeats (Fig. 2a). Importantly, the increase in the abortive transcripts causes a concomitant decrease in full-length transcripts. To quantify this observation, we compared the levels of full-length transcripts, which contain regions downstream (3′) of the repeats, with the total levels of all transcripts, which contain regions upstream (5′) of the repeats (Fig. 2a). The plot of this 3′/5′ ratio shows that there is a transition to more abortive transcripts relative to full-length products with increasing repeats (Fig. 2b). Furthermore, we verified that the repeat templates were not extensively modified by depurination of the DNA sequence, excluding this potential modification as the cause of the observed abortive transcription (Extended Data Fig. 2). Therefore, these *in vitro* results demonstrate that the *C9orf72* HRE causes RNA polymerase processivity to be impaired in the repeat region, leading to an accumulation of repeat-containing abortive transcripts and loss of full-length transcripts.

Next, we investigated the molecular mechanisms underlying abortive RNA transcript production. We examined the structural features unique to the *C9orf72* HRE region on the plasmid DNA and found that, consistent with data on the DNA oligonucleotides in Fig. 1, the plasmid also forms stable G-quadruplexes in the presence of KCl, which was directly assessed by the preferential binding of the G-quadruplex-specific BG4 nanobody[28] to the DNA G-quadruplex structure (Extended Data Fig. 3a, b). The formation of these G-quadruplexes leads to transcripts being aborted earlier and to a further decrease in full-length transcripts in the *in vitro* transcription assay; these repeat-containing abortive transcripts accumulate over time (Extended Data Fig. 3c, d). These results indicate that the formation of G-quadruplexes impairs polymerase processivity within the *C9orf72* HRE region.

Next, we examined the nascent RNA transcript to look for possible mechanistic contributions to abortive transcription. First, using a

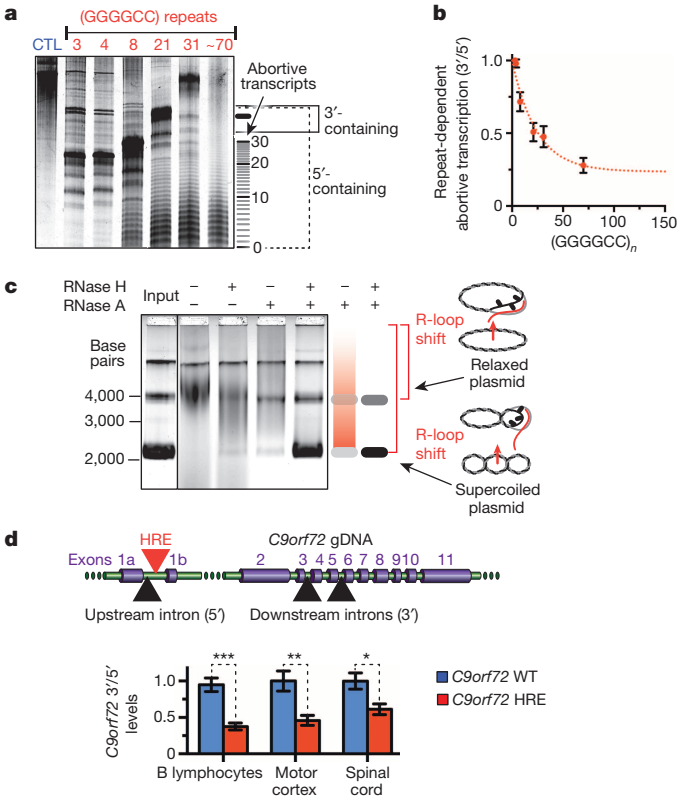


Figure 2 | Abortive transcription in the *C9orf72* HRE. **a**, Increasing lengths of GGGGCC repeats cause accumulation of abortive transcripts in a length-dependent manner *in vitro*. The transcriptional products were separated on a denaturing gel with a 500-nucleotide ssRNA control (CTL). **b**, Transcript levels shown in **a** were densitometrically quantified and then plotted as the ratio of full-length transcripts that contain regions 3′ of the repeat to all transcripts that contain 5′ regions. The curve was fit to a single exponential. Data are means ± s.d. $n = 4$. **c**, The *C9orf72* HRE induces the formation of R-loops on *C9orf72* HRE-containing plasmids with $(GGGGCC)_{\sim 70}$. Treatment of the *in vitro* transcription products with RNase A and H digests the RNA still hybridized with relaxed or supercoiled plasmid and reduces the smearing that was caused by the size heterogeneity of RNA•DNA hybrids. Genomic DNA (top band) serves as an internal loading control. **d**, Patients carrying the *C9orf72* HRE have reduced pre-mRNA 3′/5′ ratios relative to *C9orf72* wild type (WT), consistent with the HRE-induced abortive transcription reducing full-length transcript levels. Data are means ± s.e.m. $n = 5/6$ (B lymphocytes), $n = 12/10$ (motor cortex), $n = 8/5$ (spinal cord) for *C9orf72* WT/HRE samples, respectively. $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$.

G-quadruplex•haemin complex-catalysed peroxidase-like colorimetric assay[29], we demonstrated that the repeat-containing transcripts also form RNA G-quadruplexes, but the formation of these structures on nascent RNA during transcription has negligible effects on the decrease in polymerase processivity (Extended Data Fig. 4a, b and Supplementary Results). However, we noted that treatment with RNase H, which specifically digests RNA in an RNA•DNA hybrid, decreases the accumulation of abortive transcripts and increases longer transcripts during the assay (Extended Data Fig. 4c). To test whether nascent RNA•DNA template hybrids (R-loops) form on the GGGGCC repeats, we performed the *in vitro* transcription assay, and then treated the samples with RNase H and RNase A to remove the RNA in this hybrid state and the excess single-stranded RNA, respectively. In the absence of the RNases, the mobility of the plasmid decreases, and consistent with the heterogeneity of their sizes as a result of R-loop formation, the plasmids migrate as a smear above the previously compact supercoiled/circular plasmid band (Fig. 2c and Supplementary Results). However, upon simultaneous treatment with RNase H and RNase A, the plasmids then migrate as distinct compact bands. The control plasmid shows a minimally altered mobility when compared to the HRE-containing plasmid after treatment with RNase H and/or RNase A (Extended Data Fig. 4d, e). Together, these results provide evidence that the formation of R-loops is an additional structurally polymorphic feature of the *C9orf72* HRE that contributes to the abortive transcription mechanism.

To further extend our observations concerning *in vitro* abortive transcription to possible truncation of the endogenous *C9orf72* transcript, we examined the production of pre-messenger RNA transcripts at different positions along the *C9orf72* gene in patient cells. Primers targeting intronic regions immediately upstream (5′) of the HRE versus downstream (3′) of the HRE region were used with quantitative reverse transcription polymerase chain reactions (qRT–PCR) to measure pre-mRNA levels. These results indicate that patient B lymphocytes carrying the *C9orf72* HRE show a significant decrease in the ratio of 3′/5′ pre-mRNA levels relative to the controls (Fig. 2d). This observation was extended to pathologically relevant tissue from patients such as the motor cortex and cervical spinal cord. Quantitative measurements of pre-mRNA levels using NanoString technology[20] show a similar decrease in the ratio of 3′/5′ pre-mRNA levels in patients carrying the HRE. These results indicate that the endogenous *C9orf72* HRE in patients induces abortive transcription and results in unequal transcriptional efficiency between regions upstream and downstream of the HRE, in agreement with our *in vitro* abortive transcript results (Fig. 2a, b).

Structure-dependent HRE-binding proteins

Because the *C9orf72* HRE can generate both full-length and abortive transcripts capable of forming RNA G-quadruplexes, we explored conformation-dependent protein interactions associated with the RNA repeat sequences. Biotinylated $(GGGGCC)_4$ RNAs, in either G-quadruplex or hairpin conformations, and antisense $(CCCCGG)_4$ hairpin RNAs were conjugated to streptavidin beads (Extended Data Fig. 1h). To comprehensively and quantitatively identify the conformation-dependent ribonucleoprotein complexes formed by these HRE RNAs, we used stable isotope labelling by amino acids in cell culture (SILAC)[30] with HEK293T cells grown in medium containing normal, medium or heavy amino acids before performing the RNA pull-down as shown in Extended Data Fig. 5. In brief, labelled cell lysates were incubated with RNA-conjugated streptavidin beads and the isolated ribonucleoprotein complexes were washed with increasing KCl concentrations to remove proteins that were loosely associated with the RNA or with other proteins. Then the final eluted fractions were analysed for complexes binding the $(GGGGCC)_4$ G-quadruplex, $(GGGGCC)_4$ hairpin, and $(CCCCGG)_4$ hairpin with the respective labels by mass spectrometry. The SILAC analysis provided a list of 288 proteins with quantification of their binding preferences (Supplementary Table 1) for the various RNA structures formed by the sense GGGGCC or the antisense CCCCGG sequences.

To confirm the specificity of the major proteins in binding to the various RNA structures identified by SILAC, we performed western blot analysis on the increasingly stringent fractions from the RNA pull-down (Fig. 3a). NCL and heterogeneous nuclear ribonucleoprotein (hnRNP) U preferentially recognize the RNA G-quadruplex motif (Fig. 3 and Extended Data Fig. 6). The proteins hnRNP F and RPL7, a heterogeneous nuclear ribonucleoprotein and a ribosomal protein, respectively, prefer guanine-rich RNA and bind both the G-quadruplex and the alternative hairpin of the GGGGCC repeat. The protein hnRNP K, an RNA-binding protein that binds cytosine-rich RNA, shows preferential binding to the antisense CCCCGG repeat. Differential KCl salt elution indicates that whereas the ribonucleoprotein complexes with hnRNP U or hnRNP F are more easily destabilized with increasing KCl, the interactions with NCL or RPL7 are more robust and survive high KCl concentrations. These results are in accordance with the SILAC quantitative analysis. To further demonstrate that NCL binds directly to the G-quadruplex motif, we used an *in vitro* RNA pull-down with

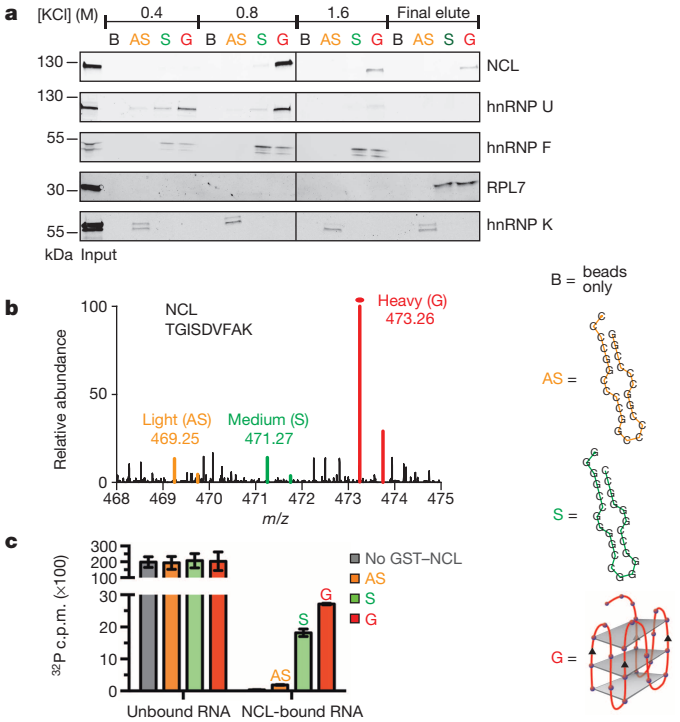


Figure 3 | Identification of conformation-dependent *C9orf72* HRE RNA-binding proteins. **a**, Western blotting analyses of the sequential fractions eluted from a biotinylated-RNA pull-down of HEK293T cells with increasing KCl concentrations. The RNA-binding proteins identified using LC-MS were differentiated among those that recognize different structural motifs, including the RNA antisense (AS) hairpin, $(CCCCGG)_4$; the RNA sense (S) hairpin $(GGGGCC)_4$; and the sense G-quadruplex (G). NCL and hnRNP U have binding preferences for the G motif of the sense RNA. hnRNP F and RPL7 bind both guanine-rich sense sequences, regardless of the underlying RNA structure. hnRNP K prefers binding to the cytosine-rich AS. **b**, A representative spectrum from SILAC analysis is shown for the preferential binding of NCL to the G motif formed by $(GGGGCC)_4$ compared to the hairpin motifs formed by the sense or antisense sequences. **c**, An RNA pull-down performed with GST–NCL demonstrates that NCL directly binds $(GGGGCC)_4$ RNA *in vitro* with highest affinity for the G-quadruplex. Data are means ± s.d. $n = 3$.

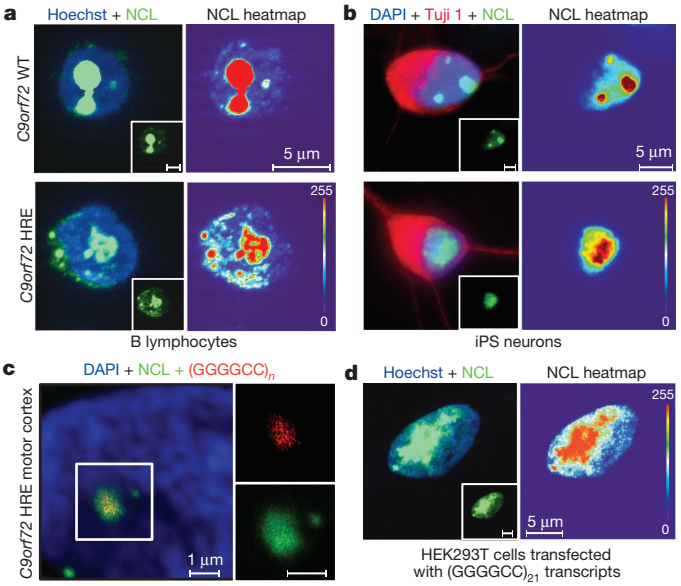


Figure 4 | Nucleolar stress is a result of repeat-containing RNA transcripts from the *C9orf72* HRE. **a**, In the control *C9orf72* WT B lymphocytes, NCL (green) is localized to the condensed nucleolus. In contrast, the cells of patients with the *C9orf72* HRE show an increased NCL diffusion and fractured nucleoli in the nucleus (Hoechst, blue). A heat map of NCL intensities marks the difference between cells. **b**, iPS motor neurons derived from patients carrying the *C9orf72* HRE also demonstrate NCL mislocalization. β-III tubulin (Tuj1) (red) was used to identify neurons. **c**, NCL colocalizes with RNA foci (red) formed in motor cortex tissue from patients carrying the *C9orf72* HRE. A $(CCCCGG)_{2.5}$ probe was used to detect the $(GGGGCC)_n$ RNA foci, a previously identified pathological feature of the *C9orf72* HRE tissues. **d**, Transfection of $(GGGGCC)_{21}$ abortive transcripts (Fig. 2a) recapitulates NCL pathological features observed in patient cells with the *C9orf72* HRE.

a recombinant glutathione *S*-transferase (GST)–NCL fusion protein. NCL was shown to preferentially pull down the $(GGGGCC)_4$ RNA G-quadruplexes (Fig. 3c). These results establish NCL as a specific and enriched interactor that directly and preferentially binds the RNA G-quadruplexes formed by the *C9orf72* HRE.

The *C9orf72* HRE causes nucleolar stress

To investigate the consequences of the specific binding of NCL, a principal component of the nucleolus[31] to G-quadruplex complexes formed by the *C9orf72* HRE RNA, we examined phenotypic differences in the cellular localization of NCL in *C9orf72* HRE patient cells and controls. In control B lymphocyte cells, endogenous NCL immunofluorescent staining is condensed in the nucleolus. In contrast, the nucleolus appears more fractured and NCL is more dispersed throughout the nucleus in patient cells carrying the *C9orf72* HRE, as shown in heat maps indicating the NCL density (Fig. 4a). Quantification of NCL distribution in the nuclei of patient and control cells confirms a significant shift of NCL away from dense nucleoli to a more dispersed localization (Extended Data Fig. 7a). This NCL pathology is also observed in *C9orf72* patient fibroblasts, but not in fibroblasts from non-ALS or non-*C9orf72* ALS controls (Extended Data Fig. 7a). Next, we examined the NCL localization pattern in induced pluripotent stem (iPS) motor neuron cells derived from patient fibroblasts[20]. As seen for the B lymphocytes, immunofluorescent staining for NCL in disease-relevant motor neurons from patients with the *C9orf72* HRE shows NCL dispersion that occupies a significantly increased area of the nucleus (Fig. 4b and Extended Data Fig. 7a). No obvious phenotypic differences were observed for hnRNP F, hnRNP U, or hnRNP K between *C9orf72* HRE B lymphocytes and those of controls (Extended Data Fig. 7d).

To confirm the association of NCL with *C9orf72* repeat RNA in pathologically relevant tissues of ALS/FTD patients with the *C9orf72* HRE, we performed RNA fluorescence *in situ* hybridization (FISH) in combination with NCL immunostaining in motor cortex tissues from these patients. The distribution of NCL in post-mortem tissues seems variable; however, it is evident that NCL frequently colocalizes with the GGGGCC RNA foci in the neurons of the motor cortex only in *C9orf72* HRE ALS patients (Fig. 4c and Extended Data Fig. 7c). These results support the interaction of NCL and the GGGGCC repeat transcripts *in vivo* and indicate the possible occurrence of functional defects associated with nucleolar stress in ALS/FTD patients.

Next, we examined whether the NCL mislocalization and nucleolar stress in the patient cells were caused by a gain in toxicity of the aberrant *C9orf72* RNA. To address this question directly, we transfected HEK293T cells with the abortive transcripts generated in the aforementioned *in vitro* reactions. Treatment with the 21-repeat-containing abortive transcripts recapitulates the NCL pathology observed in patient cells carrying the *C9orf72* HRE: NCL is significantly more dispersed in the nucleus (Fig. 4d and Extended Data Fig. 7e, f). Furthermore, the transfected HRE transcriptional products show a concentration-dependent decrease in cell viability compared to a control transcript (Extended Data Fig. 7g). Thus, our results establish a direct link between abortive *C9orf72* HRE transcripts, cytotoxicity and patient pathology.

The mislocalization of another nucleolar component, nucleophosmin (also known as B23), in B lymphocytes carrying the *C9orf72* HRE

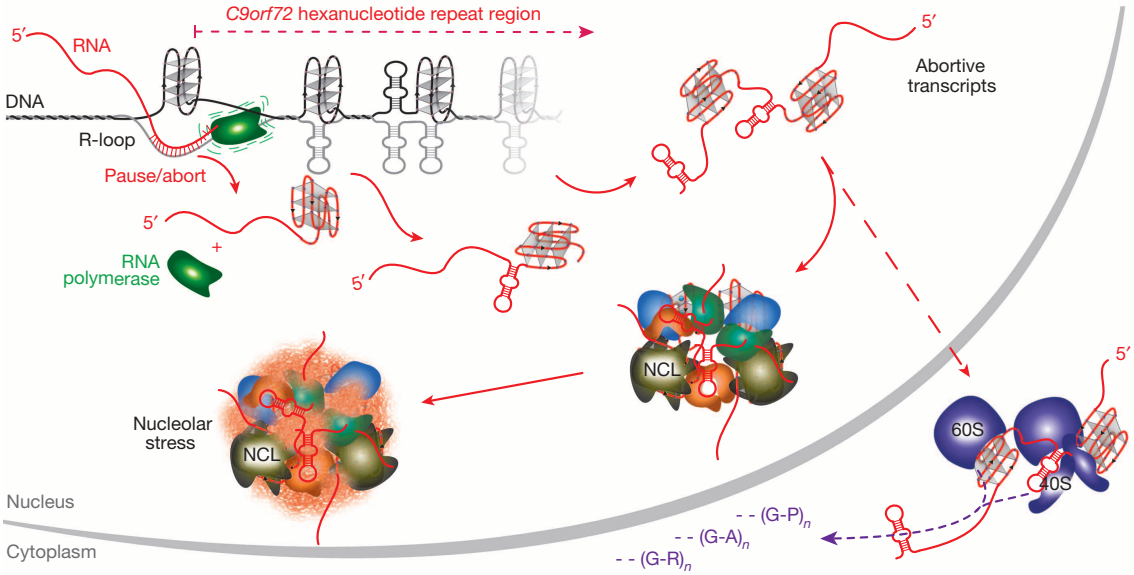


Figure 5 | A model for the molecular cascade resulting from the *C9orf72* HRE structural polymorphism. The DNA and RNA•DNA structures formed in the GGGGCC repeat region impede RNA polymerase transcription, which results in transcriptional pausing and abortion. This leads to a loss of full-length products and an accumulation of abortive transcripts. Abortive transcripts that contain the hexanucleotide repeats form G-quadruplexes and hairpins and bind essential proteins in a conformation-dependent manner. Sequestration of these proteins leads to nucleolar stress and other downstream defects. The repeat-containing transcripts can also escape the nucleus and be bound by ribosomal complexes, thereby increasing repeat-associated non-ATG-dependent translation that results in aggregative polydipeptides.

further indicates general nucleolar stress (Extended Data Fig. 8a, b). To determine whether the functions of the nucleolus are impaired in patient cells and tissues, we used qRT–PCR to measure the processing and maturation of the 45S ribosomal RNA (Extended Data Fig. 8c). The results show a decrease in the processing of the precursor 45S rRNA into the mature cleaved 28S, 18S and 5.8S rRNAs in *C9orf72* HRE patient B lymphocytes (Extended Data Fig. 8d). Furthermore, this rRNA maturation is significantly decreased in the motor cortex tissues from ALS patients carrying the *C9orf72* HRE (Extended Data Fig. 8d). Taken together, these results identify a functional defect involving the biochemical effects of the *C9orf72* HRE in nucleolar stress and directly link this defect to disease pathology.

Next, to examine other RNA-related functional defects that could arise in addition to the chronic nucleolar stress identified in *C9orf72* HRE patients, we measured the abundance of processing bodies (P bodies) in patient-derived motor neurons. P bodies, which are composed of ribonucleoprotein complexes involved in the degradation of untranslated mRNAs[32], are significantly increased in number, but not size, in iPS motor neurons from patients carrying the *C9orf72* HRE relative to controls (Extended Data Fig. 9a, b). This increase in P bodies is consistent with observations of decreased ribosomal maturation caused by nucleolar stress, which can then lead to increased untranslated mRNA in the cytoplasm and to a global perturbation in RNA processing.

Finally, to test whether the cells of patients harbouring the *C9orf72* HRE are more sensitive to proteotoxic stress, we treated the iPS motor neurons with tunicamycin, which induces endoplasmic reticulum stress and unfolded protein response. *C9orf72* HRE patient neurons show a dose-dependent increase in sensitivity to tunicamycin compared to controls (Extended Data Fig. 9c, d). These results are consistent with the perturbation in protein homeostasis, which is possibly linked to the nucleolar stress identified in ALS patients carrying the *C9orf72* HRE.

Discussion

The biological activity of nucleic acids is determined not only by their linear sequence of nucleotides, but also by their structural diversity. Here we found that the generation of distinct polymorphic structures of the *C9orf72* HRE DNA and RNA, such as G-quadruplexes and RNA•DNA hybrids (R-loops), underlie HRE-dependent molecular events, including abortive transcription and sequestration of unique RNA-binding proteins. The identification of conformation-dependent ribonucleoprotein complexes and a specific nucleolar pathology provides a cohesive mechanism for the disease (Fig. 5 and Supplementary Discussion): An impairment in transcriptional processivity by the HRE leads to accumulation of abortive transcripts, and these repeat-containing RNAs sequester proteins that recognize their distinctive conformations, sensitizing cells for chronic neurodegenerative damage. Our findings concerning NCL represent a direct link between the characteristic G-quadruplexes of the *C9orf72* HRE and the resulting cascade of pathological defects in ALS/FTD patients.

The nucleolus is a central hub in cellular stress responses[33], and NCL has been shown to have a critical role in the long-term maintenance of mature neurons[34]. Our results demonstrate that motor neurons harbouring the *C9orf72* HRE show defects in rRNA processing and increased sensitivity to stresses related to protein homeostasis, suggesting that affected patient neurons are more vulnerable to age-dependent chronic stresses. Together, these results point to the targeting of these toxic nucleic acid conformations as a possible intervention at the root of the pathogenic cascade, and also suggest a mechanistic model for similar repeat expansion neurodegenerative diseases that share nucleic acid features of the structurally polymorphic *C9orf72* HRE.

METHODS SUMMARY

CD analysis of nucleic acid structures was performed as described[35], EMSAs were performed as recommended[36,37], and DMS protection assays followed a previously described protocol[26,36]. The RNase protection assays were performed following Ambion's recommendations. Plasmids were constructed in a pCR8-TOPO vector (Invitrogen), and GGGGCC HRE inserts were generated using a self-templating PCR protocol[38]. *In vitro* transcription reactions were performed with these plasmids and analysed on sequencing gels. R-loop assays were adapted from previously described methods[39]. The complementary DNAs from B lymphocytes or RNAs from human tissues with or without the *C9orf72* HRE were generated from total RNA following manufacturer's protocols and relative levels were then measured. NanoString RNA analysis followed standard protocols as previously described[20]. The RNA pull-down with isotopically labelled HEK293T lysates and biotinylated RNA conjugated to streptavidin beads followed a previously described protocol[40], with an additional KCl gradient wash. Quantitative mass spectrometry

was performed by using a three-state SILAC analysis using a filter-aided sample preparation (FASP) method followed by analysis on an LTQ-Orbitrap Elite mass spectrometer[30,41]. Peptides were identified using the Mascot search algorithm. Western blotting was performed on RNA pull-down fractions according to the manufacturer's recommendations for each antibody. Immunofluorescent staining of lymphocytes, HEK293T cells, fibroblasts and iPS motor neurons followed a standard protocol described in detail in Methods. RNA FISH with immunofluorescent on human motor cortex tissue was performed essentially as previously described[20].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April 2013; accepted 5 February 2014.
Published online 5 March 2014.

1. Richard, G. F., Kerrest, A. & Dujon, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* **72,** 686–727 (2008).
2. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–256 (2011).
3. Renton, A. E. *et al.* A hexanucleotide repeat expansion in *C9ORF72* is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
4. Kiernan, M. C. *et al.* Amyotrophic lateral sclerosis. *Lancet* **377,** 942–955 (2011).
5. Majounie, E. *et al.* Frequency of the *C9orf72* hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol.* **11,** 323–330 (2012).
6. Rademakers, R., Neumann, M. & Mackenzie, I. R. Advances in understanding the molecular basis of frontotemporal dementia. *Nature Rev. Neurol.* **8,** 423–434 (2012).
7. van Langenhove, T., van der Zee, J. & van Broeckhoven, C. The molecular basis of the frontotemporal lobar degeneration–amyotrophic lateral sclerosis spectrum. *Ann. Med.* **44,** 817–828 (2012).
8. Rollinson, S. *et al.* Analysis of the hexanucleotide repeat in *C9ORF72* in Alzheimer's disease. *Neurobiol. Aging* **33,** 1846.e5–1846.e6 (2012).
9. Kohli, M. A. *et al.* Repeat expansions in the C9ORF72 gene contribute to Alzheimer's disease in Caucasians. *Neurobiol. Aging* **34,** 1519.e5–1519.e12 (2013).
10. Majounie, E. *et al.* Repeat expansion in *C9ORF72* in Alzheimer's disease. *N. Engl. J. Med.* **366,** 283–284 (2012).
11. Hensman-Moss, D. J. *et al. C9orf72* expansions are the most common genetic cause of Huntington disease phenocopies. *Neurology* **82,** 292–299 (2014).
12. Rutherford, N. J. *et al.* Length of normal alleles of *C9ORF72* GGGGCC repeat do not influence disease phenotype. *Neurobiol. Aging* **33,** 2950.e5–2950.e7 (2012).
13. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447,** 932–940 (2007).
14. La Spada, A. R. & Taylor, J. P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature Rev. Genet.* **11,** 247–258 (2010).
15. Fratta, P. *et al. C9orf72* hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.* **2,** 1016 (2012).
16. Reddy, K., Zamiri, B., Stanley, S. Y. R., Macgregor, R. B. & Pearson, C. E. The Disease-associated r(GGGGCC)$_n$ repeat from the *C9orf72* gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.* **288,** 9860–9866 (2013).
17. Xu, Z. *et al.* Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl Acad. Sci. USA* **110,** 7778–7783 (2013).
18. Mori, K. *et al.* The *C9orf72* GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339,** 1335–1338 (2013).
19. Ash, P. E. A. *et al.* Unconventional translation of *C9ORF72* GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77,** 639–646 (2013).
20. Donnelly, C. J. *et al.* RNA toxicity from the ALS/FTD *C9ORF72* expansion is mitigated by antisense intervention. *Neuron* **80,** 415–428 (2013).
21. Lagier-Tourenne, C. *et al.* Targeted degradation of sense and antisense *C9orf72* RNA foci as therapy for ALS and frontotemporal degeneration. *Proc. Natl Acad. Sci. USA* **110,** E4530–E4539 (2013).
22. Sareen, D. *et al.* Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a *C9ORF72* repeat expansion. *Sci. Transl. Med.* **5,** 208ra149 (2013).
23. Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in *C9ORF72* ALS and frontotemporal dementia. *Proc. Natl Acad. Sci. USA,* (2013).
24. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA* **48,** 2013 (1962).
25. Huppert, J. L. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* **37,** 1375 (2008).
26. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA* **74,** 560–564 (1977).
27. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5′ UTR of the *NRAS* proto-oncogene modulates translation. *Nature Chem. Biol.* **3,** 218–221 (2007).
28. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chem.* **5,** 182–186 (2013).
29. Li, H. *et al.* Ultrasensitive label-free amplified colorimetric detection of p53 based on G-quadruplex MBzymes. *Biosens. Bioelectron.* **50,** 180–185 (2013).
30. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386 (2002).
31. Abdelmohsen, K. & Gorospe, M. RNA-binding protein nucleolin in disease. *RNA Biol.* **9,** 799–808 (2012).
32. Eulalio, A., Behm-Ansmant, I. & Izaurralde, E. P bodies: at the crossroads of post-transcriptional pathways. *Nature Rev. Mol. Cell Biol.* **8,** 9–22 (2007).
33. Boulon, S., Westman, B. J., Hutten, S., Boisvert, F.-M. & Lamond, A. I. The nucleolus under stress. *Mol. Cell* **40,** 216–227 (2010).
34. Hetman, M. & Pietrzak, M. Emerging roles of the neuronal nucleolus. *Trends Neurosci.* **35,** 305–314 (2012).
35. Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126,** 16405–16415 (2004).
36. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols* **2,** 1849–1861 (2007).
37. Sun, D. & Hurley, L. H. in *Methods in Molecular Biology* (ed. Baumann, P.) **608,** 65–79 (Humana, 2009).
38. Hsu, R.-J. *et al.* Long tract of untranslated CAG repeats is deleterious in transgenic mice. *PLoS ONE* **6,** e16417 (2011).
39. Reddy, K. *et al.* Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.* **39,** 1749–1762 (2011).
40. Iioka, H., Loiselle, D., Haystead, T. A. & Macara, I. G. Efficient detection of RNA-protein interactions using tethered RNAs. *Nucleic Acids Res.* **39,** e53 (2011).
41. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature Methods* **6,** 359–362 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. J. Traynor for the information on the ALS patient cells, M. DeJesus-Hernandez and R. Rademakers for the genotyping PCR protocols on patient cells, P. J. Tienari for providing additional fibroblast/iPS cell lines, M. J. Matunis for the hnRNP F/H antibody, R. S. Nirujogi for his assistance with the SILAC LC-MS analysis, T. K. Chakravarty for his assistance with transfection of HEK293T cells, the S. Balasubramanian laboratory for generously providing the BG4 plasmid, P. L. Steinwald for purification and assistance in the validation of the BG4 protein, R. M. Powers for assistance in the staining and quantification of P bodies in iPS neurons, and O. Pletnikova for processing human tissues. This work was supported by National Institutes of Health grants NS07432 (J.W.) and NS085207 (R.S. and J.D.R.), the Robert Packard Center for ALS Research at Johns Hopkins and the Muscular Dystrophy Association (J.W.), Target ALS (J.D.R.), the ALS Association (J.W.), the Johns Hopkins Alzheimer's Disease Research Center NIH P50AG05146 (J.C.T.), National Cancer Institute training grant 5T32CA009110-36 (A.R.H. and E.A.J.S.), Maryland Stem Cell Research Fund (C.J.D.), Judith & Jean Pape Adams Charitable Foundation (R.S.), and the Samuel I. Newhouse Foundation (J.C.T.). We thank P. A. Coulumbe, V. C. Culotta, M. J. Matunis, P. S. Miller, B. A. Learn and members of the J. Wang laboratory for manuscript advice and discussions.

Author Contributions A.R.H. performed the nucleic acid structure analysis and the *in vitro* transcription studies. G.P. constructed plasmids for *in vitro* transcription studies. E.A.J.S. and A.R.H performed and analysed the R-loop assay. A.R.H. performed and analysed the RNA pull-down. M.-S.K., P.G.S. and A.P. performed MS analysis on pull-down fractions. A.R.H. performed all B lymphocyte immunofluorescence and analysis. C.J.D., R.S. and J.D.R. performed all patient-derived iPS motor neuron and human tissue analysis. N.J.M. maintained patient cell cultures. J.D.R. and J.C.T. provided patient tissue samples. A.R.H. and J.W. designed the studies, developed the model, and wrote the paper with insights from G.P. and C.J.D. A.R.H. developed and arranged the figures with insights from J.W. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W. (jiouw@jhmi.edu).

ARTICLE

doi:10.1038/nature12966

Proof of principle for epitope-focused vaccine design

Bruno E. Correia[1,2,3], John T. Bates[4], Rebecca J. Loomis[5], Gretchen Baneyx[1], Chris Carrico[6], Joseph G. Jardine[1,7,8,9], Peter Rupert[6], Colin Correnti[6], Oleksandr Kalyuzhniy[1,8,9], Vinayak Vittal[1], Mary J. Connell[5], Eric Stevens[1], Alexandria Schroeter[1], Man Chen[10], Skye MacPherson[1,7,8,9], Andreia M. Serra[1,8,9], Yumiko Adachi[1,8,9], Margaret A. Holmes[6]‡, Yuxing Li[7,8,9], Rachel E. Klevit[1], Barney S. Graham[10], Richard T. Wyatt[7,8,9], David Baker[1], Roland K. Strong[6], James E. Crowe Jr[4,11,12], Philip R. Johnson[5] & William R. Schief[1,7,8,9]

Vaccines prevent infectious disease largely by inducing protective neutralizing antibodies against vulnerable epitopes. Several major pathogens have resisted traditional vaccine development, although vulnerable epitopes targeted by neutralizing antibodies have been identified for several such cases. Hence, new vaccine design methods to induce epitope-specific neutralizing antibodies are needed. Here we show, with a neutralization epitope from respiratory syncytial virus, that computational protein design can generate small, thermally and conformationally stable protein scaffolds that accurately mimic the viral epitope structure and induce potent neutralizing antibodies. These scaffolds represent promising leads for the research and development of a human respiratory syncytial virus vaccine needed to protect infants, young children and the elderly. More generally, the results provide proof of principle for epitope-focused and scaffold-based vaccine design, and encourage the evaluation and further development of these strategies for a variety of other vaccine targets, including antigenically highly variable pathogens such as human immunodeficiency virus and influenza.

Vaccination is a proven, safe and cost-effective way to protect against infectious disease[1,2], but potentially vaccine-preventable illnesses continue to place a heavy burden on the human population. Data from recent epidemiological studies indicate that in 2010, infectious diseases caused 18.5% of all human deaths and 23% of disability-adjusted life years[3,4]. This burden could be reduced by broader deployment and use of existing vaccines or by other prevention modalities or treatment regimens. However, for maximal, affordable and sustainable gains in global health, new or improved vaccines are needed for several major pathogens including human immunodeficiency virus (HIV)-1 (ref. 5), malaria[6], *Mycobacterium tuberculosis*[7], influenza virus[8], dengue virus[9] and respiratory syncytial virus (RSV)[10]. One likely impediment to vaccine development in these cases is the limited set of antigen design or presentation methods available to vaccine engineers. For example, current licensed vaccines in the United States[11] derive from strategies that have been available for many years: viral vaccines are composed of recombinant virus-like particles or live, live-attenuated or whole inactivated viruses or subunit vaccines, and bacterial vaccines are composed of bacterial surface proteins, detoxified toxins or polysaccharides with or without conjugation to a carrier protein.

Epitope-focused vaccine design is a conceptually appealing but unproven method in which immunogens are designed to elicit protective antibody responses against structural epitopes that are defined by protective antibodies isolated from infected patients or animal models[12]. This strategy, if validated, could offer a potential route to vaccines for many pathogens that have resisted traditional vaccine development, including highly antigenically variable viruses such as HIV, influenza and hepatitis C virus, for which broadly neutralizing antibodies have been discovered and characterized structurally with their target epitopes[13]. We tested the feasibility of this strategy using an epitope from RSV, a virus that causes lower respiratory tract infections in children and the elderly. In 2010 RSV was estimated to be responsible for 6.7% of all deaths in children of ages 1 month to 1 year[3]. We focused on the epitope targeted by the licensed, prophylactic neutralizing antibody palivizumab (also known as Synagis, pali) and an affinity-matured variant, motavizumab (mota)[14]. A crystal structure of mota in complex with its epitope from the RSV Fusion (F) glycoprotein revealed that the antibody-bound epitope attains a helix-turn-helix conformation[15].

We previously developed 'side-chain grafting' and 'backbone grafting' methods to transplant continuous or discontinuous epitopes to scaffold proteins of known structure, for epitope conformational stabilization and immune presentation[16–20]. Epitope scaffold immunogens designed by these methods for epitopes from HIV or RSV (including the mota epitope) have in some cases induced structure-specific antibodies but have failed to induce neutralizing antibodies[16–18]. Because these methods are restricted to scaffold proteins of predetermined structure, we have developed a new computational method to design scaffold proteins with full backbone flexibility, to allow greater precision in tailoring scaffold structures for particular epitope structures. We used this method to design scaffolds for the mota epitope, and we found that the scaffolds had favourable biophysical and structural properties and that scaffold immunization of rhesus macaques induced RSV-neutralizing activity (Fig. 1).

Computational method

Great strides have been made in developing *de novo* methods to design arbitrary, idealized protein structures[21,22], but the resulting proteins have lacked functional activity. We devised a computational method

[1]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [2]PhD Program in Computational Biology, Instituto Gulbenkian Ciência and Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras 2780-157, Portugal. [3]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA. [4]The Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, Tennessee 37232, USA. [5]The Children's Hospital of Philadelphia Research Institute, Philadelphia, Pennsylvania 19104, USA. [6]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024, USA. [7]Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California 92037, USA. [8]IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California 92037, USA. [9]Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, California 92037, USA. [10]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [11]Department of Pathology, Microbiology and Immunology, Vanderbilt Medical Center, Nashville, Tennessee 37232, USA. [12]Department of Pediatrics, Vanderbilt Medical Center, Nashville, Tennessee 37232, USA.
‡Deceased.

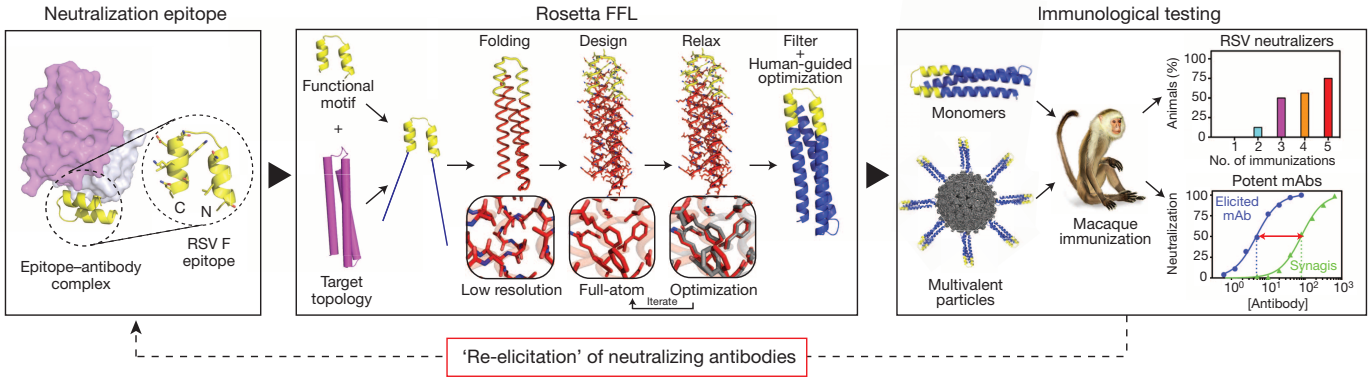


Figure 1 | A new computational method to design epitope-focused vaccines, illustrated with a neutralization epitope from RSV. Stages of computational design and immunological evaluation are shown; biophysical and structural evaluation are also important (see text). mAbs, monoclonal antibodies.

to allow *de novo* folding and design of scaffold proteins stabilizing functional motifs (Extended Data Fig. 1). This procedure, called Fold From Loops (FFL), has four stages: (1) selection of the functional motif and target topology to be folded around the motif; (2) *ab initio* folding to build diverse backbone conformations consistent with the target topology; (3) iterative sequence design and structural relaxation to select low-energy amino-acid sequences for the given backbone conformations; (4) filtering and human-guided optimization, in which the best designs are identified by structural metrics and then subjected to optional human-guided sequence design to correct remaining flaws.

Design of epitope scaffolds

To design scaffolds for the helix-turn-helix conformation of the mota epitope (PDB accession 3IXT, chain P), we selected a three-helix bundle (PDB 3LHP, chain S) as the template topology. Knowing that the template protein folds into thermally stable, soluble monomers[16,23], we designed scaffolds of similar length and position-dependent secondary structure. We produced 40,000 designs using FFL stages 1–3 and then used multiple structural filters to select eight designs for human-guided optimization. Additional modifications were made to those designs as follows: first, to optimize solubility, nearly all surface residues outside the epitope were replaced with those from the template protein; and second, to optimize side-chain packing in the buried protein core, computational design was used to design larger hydrophobic residues at selected buried positions of most designs (Extended Data Fig. 2). The final eight FFL designs had similar but non-identical backbone conformations (pairwise root mean squared deviation (r.m.s.d.) ranging from 0.5 to 3.0 Å) with correspondingly diverse core packing solutions differing from each other by 8 to 42 mutations and from the template by 56 mutations on average (Extended Data Fig. 2). All eight FFL designs had identical surface residues (including non-epitope residues taken from the template, as well as the epitope itself). To create fully artificial scaffolds with different antigenic surfaces that could be used in heterologous prime-boost regimens with FFL scaffolds or to map immune responses to FFL scaffolds, we resurfaced[23] the FFL_001 design; this produced the 'FFL_surf' designs (Extended Data Fig. 2) that differed from FFL_001 by 36 mutations on average and had no significant sequence similarity (BLAST *E* value $< 10^{-3}$) to any known protein except the RSV F protein.

Biophysical and structural characterization

Six out of eight FFL designs and three out of four FFL_surf designs could be expressed in *Escherichia coli* and purified, with yields ranging from 3 to 5 mg l^{-1}. These nine scaffolds were monomeric in solution, showed circular dichroism spectra typical for properly folded helical proteins, and all but one were highly thermally stable with melting temperatures (T_m) greater than 75 °C (Fig. 2a, b, Table 1 and Extended Data Fig. 3). ^{15}N heteronuclear single quantum coherence (HSQC) spectra were collected for four FFL designs, and these data showed reasonable to good peak dispersion, typical of well-behaved, globular proteins with high α-helical content in solution (Fig. 2c, Table 1 and Extended Data Fig. 3).

The nine purifiable FFL and FFL_surf scaffolds all had high affinity for mota, as assessed by surface plasmon resonance (SPR) ($K_D < 800$ pM, Fig. 2d, Table 1 and Extended Data Fig. 4). In particular, six out of nine scaffolds had very high mota affinities ($K_D = 6$–94 pM) and slow dissociation rates ($k_{off} \sim 10^{-4}$ s^{-1}) comparable to those of the mota interaction with the RSV F glycoprotein ($K_D = 35$ pM and $k_{off} = 0.31 \times 10^{-4}$ s^{-1})[14]. The mota–scaffold interaction was also specific—the point mutation K82E on scaffold FFL_001, analogous to the K272E mota escape mutation on RSV F[24], reduced mota binding by more than a factor of 1,000 (Extended Data Fig. 4). These results suggested that the conformation of the native epitope was reproduced accurately on the scaffolds. The mota affinities for FFL scaffolds were three to four orders of magnitude higher than the mota affinities for the free epitope peptide ($K_D = 210$–240 nM[18]) or for the best side-chain-grafting epitope scaffold previously reported ($K_D = 90$–125 nM[18]).

To evaluate the degree to which high-resolution structures of the FFL designs recapitulated the design models and the mota epitope, we solved two crystal structures: unliganded FFL_005 and the complex of FFL_001 bound to mota Fab (Fig. 2e, f), to resolutions of 2.0 and 2.7 Å, respectively (Extended Data Fig. 5). The crystal structures showed good overall agreement with the design models—the backbone r.m.s.d. over all residues was 1.7 Å for FFL_005 and 1.2 Å for FFL_001 (Fig. 2e, f), and the all-atom r.m.s.d. for the core side chains was 2.5 Å for FFL_005 and 1.8 Å for FFL_001. Consistent with the biophysical data, both unliganded and mota-bound structures revealed a high degree of epitope mimicry. Compared to the structure of peptide in complex with mota (PDB 3IXT), the epitope backbone r.m.s.d. was 0.5 Å for FFL_005 (Fig. 2g) and 0.4 Å for FFL_001. Compared to structures of pre- and post-fusion RSV F trimer (PDB 4JHW and 3RRR), which were not available at the time the designs were carried out, epitope backbone r.m.s.d. was 0.3 and 0.4 Å for FFL_005, respectively. Furthermore, the interaction of FFL_001 with mota accurately recapitulated the interaction of mota with peptide; superposition of the epitope and paratope of both complexes gave an all-atom r.m.s.d. of 0.8 Å (Fig. 2h).

Immunological evaluation

To assess whether humans can make antibodies specific for the RSV epitope structure stabilized on the scaffolds, we tested the binding of sera from six RSV-seropositive humans to RSV F, FFL_001 and FFL_001 variants with two different epitope mutations (N72Y and K82E) corresponding to RSV escape mutations for pali (N262Y and K272E) and mota (K272E) (Fig. 2g and Extended Data Fig. 6). Although all sera reacted with RSV F and none reacted to the scaffold escape mutants, three sera displayed reactivity to FFL_001. These data confirmed that

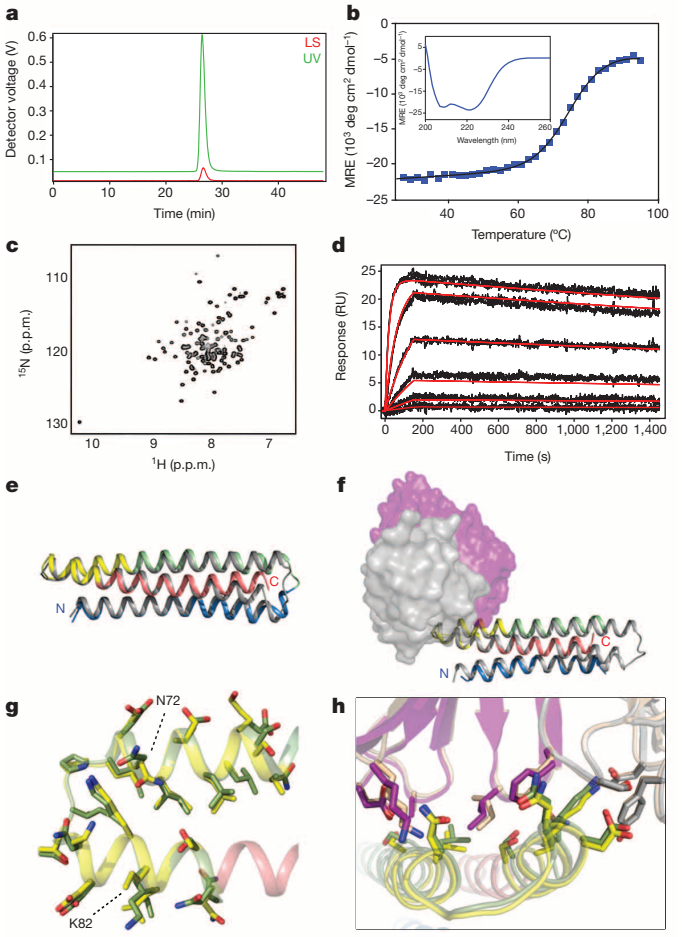


Figure 2 | Biophysical and structural characterization of scaffold FFL_001. **a**, Size-exclusion chromatography coupled in-line with multi-angle light scattering measured a molecular weight in solution of ~15 kDa, corresponding to a monomer. LS, light scattering; UV, ultraviolet absorbance. **b**, Circular dichroism data fit with a two-state model showed that the protein had a melting temperature of 74 °C. Inset, the wavelength scan at 25 °C exhibited two minima characteristic of an all-helical protein. Deg, degrees; MRE, mean residue ellipticity. **c**, Two-dimensional ^{1}H-^{15}N HSQC spectrum at 25 °C and 600 MHz showed good peak dispersion typical of well-folded, α-helical proteins. **d**, SPR data and model fits (red lines) of the interaction with mota Fab analyte, from which the dissociation constant (K_D) was measured to be 29.9 pM. Similar results were obtained with scaffold analyte and mota IgG ligand. RU, response units. **e**, Crystal structure of unliganded FFL_005 (blue, green and salmon helices, with yellow epitope), superimposed with the design model (grey with yellow epitope). **f**, Crystal structure of FFL_001 bound to mota Fab, superimposed with the design model. Colouring as in **e**, but with the Fab light and heavy chains in grey and purple, respectively. **g**, Superposition of the epitope structure from unliganded FFL_005 (yellow) and the complex of peptide (green) bound to mota (PDB 3IXT). FFL_005 is coloured salmon outside the epitope. The positions of escape mutations for pali (262 and 272, RSV numbering) or mota (272) are noted. **h**, Superposition of the mota-liganded structures of FFL_001 and peptide (PDB 3IXT). The antibody chains of 3IXT are coloured in wheat, and the interfacial side chains of both epitope and antibody are shown in stick representation.

the FFL scaffolds presented a clinically relevant epitope conformation and illustrated that epitope scaffolds have promise as reagents to assay levels of epitope-structure-specific antibodies in polyclonal sera.

We tested whether the FFL scaffolds could induce RSV-neutralizing antibodies by vaccination in both BALB/c mice and rhesus macaques. Four immunogens were tested: monomeric scaffolds FFL_001, FFL_005 and FFL_007, and a virus-like particle consisting of hepatitis B core antigen (HBcAg) particles conjugated with multiple copies of FFL_001 (refs 25, 26). Mice produced robust binding antibody responses against the autologous antigens, but binding antibody responses against RSV F protein or RSV viral lysate were detected in only a few animals (Extended Data Fig. 6), and neutralizing activity as judged by a plaque reduction assay was not detected (not shown). In contrast to the mouse results, after three immunizations all macaques produced robust binding responses not only against the autologous antigens (Fig. 3a) but also recombinant RSV F protein (Extended Data Fig. 6), and most animals responded to RSV viral lysate (Fig. 3a and Supplementary Table 1). Neutralizing activity was detected by the plaque assay in 7 out of 16 macaques after three immunizations and in 12 of 16 macaques after five immunizations (Figs 1 and 3a and Supplementary Table 2). Neutralizing activities were confirmed at selected time points using two different assays (microneutralization and a flow cytometry-based assay) in different laboratories, and included measurement of neutralizing activity against RSV subtype B[27] as well as subtype A (Extended Data Figs 6, 7 and Supplementary Table 2). To benchmark the neutralization potency, selected macaque sera were tested side by side with sera from seropositive human adults, in both the plaque reduction and flow cytometry assays (Fig. 3b, c). The results in both assays demonstrate that the best-responding macaques, including two out of four animals in the particle group at week 20 and one animal in that group at week 12, have neutralization titres comparable to those induced by natural human infection. This is noteworthy given that natural infection exposes multiple epitopes on the RSV F and G glycoproteins, whereas the scaffolds exposed only one epitope.

Monoclonal antibody characterization

To study the molecular basis for the vaccine-induced neutralizing activity, we used single-B-cell sorting to isolate epitope-specific monoclonal antibodies[28] from memory B cells of one animal from the particle group with potent serum neutralizing activity. We isolated B cells that bound strongly to FFL_001 but not to a double mutant of FFL_001 (FFL_001_N72Y_K82E) containing both pali escape mutations. Following DNA sequencing of antibody variable genes in those cells, we produced 11 recombinant monoclonal antibodies, of which eight bound with high avidity to FFL_001 and two (17-HD9 and 31-HG7) bound with high avidity to RSV F protein (Fig. 4a and Extended Data Fig. 8). SPR revealed that these two monoclonal antibodies, which are clonal relatives, have extremely high affinities ($K_D \approx 3$ pM) for the scaffold FFL_001 that elicited them when mounted on the particle (Extended Data Fig. 8). Concomitant with high affinities, these two monoclonal antibodies have neutralization potencies similar to mota and higher than pali by nearly an order of magnitude (Fig. 4b and Extended Data Fig. 8).

To map the epitopes for 17-HD9 and 31-HG7, we assessed binding to several scaffold variants (Extended Data Fig. 9). Both monoclonal antibodies: (1) bound with very high affinity (K_D = 40–50 pM) to FFL_001_surf1, which has an antigenically distinct surface from FFL_001 outside the RSV epitope; (2) retained high affinity (K_D = 180–330 pM) for the FFL_001_K82E mota escape mutant; (3) retained modest affinity (K_D = 60–140 nM) for the FFL_001_N72Y_K82E double escape mutant; and (4) lacked detectable affinity for FFL_MPV_001, which swaps RSV residues on FFL_001 to those at the analogous positions on human metapneumovirus, which has a similar helix-turn-helix conformation (r.m.s.d. 0.9 Å, PDB 4DAG[29]) but very different amino acid sequence. These results indicate that the two macaque monoclonal antibodies target the same helix-turn-helix epitope as mota and pali but have different fine specificities.

To understand the structural basis for the binding and neutralizing potency of these macaque monoclonal antibodies, we pursued crystallography of 17-HD9 and 31-HG7 complexes with FFL_001. We obtained crystals of the 31-HG7–FFL_001 complex that diffracted to 3.8 Å, which was sufficient to determine a molecular replacement solution using the FFL_001 crystal structure and a composite Fab model, but insufficient to perform detailed rebuilding and refinement. The molecular replacement solution allowed determination of the rigid-body orientation of 31-HG7 relative to FFL_001 and demonstrated that 31-HG7 approaches the helix-turn-helix from a different angle than mota (angle difference

Table 1 | Biophysical properties of scaffolds and scaffold–antibody interactions

Molecule	Multimeric state	T_m (°C)	ΔG (kcal mol^{-1})	SPR motavizumab			NMR-HSQC dispersion
				k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	k_{off}/k_{on} (pM)	
FFL_001	Mon	76	ND	3.99×10^6	1.19×10^{-4}	29.9	Dispersed
FFL_002	Mon	49	ND	1.56×10^6	7.34×10^{-4}	469.9	ND
FFL_004	Mon	>85	ND	1.05×10^6	8.32×10^{-4}	795.0	ND
FFL_005	Mon	>100	15.0	2.97×10^6	2.09×10^{-4}	70.3	Partially dispersed
FFL_006	Mon	>85	ND	3.57×10^6	2.32×10^{-4}	651.9	Dispersed
FFL_007	Mon	>85	ND	1.45×10^6	1.36×10^{-4}	94.1	Partially dispersed
FFL_001_surf1	Mon	84	8.2	7.43×10^6	4.70×10^{-4}	6.32	ND
FFL_001_surf2	Mon	>85	8.1	5.32×10^6	1.58×10^{-4}	29.6	ND
FFL_001_surf4	Mon	>85	9.0	4.80×10^6	1.58×10^{-4}	32.9	ND

Mon, monomer; ND, not done.

~56°, Fig. 4c). We also obtained crystals and determined the structure of the 17-HD9–FFL_001 complex (resolution = 2.5 Å), which contained four complexes of 17-HD9 bound to a 35-residue helix-turn-helix peptide (scaffold substructure) in the asymmetric unit (Fig. 4c and Extended Data Fig. 10). The 17-HD9 complex structures demonstrated that 17-HD9 recognizes essentially the same helix-turn-helix epitope as mota and pali—the conformation of the epitope in the 17-HD9 complexes is very similar to that in the structures of mota–FFL_001 (r.m.s.d. 0.5–0.7 Å), RSV F pre-fusion (r.m.s.d. 0.3–0.4 Å) and RSV F post-fusion (r.m.s.d. 0.5 Å), and 85% of the epitope residues buried by either mota or 17-HD9 are also buried by the other (Fig. 4d and Supplementary Table 3). Although 17-HD9 and mota bury a similar amount of area on the epitope (690 Å^2 versus 683 Å^2), 17-HD9 uses a different paratope to make more hydrogen bonds (15–18 versus 7) that plausibly contribute to its higher scaffold affinity and higher neutralization potency (Fig. 4e and Supplementary Tables 4, 5). The 17-HD9 complexes are also consistent with the ability of 17-HD9 to bind to the K82E mota escape mutant: density for the K82 side chain is absent in two out of four 17-HD9 complexes, and K82 is only 37% buried by 17-HD9 in the other two complexes (Fig. 4e); by contrast, K82 is 65% buried by mota and makes a buried salt bridge to mota light-chain residue D50. Taken together, these results demonstrate that epitope scaffold immunization can 're-elicit' neutralizing antibodies that target with high precision an epitope predefined by a protective antibody.

Discussion

We have demonstrated that small, thermally and conformationally stable protein scaffolds that accurately mimic the structure of a viral neutralization epitope can induce neutralizing activity in a majority of vaccinated macaques. The results establish the feasibility of epitope-focused and scaffold-based vaccine design, and encourage the application of these strategies for a variety of vaccine targets. The biophysical, structural and functional data on the mota scaffolds validate the computational design method (FFL), and support its continued development and application to other vaccine epitopes and other types of functional sites. Indeed, the data should encourage the general use of methods using protein backbone flexibility to design novel functional proteins.

The scaffolds themselves represent promising leads for RSV vaccine research and development (particularly the scaffolds presented on virus-like particles). Non-replicating RSV vaccine candidates are not tested in RSV naive young infants, the highest priority target population, owing to vaccine-mediated disease enhancement in early clinical trials of formalin-inactivated RSV[10]. Scaffold immunogens that focus antibody responses to a known protective epitope but are otherwise

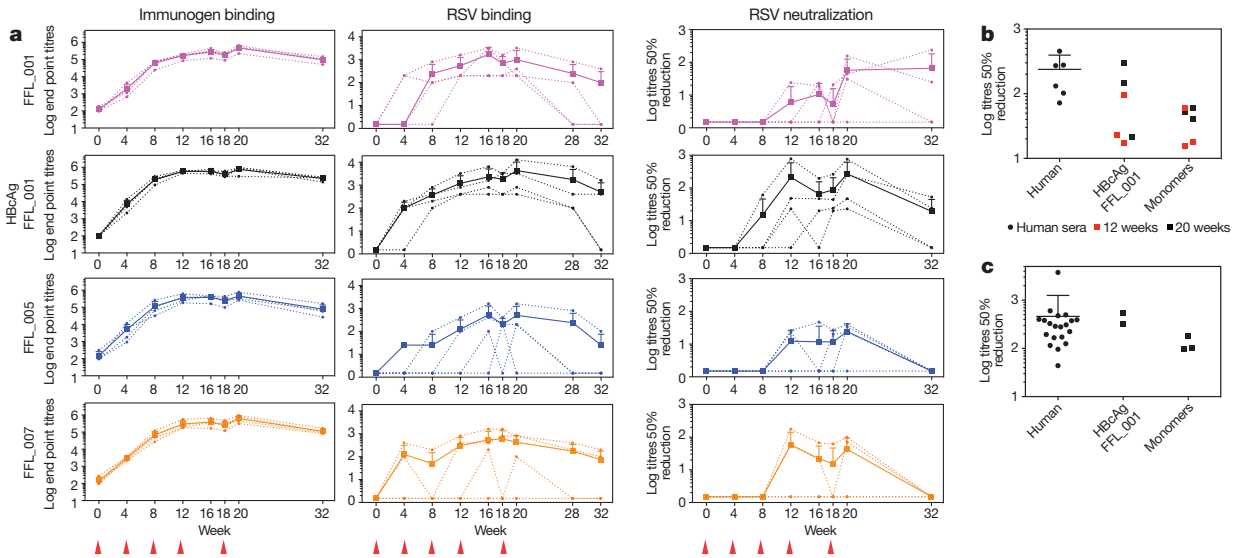


Figure 3 | Serological analysis of immunized macaques. **a**, ELISA end point titres measured against the autologous immunogen (left) or against RSV whole viral lysate (middle), and 50% neutralization titres as determined by the plaque reduction assay (right). The immunization groups are shown on the far left, and the schedule is indicated at the bottom. Small symbols connected with dashed lines indicate individual animals. Large symbols connected with solid lines report group averages, with error bars showing standard deviations, measured over the four animals in each group at each time point. **b**, Comparison of 50% neutralization titres for sera from six RSV-seropositive humans and sera from eight macaques from weeks 12 and 20, measured side by side in the plaque reduction assay. Mean ± standard deviation for the human data is 218 ± 145. Two macaque data points at both week 12 and week 20 are not visible in the graph because no neutralizing activity was detected. **c**, Comparison of 50% neutralization titres for sera from 20 RSV-seropositive humans and sera from five macaques from week 20, measured side by side in the flow cytometry assay. Mean ± standard deviation for the human data is 462 ± 792.

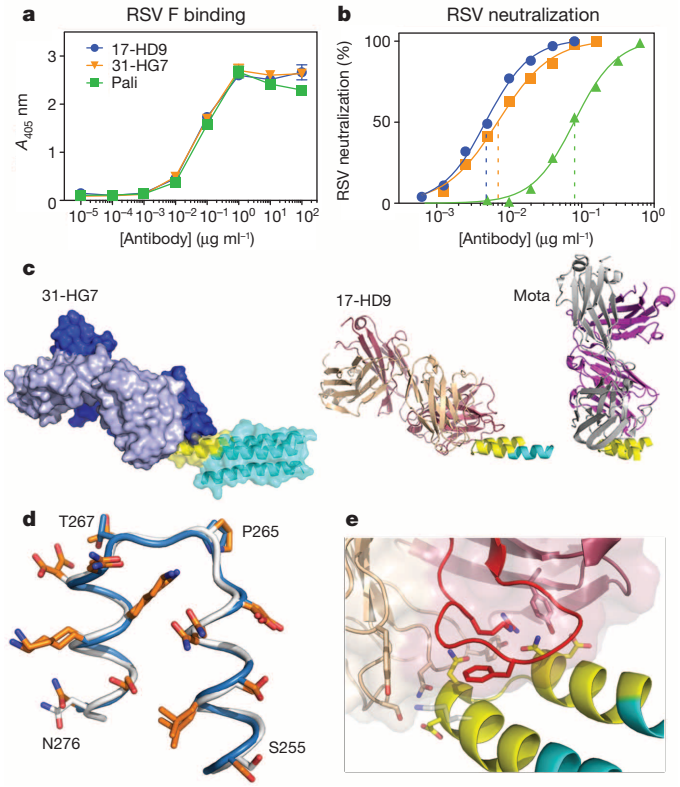


Figure 4 | Analysis of monoclonal antibodies isolated from an immunized macaque. **a**, Enzyme-linked immunosorbent assay (ELISA) binding of the macaque monoclonal antibodies (17-HD9 and 31-HG7) and pali to RSV F. **b**, Neutralization of RSV by the macaque monoclonal antibodies and pali, measured by a microneutralization assay. The half-maximum inhibitory concentrations (IC_{50}) for pali, 17-HD9 and 31-HG7 were 0.08, 0.005 and 0.007 μg ml^{-1}, respectively. **c**, Molecular replacement model of 31-HG7 bound to FFL_001 (left), a crystal structure of 17-HD9 bound to a 35-residue helix-turn-helix peptide from FFL_001 (middle) and the crystal structure of mota (PDB: 3IXT) bound to peptide. The three structures are aligned with respect to the helix-turn-helix epitope. **d**, Structural alignment of the helix-turn-helix epitopes bound to mota (blue) and 17-HD9 (white), in which side chains are coloured orange if at least 15% of the total area (backbone plus side chain) of that residue is buried by the respective antibody. Nine positions are buried by both antibodies, two positions in the turn are buried only by 17-HD9 (P265 and T267, RSV numbering), and two positions near the peptide termini are buried only by mota (S255 and N276). **e**, Close-up view of the interface between 17-HD9 and helix-turn-helix epitope. Interaction residues are shown in stick, and the complementary determining region H3 (CDRH3) is coloured red. K82/K272 (scaffold numbering/RSV numbering), at the edge of the interface, is coloured grey.

unrelated to RSV may have a lower safety barrier to testing in human infants than other non-replicating RSV vaccine approaches. The frequency, magnitude and durability of neutralizing responses to these scaffolds remain to be optimized, by varying such parameters as adjuvant, dose, schedule, particle display system and mode of delivery. In the context of vaccinating a diverse range of humans (or nonhuman primates) with different, time-dependent B-cell repertoires, neutralizing antibody responses to a single epitope may be more variable than responses to whole antigens or pathogens containing multiple neutralization epitopes. (Indeed, our discordant results in mice compared to macaques may reflect differences in species-dependent repertoires.) Thus, epitope-focused or scaffold-based vaccines for RSV or other pathogens may also be improved by inclusion of more than one epitope.

Features of the very potent neutralizing monoclonal antibodies isolated from one vaccinated macaque offer implications for vaccine design. These monoclonal antibodies had unusually high affinity for the eliciting antigen, for example when compared to monoclonal antibodies isolated from macaques that had been immunized with different, more conformationally labile antigens using similar regimens[28]. This suggests that rigid epitope structures may more efficiently induce extremely high-affinity antibodies, a possibility that merits further investigation. In cases of antigenically highly variable pathogens such as HIV, influenza or hepatitis C virus, the vaccine challenge is to induce responses to conserved but immunorecessive epitopes instead of the strain-specific epitopes that dominate the response to native antigens. Such conserved epitopes—the sites of vulnerability targeted by broadly neutralizing antibodies—are typically in close physical proximity to variable residues, making precision of immuno-focusing a vaccine requirement. Our crystallographic finding that scaffold-elicited monoclonal antibodies recapitulate the mota neutralization specificity with high precision provides proof of principle that epitope-focused vaccine design can meet this immuno-focusing challenge.

METHODS SUMMARY

Details of the FFL computational design protocol are provided in Methods. Protocols for protein expression and purification, biophysical characterization, virus-like particle preparation, X-ray crystallography, NMR, animal immunization, enzyme-linked immunosorbent assays, neutralization assays and monoclonal isolation are provided in Methods and Supplementary Information.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 July; accepted 19 December 2013.
Published online 5 February 2014.

1. Plotkin, S. A., Orenstein, W. A. & Offit, P. A. *Vaccines* 6th edn (Elsevier, 2012).
2. Rappuoli, R., Mandl, C. W., Black, S. & De Gregorio, E. Vaccines for the twenty-first century society. *Nature Rev. Immunol.* **11,** 865–872 (2011).
3. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380,** 2095–2128 (2012).
4. Murray, C. J. *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380,** 2197–2223 (2012).
5. Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* **361,** 2209–2220 (2009).
6. The RTS,S Clinical Trials Partnership. A phase 3 trial of RTS,S/AS01 malaria vaccine in African infants. *N. Engl. J. Med.* **367,** 2284–2295 (2012).
7. Tameris, M. D. *et al.* Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet* **381,** 1021–1028 (2013).
8. Osterholm, M. T., Kelley, N. S., Sommer, A. & Belongia, E. A. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *Lancet Infect. Dis.* **12,** 36–44 (2012).
9. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).
10. Anderson, L. J. *et al.* Strategic priorities for respiratory syncytial virus (RSV) vaccine development. *Vaccine* **31** (Suppl. 2), B209–B215 (2013).
11. U.S. Food and Drug Administration. *Vaccines, Blood & Biologics* http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/
12. Burton, D. R. Antibodies, viruses and vaccines. *Nature Rev. Immunol.* **2,** 706–713 (2002).
13. Burton, D. R., Poignard, P., Stanfield, R. L. & Wilson, I. A. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337,** 183–186 (2012).
14. Wu, H. *et al.* Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract. *J. Mol. Biol.* **368,** 652–665 (2007).
15. McLellan, J. S. *et al.* Structural basis of respiratory syncytial virus neutralization by motavizumab. *Nature Struct. Mol. Biol.* **17,** 248–250 (2010).
16. Correia, B. E. *et al.* Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure* **18,** 1116–1126 (2010).
17. Ofek, G. *et al.* Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl Acad. Sci. USA* **107,** 17880–17887 (2010).
18. McLellan, J. S. *et al.* Design and characterization of epitope-scaffold immunogens that present the motavizumab epitope from respiratory syncytial virus. *J. Mol. Biol.* **409,** 853–866 (2011).
19. Azoitei, M. L. *et al.* Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science* **334,** 373–376 (2011).
20. Azoitei, M. L. *et al.* Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J. Mol. Biol.* **415,** 175–192 (2012).
21. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364–1368 (2003).

22. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491,** 222–227 (2012).
23. Correia, B. E. *et al.* Computational protein design using flexible backbone remodeling and resurfacing: case studies in structure-based antigen design. *J. Mol. Biol.* **405,** 284–297 (2011).
24. Zhu, Q. *et al.* Analysis of respiratory syncytial virus preclinical and clinical variants resistant to neutralization by monoclonal antibodies palivizumab and/or motavizumab. *J. Infect. Dis.* **203,** 674–682 (2011).
25. Clarke, B. E. *et al.* Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. *Nature* **330,** 381–384 (1987).
26. Jegerlehner, A. *et al.* A molecular assembly system that renders antigens of choice highly repetitive for induction of protective B cell responses. *Vaccine* **20,** 3104–3112 (2002).
27. Jafri, H. S., Wu, X., Makari, D. & Henrickson, K. J. Distribution of respiratory syncytial virus subtypes A and B among infants presenting to the emergency department with lower respiratory tract infection or apnea. *Pediatr. Infect. Dis. J.* **32,** 335–340 (2013).
28. Sundling, C. *et al.* High-resolution definition of vaccine-elicited B cell responses against the HIV primary receptor binding site. *Sci. Transl. Med.* **4,** 142ra196 (2012).
29. Wen, X. *et al.* Structure of the human metapneumovirus fusion protein with neutralizing antibody identifies a pneumovirus antigenic site. *Nature Struct. Mol. Biol.* **19,** 461–463 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Menis, D. Kulp and D. Burton for comments on the manuscript, Y.-E. Ban, D. Alonso and K. E. Laidig for computing assistance, E. Gribben and R. Carnahan for assistance with mouse immunizations and C. Slaughter for assistance in statistical analysis. Adjuplex adjuvant was a gift from Advanced BioAdjuvants. The University of Washington has filed patents relating to immunogens in this manuscript. Materials and information will be provided under Materials Transfer Agreement (MTA). Support for this work was provided by Fundação para a Ciência e a Tecnologia fellowship SFRH/BD/32958/2006 (B.E.C.), National Institutes of Health NRSA Training Grant fellowship T32CA080416 (J.G.J.), The Children's Hospital of Philadelphia (P.R.J.), a Bill and Melinda Gates Foundation CAVD award (W.R.S., R.K.S. and D.B.), the International AIDS Vaccine Initiative Neutralizing Antibody Consortium (W.R.S. and D.B.), the International AIDS Vaccine Initiative Neutralizing Antibody Center (W.R.S. and R.T.W.), a grant from the March of Dimes (J.E.C.), National Institutes of Health grants 2T32 GM007270 (V.V.) and U54 AI 005714 (R.E.K.), National Institute of Allergy and Infectious Diseases grants P01AI094419 (W.R.S. and R.K.S.), 5R21AI088554 (W.R.S.), 1UM1AI100663 (W.R.S. and R.T.W.), 1R01AI102766-01A1 (Y.L. and R.T.W.), P30AI36214 (from the Center for AIDS Research, University of California, San Diego, to Y.L.), and the National Institute of Allergy and Infectious Diseases Intramural Research Program (B.S.G.). This is manuscript 26069 from The Scripps Research Institute.

Author Contributions B.E.C., J.T.B., R.J.L., R.E.K., B.S.G., R.T.W., D.B., R.K.S., J.E.C, P.R.J. and W.R.S. designed research. B.E.C. wrote the code for FFL, designed proteins and performed biophysical characterization. J.T.B., R.J.L., M.C. and M.J.C. performed serological analysis. G.B. and E.S. prepared and characterized particle immunogens. C. Carrico, J.G.J., P.R., C. Correnti and M.A.H. performed X-ray crystallography. O.K. performed biophysical characterization. V.V. performed NMR studies. M.J.C. performed NHP immunizations. O.K., A.S., S.M., A.M.S., Y.A. and E.S. performed protein expression and purification. Y.L. performed B-cell sorting and RT–PCR. B.E.C. and W.R.S. wrote, and all co-authors edited, the manuscript.

Author Information Coordinates and structure factors for FFL_005, FFL_001–mota, and FFL_001–17-HD9 structures have been deposited in the Protein Data Bank with accession codes 4L8I, 4JLR and 4N9G, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.R.S. (schief@scripps.edu).

LETTER

doi:10.1038/nature13031

Reflection from the strong gravity regime in a lensed quasar at redshift $z = 0.658$

R. C. Reis[1], M. T. Reynolds[1], J. M. Miller[1] & D. J. Walton[2]

The co-evolution of a supermassive black hole with its host galaxy[1] through cosmic time is encoded in its spin[2–4]. At $z > 2$, supermassive black holes are thought to grow mostly by merger-driven accretion leading to high spin. It is not known, however, whether below $z \approx 1$ these black holes continue to grow by coherent accretion or in a chaotic manner[5], though clear differences are predicted[3,4] in their spin evolution. An established method[6] of measuring the spin of black holes is through the study of relativistic reflection features[7] from the inner accretion disk. Owing to their greater distances from Earth, there has hitherto been no significant detection of relativistic reflection features in a moderate-redshift quasar. Here we report an analysis of archival X-ray data together with a deep observation of a gravitationally lensed quasar at $z = 0.658$. The emission originates within three or fewer gravitational radii from the black hole, implying a spin parameter (a measure of how fast the black hole is rotating) of $a = 0.87^{+0.08}_{-0.15}$ at the 3σ confidence level and $a > 0.66$ at the 5σ level. The high spin found here is indicative of growth by coherent accretion for this black hole, and suggests that black-hole growth at $0.5 \leq z \leq 1$ occurs principally by coherent rather than chaotic accretion episodes.

When optically thick material, for example, an accretion disk, is irradiated by hard X-rays, some of the flux is reprocessed into an additional 'reflected' emission component, which contains both continuum emission and atomic features. The most prominent signature of reflection from the inner accretion disk is typically the relativistic iron Kα line (6.4–6.97 keV; rest frame)[8] and the Compton reflection hump[7], often peaking at 20–30 keV (rest frame). However, the deep gravitational potential and strong Doppler shifts associated with regions around black holes will also cause the forest of soft X-ray emission lines in the approximately 0.7–2.0-keV range to be blended into a smooth emission feature, providing a natural explanation for the 'soft excess' observed in the X-ray spectra of nearby active galactic nuclei (AGNs)[9,10]. Indeed, both the iron (Fe) line and the soft excess can be used to provide insight into the nature of the central black hole and to measure its spin[10]. Previous studies have revealed the presence of a soft excess in over 90% of quasars at[11,12] around $z < 1.7$, and broad Fe-lines are also seen in $\lesssim 25\%$ of these objects[11,13], suggesting that reflection is also prevalent in these distant AGNs. However, owing to the inadequate signal-to-noise ratio resulting from their greater distances, the X-ray spectra of these quasars were necessarily modelled using simple phenomenological parameterizations[12,14].

Gravitational lensing offers a rare opportunity to study the innermost relativistic region in distant quasars[15,16]; the 'lens' galaxy acts as a natural telescope, magnifying the light from the faraway quasar. Quasars located at $0.5 \lesssim z \lesssim 1$ are considerably more powerful than local Seyfert galaxies (relatively nearby active galaxies) and are known to be a major contributor to the cosmic X-ray background[17], making them objects of particular cosmological importance. 1RXS J113151.6−123158 (hereafter RX J1131−1231) is a quadruply imaged quasar at redshift $z = 0.658$ hosting a supermassive black hole ($M_{\rm bh} \approx 2 \times 10^8 M_{\odot}$, where $M_{\odot}$ is the mass of the Sun)[18] gravitationally lensed by an elliptical galaxy at $z = 0.295$ (ref. 19). X-ray and optical observations exhibited an intriguing flux-ratio variability between the lensed images, which was subsequently revealed to be due to significant gravitational micro-lensing by stars in the lensing galaxy[15,16].

Taking advantage of gravitational micro-lensing techniques, augmented by substantial monitoring with the Chandra X-ray observatory, a tight limit of the order of about ten gravitational radii ($r_{\rm g} = GM/c^2$, where G is the gravitational constant, M is the mass of the black hole and c is the speed of light in vacuum) was placed[20] on the maximum size of the X-ray emitting region in RX J1131−1231, indicative of a highly compact[21] source of emission (less than about three billion kilometres, or 20 astronomical units, AU). The lensed nature of this quasar provides an excellent opportunity to study the innermost regions around a black hole at a cosmological distance (the look-back time for RX J1131−1231 is about 6.1 billion years), and with this aim Chandra and XMM-Newton have accumulated nearly 500,000 s of data on RX J1131−1231.

Starting with the Chandra data, fits with a model consisting of a simple absorbed power-law continuum, both to the data from individual lensed images (Extended Data Fig. 1) and to the co-added data (Extended Data Figs 2, 3 and 4), reveal broad residual emission features at low energies ($\lesssim$2-keV rest frame, the soft excess) and also around the Fe K energies (3.5–7-keV rest frame), which are characteristic signatures of relativistic disk reflection[10,22]. To treat these residuals, we consider two template models based on those commonly used to fit the spectra of Seyfert galaxies[6,23] and stellar-mass black-hole binaries[24], and which have also at times been used to model local quasars[25]. The first is a simple phenomenological combination of a power-law, a soft-thermal disk and a relativistic Fe-line component ('baseline-simple'), and the second

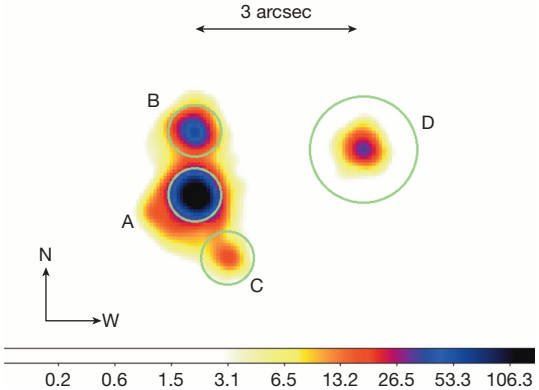


Figure 1 | Chandra image of RX J1131−1231. This representative image of a single epoch was made using subpixel techniques in the 0.3–8-keV energy range (Supplementary Information) and is shown here smoothed with a Gaussian ($\sigma = 0.25''$). The green circles show the source extraction regions. For images A–C, we used a radius of 0.492″, whereas the source region for image D was set to 0.984″. Individual source and background regions were made for all 30 observations and spectra were extracted from the unsmoothed images. The colour (logarithmic) scale reflects the number of counts in each particular pixel ranging from 0 to a maximum of 212 counts.

[1]Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA. [2]Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, California 91125, USA.

employs a self-consistent blurred-reflection model together with a power law ('baseline-reflection'). In addition, both models include two neutral absorbers; the first to account for possible intrinsic absorption at the redshift of the quasar; and the second to account for Galactic absorption.

We first statistically confirm the presence of reflection features in RX J1131-1231 using the baseline-simple model. Least-squares fits were made to all the individual Chandra spectra of image B (the lens galaxy creates four images, which we label A–D; see Fig. 1) simultaneously, allowing only the normalizations of the various components and the power-law indices to vary (Extended Data Fig. 1). The thermal component, used here as a proxy for the soft excess, is required at $>10\sigma$ and a Fisher's F-test indicates that the addition of a relativistic emission line to the combined Chandra data of image-B is significant at greater than the 99.9% confidence level. Tighter constraints ($>5\sigma$) on the significance of the relativistic Fe line can be obtained by co-adding all Chandra data to form a single, time-averaged spectrum representative of the average behaviour of the system (Fig. 2 and Extended Data Fig. 3).

The XMM-Newton observation also shows the clear presence of a soft excess below about 1.2 keV, again significant at $>5\sigma$, and thanks to its high effective area above approximately 5 keV, it also displays the presence of a hardening (more high-energy photons) at high energies (Extended Data Fig. 6). An F-test indicates that a break in the power law at around 5 keV (Fig. 2) is significant at the 3.6σ level of confidence. This hardening is consistent with the expectation of a reflection spectrum and can be characterized with the Compton reflection hump[7].

The unprecedented data quality for this moderate-z quasar (about 100,000 counts in the 0.3–8 keV energy range from each of the Chandra and XMM-Newton data sets) enables us to apply physically motivated, self-consistent models for the reflection features. We proceed by using the baseline-reflection model to estimate the spin parameter through a variety of analyses, including time-resolved and time-averaged analyses of individual Chandra images, using its superior angular resolution, and through analysis of the average spectrum obtained from all four lensed images with XMM-Newton. During the time-resolved analysis, the black-hole spin parameter as well as the disk inclination and emissivity profile were kept constant from epoch to epoch, whereas the normalizations of the reflection and power-law components, as well as the ionization state of the disk and the power-law indices, were allowed to vary between epochs (see the Supplementary Information for further details). In all cases, we obtain consistent estimates for the black-hole spin, which imply that RX J1131−1231 hosts a rapidly rotating black hole (Extended Data Fig. 5 and Extended Data Tables 1 and 2). Finally, to optimize the signal-to-noise ratio and obtain the best estimate of the spin parameter we fit the combined Chandra and XMM-Newton data of RX J1131-1231 simultaneously with the baseline-reflection model and find $a=0.87^{+0.08}_{-0.15}\ Jc/GM^2$ (where J is the angular momentum) at the 3σ level of confidence (and $a>0.66\ Jc/GM^2$ at the 5σ confidence level; Fig. 3).

The tight constraint on the spin of the black hole in this gravitationally lensed quasar represents a robust measurement of black-hole spin beyond our local Universe. The compact nature of the X-ray corona returned by the relativistic reflection model used herein confirms the prior micro-lensing analysis[15,16], and hence moves the basic picture of X-ray emission in quasars away from large X-ray coronae[26] that may blanket at least the inner disk, and more towards a compact emitting region in the very innermost parts of the accretion flow, consistent with models for the base of a jet[27].

In addition to constraining the immediate environment and spin of the black hole, the analysis presented here has implications for the nature of the cosmic X-ray background. The best-fitting baseline-reflection model to the time-averaged Chandra and XMM-Newton spectra (Extended Data Figs 3 and 6) suggest that the source is at times reflection-dominated, that is, we find the ratio of the reflected to the illuminating continuum in the Chandra and XMM-Newton data to be $f_{\rm reflect}/f_{\rm illum}=2.3\pm1.2$

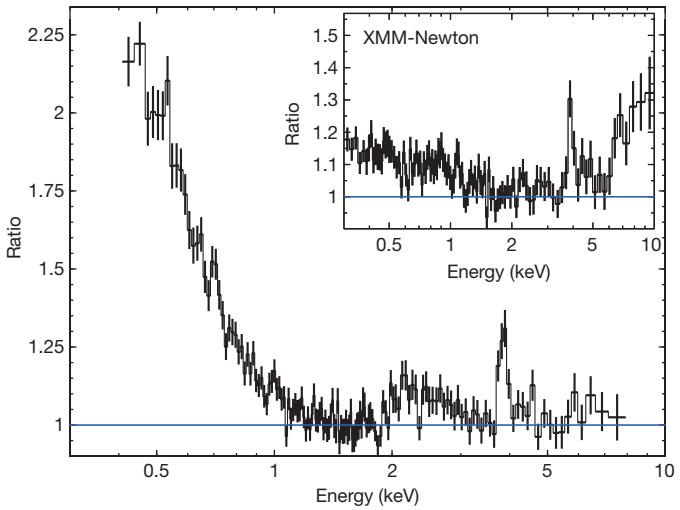


Figure 2 | Broad Fe line, soft and hard excess in RX J1131−1231. The figure shows the co-added Chandra data over all epochs for all four lensed images. The blue line is a guide to the eye of a perfect model fit to the data. The data was fitted in a phenomenological manner with a model consisting of an absorbed power law with an index of $\Gamma=1.60\pm0.04$ for the continuum, a thermal disk component with a temperature of 0.19 ± 0.02 keV to account for the soft excess, and a broad relativistic line with energy constrained to lie between 6.4 keV and 6.97 keV (rest-frame is the baseline-simple model). The ratio (of the data to the model) is shown after setting the normalization of the disk, the relativistic line and the narrow line component to zero, in order to highlight these features. The inset shows the XMM-Newton data fitted with a $\Gamma=1.83$ power law. The best-fitting, phenomenological model for the XMM-Newton data requires the presence of a soft excess, which can again be characterized by a thermal disk component with a temperature of 0.22 ± 0.03 keV, a power law with an index $=1.83^{+0.07}_{-0.03}$ up to a break at $E_{\rm break}=5.5^{+0.5}_{-2.2}$ keV, at which point it hardens to $\Gamma=1.28^{+0.33}_{-0.19}$. This hardening is interpreted as the Compton reflection hump. Both co-added spectra shown here probe the time-averaged behaviour of RX J1131–1231. Quoted errors refer to the 90% confidence limit and the error bars are 1σ.

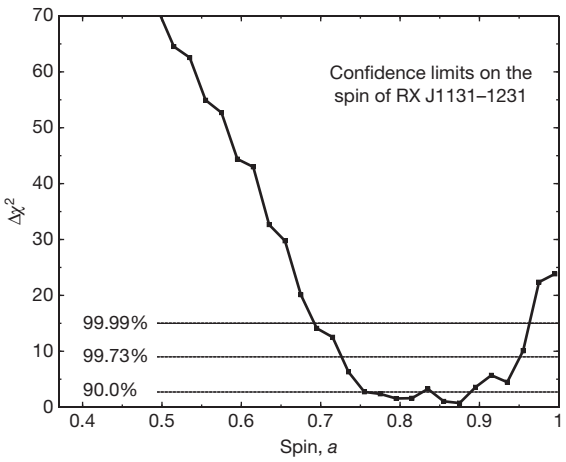


Figure 3 | Goodness of fit versus the spin parameter of the supermassive black hole in RX J1131−1231. Fits were made with the spin parameter a varying from 0.495 to 0.995 in steps of 0.02 with all parameters of the model allowed to vary. The contour was generated by adopting a model consisting of a power law together with a relativistic blurred reflection by an accretion disk, as well as two neutral absorbers: one at the redshift of the quasar and another local to our Galaxy (the baseline-reflection model). The fit was made to the co-added spectra from both XMM-Newton and Chandra simultaneously, with the assumption that the spin of the black hole, the inclination of the accretion disk and the total hydrogen in our line of sight does not change between observations. The dotted lines show the 99.99%, 99.73% (3σ) and 90% confidence limits where it becomes clear that the supermassive black hole in RX J1131−1231 must be rotating with a spin $a=0.87^{+0.08}_{-0.15}\ Jc/GM^2$ at the 3σ level of confidence.

and 0.47 ± 0.15, respectively, in the 0.1–10-keV band (local frame; Extended Data Table 2). However, it must be noted that uncertainties in the size of the microlensed regions could affect the absolute value of this ratio. Nonetheless, this analysis clearly demonstrates the presence of a significant contribution from a reflection component to the X-ray spectrum of this $z = 0.658$ quasar. The properties of RX J1131−1231 are consistent[11–14,16] with the known observational characteristics of quasars at $0.5 \lesssim z \lesssim 1$, and our results suggest that the relativistic reflection component from the large population of unobscured quasars expected in this epoch[17] could significantly contribute in the 20–30-keV band of the cosmic X-ray background.

Although questions have previously been raised over whether reflection is a unique interpretation for the features observed in AGNs, the amassed evidence points towards this theoretical framework[23,28], and reached culmination with the launch of NuSTAR (the Nuclear Spectroscopic Telescope Array is an Explorer mission orbiting Earth that will allow us to study the Universe with high energy X-rays) in June 2012 and the strong confirmation of relativistic disk reflection from a rapidly spinning supermassive black hole at the centre of the nearby galaxy NGC 1365 (ref. 6). Nonetheless, there still remain possible systematic uncertainties, for example, owing to the intrinsic assumption that the disk truncates at the innermost stable circular orbit. Simulations have been performed that are specifically aimed at addressing the robustness of this assumption[29], which found that emission within this radius is negligible, especially for rapidly rotating black holes, as is the case here.

The ability to measure cosmological black-hole spin brings with it the potential to study directly the co-evolution of the black hole and its host galaxy[1]. The ultimate goal is to measure the spin in a sample of quasars as a function of redshift and to use the spin distribution as a window on the history of the co-evolution of black holes and galaxies[4]. Our measurement of the spin in RX J1131−1231 is a step along that path, and introduces a possible way to begin assembling a sample of supermassive black-hole spins at moderate redshift with current X-ray observatories.

METHODS SUMMARY

We produced images for all 30 individual Chandra pointings (Fig. 1; see Methods for details), and spectra were extracted over the 0.3–8.0-keV energy band for each of the four lensed images in each observation (all energies are quoted in the observed frame unless stated otherwise). Previous studies[16] have demonstrated that certain lensed images/epochs might suffer from a moderate level of pile-up[30]. We therefore exclude spectra that displayed any significant level of pile-up in all further analysis (see Methods for details and Extended Data Figs 7 and 8). The remaining spectra sample a period of approximately 8 years, which allows for both a time-resolved and time-averaged analysis of RX J1131-1231. We also analyse a deep XMM-Newton observation taken in July 2013, which provides an average spectrum of the four lensed images over the 0.3–10.0-keV energy range.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 March 2013; accepted 10 January 2014.
Published online 5 March 2014.

1. Gebhardt, K. *et al.* A relationship between nuclear black hole mass and galaxy velocity dispersion. *Astrophys. J.* **539,** L13–L16 (2000).
2. Berti, E. & Volonteri, M. Cosmological black hole spin evolution by mergers and accretion. *Astrophys. J.* **684,** 822–828 (2008).
3. Fanidakis, N. *et al.* Grand unification of AGN activity in the ΛCDM cosmology. *Mon. Not. R. Astron. Soc.* **410,** 53–74 (2011).
4. Volonteri, M., Sikora, M., Lasota, J.-P. & Merloni, A. The evolution of active galactic nuclei and their spins. *Astrophys. J.* **775,** 94 (2013).
5. King, A. R. & Pringle, J. E. Growing supermassive black holes by chaotic accretion. *Mon. Not. R. Astron. Soc.* **373,** L90–L92 (2006).
6. Risaliti, G. *et al.* A rapidly spinning supermassive black hole at the centre of NGC1365. *Nature* **494,** 449–451 (2013).
7. Ross, R. R. & Fabian, A. C. The effects of photoionization on X-ray reflection spectra in active galactic nuclei. *Mon. Not. R. Astron. Soc.* **261,** 74–82 (1993).
8. Tanaka, Y. *et al.* Gravitationally redshifted emission implying an accretion disk and massive black hole in the active galaxy MCG-6-30-15. *Nature* **375,** 659–661 (1995).
9. Crummy, J., Fabian, A. C., Gallo, L. & Ross, R. R. An explanation for the soft X-ray excess in active galactic nuclei. *Mon. Not. R. Astron. Soc.* **365,** 1067–1081 (2006).
10. Walton, D. J., Nardini, E., Fabian, A. C., Gallo, L. C. & Reis, R. C. Suzaku observations of 'bare' active galactic nuclei. *Mon. Not. R. Astron. Soc.* **428,** 2901–2920 (2013).
11. Porquet, D., Reeves, J. N., O'Brien, P. & Brinkmann, W. XMM-Newton EPIC observations of 21 low-redshift PG quasars. *Astron. Astrophys.* **422,** 85–95 (2004).
12. Piconcelli, E. *et al.* The XMM-Newton view of PG quasars. I. X-ray continuum and absorption. *Astron. Astrophys.* **432,** 15–30 (2005).
13. Jiménez-Bailón, E. *et al.* The XMM-Newton view of PG quasars. II. Properties of the Fe K_α line. *Astron. Astrophys.* **435,** 449–457 (2005).
14. Green, P. J. *et al.* A full year's Chandra exposure on Sloan Digital Sky Survey quasars from the Chandra Multiwavelength Project. *Astrophys. J.* **690,** 644–669 (2009).
15. Pooley, D., Blackburne, J. A., Rappaport, S. & Schechter, P. L. X-ray and optical flux ratio anomalies in quadruply lensed quasars. I. Zooming in on quasar emission regions. *Astrophys. J.* **661,** 19–29 (2007).
16. Chartas, G. *et al.* Revealing the structure of an accretion disk through energy-dependent X-ray microlensing. *Astrophys. J.* **757,** 137 (2012).
17. Gilli, R., Comastri, A. & Hasinger, G. The synthesis of the cosmic X-ray background in the Chandra and XMM-Newton era. *Astron. Astrophys.* **463,** 79–96 (2007).
18. Sluse, D., Hutsemékers, D., Courbin, F., Meylan, G. & Wambsganss, J. Microlensing of the broad line region in 17 lensed quasars. *Astron. Astrophys.* **544,** A62 (2012).
19. Sluse, D. *et al.* A quadruply imaged quasar with an optical Einstein ring candidate: 1RXS J113155.4–123155. *Astron. Astrophys.* **406,** L43–L46 (2003).
20. Dai, X. *et al.* The sizes of the X-ray and optical emission regions of RXJ 1131–1231. *Astrophys. J.* **709,** 278–285 (2010).
21. Reis, R. C. & Miller, J. M. On the size and location of the X-ray emitting coronae around black holes. *Astrophys. J. Lett.* **769,** 7 (2013).
22. Reynolds, C. S. Measuring black hole spin using x-ray reflection spectroscopy. *Space Sci. Rev.* http://dx.doi.org/10.1007/s11214-013-0006-6 (published online 31 August 2013).
23. Fabian, A. C. *et al.* Broad line emission from iron K- and L-shell transitions in the active galaxy 1H0707–495. *Nature* **459,** 540–542 (2009).
24. Miller, J. M. Relativistic X-ray lines from the inner accretion disks around black holes. *Annu. Rev. Astron. Astrophys.* **45,** 441–479 (2007).
25. Schmoll, S. *et al.* Constraining the spin of the black hole in Fairall 9 with Suzaku. *Astrophys. J.* **703,** 2171–2176 (2009).
26. Haardt, F. & Maraschi, L. A two-phase model for the X-ray emission from Seyfert galaxies. *Astrophys. J.* **380,** L51–L54 (1991).
27. Falcke, H. & Markoff, S. The jet model for Sgr A*: radio and X-ray spectrum. *Astron. Astrophys.* **362,** 113–118 (2000).
28. Walton, D. J., Reis, R. C., Cackett, E. M., Fabian, A. C. & Miller, J. M. The similarity of broad iron lines in X-ray binaries and active galactic nuclei. *Mon. Not. R. Astron. Soc.* **422,** 2510–2531 (2012).
29. Reynolds, C. S. & Fabian, A. C. Broad iron-Kα emission lines as a diagnostic of black hole spin. *Astrophys. J.* **675,** 1048–1056 (2008).
30. Miller, J. M. *et al.* On relativistic disk spectroscopy in compact objects with X-ray CCD cameras. *Astrophys. J.* **724,** 1441–1455 (2010); http://dx.doi.org/10.1088/0004-637X/724/2/1441.

Supplementary Information is available in the online version of the paper.

Acknowledgements R.C.R. thanks the Michigan Society of Fellows and NASA for support through the Einstein Fellowship Program, grant number PF1-120087. We thank the ESA XMM-Newton Project Scientist N. Schartel and the XMM-Newton planning team for carrying out the DDT observation. The scientific results reported in this article are based on data obtained from the Chandra Data Archive.

Author Contributions R.C.R. performed the data reduction and analysis of all the Chandra data reported here. The XMM-Newton data was reduced by both R.C.R and M.T.R. The pile-up study was carried out by R.C.R., J.M.M. and M.T.R. The text was composed, and the paper synthesised, by R.C.R., with help from D.J.W. and M.T.R. The smoothed subpixel images were made by R.C.R. and M.T.R. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.C.R. (rdosreis@umich.edu).

LETTER

doi:10.1038/nature13019

Interrogating selectivity in catalysis using molecular vibrations

Anat Milo[1], Elizabeth N. Bess[1] & Matthew S. Sigman[1]

The delineation of molecular properties that underlie reactivity and selectivity is at the core of physical organic chemistry[1–5], and this knowledge can be used to inform the design of improved synthetic methods or identify new chemical transformations[6–9]. For this reason, the mathematical representation of properties affecting reactivity and selectivity trends, that is, molecular parameters, is paramount. Correlations produced by equating these molecular parameters with experimental outcomes are often defined as free-energy relationships and can be used to evaluate the origin of selectivity and to generate new, experimentally testable hypotheses[6,10–12]. The premise behind successful correlations of this type is that a systematically perturbed molecular property affects a transition-state interaction between the catalyst, substrate and any reaction components involved in the determination of selectivity[10,11]. Classic physical organic molecular descriptors, such as Hammett[4], Taft[3] or Charton[5] parameters, seek to independently probe isolated electronic or steric effects[3–6,13,14]. However, these parameters cannot address simultaneous, non-additive variations to more than one molecular property, which limits their utility. Here we report a parameter system based on the vibrational response of a molecule to infrared radiation that can be used to mathematically model and predict selectivity trends for reactions with interlinked steric and electronic effects at positions of interest. The disclosed parameter system is mechanistically derived and should find broad use in the study of chemical and biological systems.

A molecule's structural features are embodied in its unique vibration modes, which invoke core, inherent bond force constants and atomic masses[7,15,16]. Thus, to interrogate intricate selectivity trends, we turned to vibrational energies. Kinetic isotope effects, a common reaction probe in biology and chemistry, further highlight the usefulness of relative vibrational energies for assessing mechanistic hypotheses[7,17] (Fig. 1b). With this underpinning, we postulated that infrared vibrations could serve as mechanistically meaningful molecular descriptors in the study of catalytic reaction selectivity, allowing for correlations akin to free-energy relationships[1] (Fig. 1c; see Methods Summary for details).

Here we describe three case studies that substantiate the prospect of using molecular vibrations to predict and elucidate selectivity trends. We first consider the desymmetrization of bisphenols. In our recent reports of modelling catalytic systems, the catalyst or substrate libraries, or both, were specifically designed to avoid the complexity of integrated steric and electronic effects[9,14,18]. As an example, we studied the role of substrate steric effects in the peptide-catalysed desymmetrization of bisphenols[14,19] (Fig. 2a). A specific finding was that Verloop's Sterimol parameters B_1 (minimal radius) and L (length, Fig. 1a) could be used to successfully correlate enantioselectivity to the steric impact of the substituent **R** on the bisphenol[2]. Because this study was directed towards evaluation of steric effects, obvious electronic changes were initially avoided.

To determine whether the perturbation to enantioselectivity in this reaction is purely steric in origin, we evaluated a substrate containing $-CCl_3$, which is a $-CMe_3$ (where Me is methyl) homologue according to its Sterimol values. The $-CCl_3$-containing substrate yielded a much lower enantioselectivity than would be expected from purely steric considerations, suggesting that substituent electronic effects may affect the enantioselective outcome. To explore this outlier in the steric trend further, a model was developed on the basis of Sterimol values from an eight-membered training set of the original substrates (Fig. 2b). External validation for this model using four of the original sterically perturbed substrates and six new substrates containing concurrent steric and electronic modulation reveals a poor correlation (slope, 0.82; intercept, 0.18; $R^2 = 0.60$). The poorest performers were substrates with multifaceted structural features that cannot be accounted for by the parameter selection. The steric model's shortcomings provided an impetus to explore infrared vibrations as a reaction interrogation technique.

Thus, we chose mechanistically relevant infrared frequencies to produce molecular descriptors for the prediction of enantioselectivity[16].

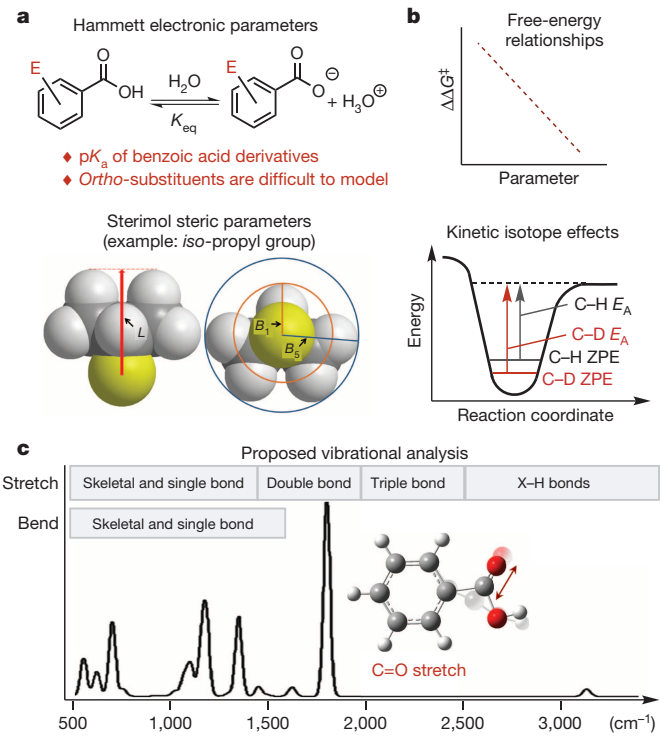
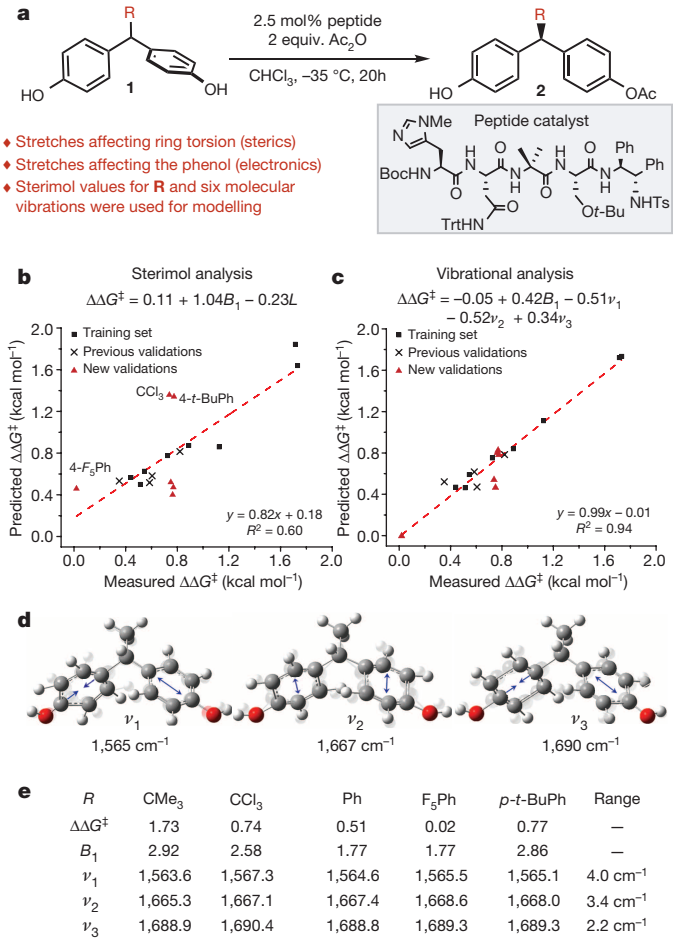


Figure 1 | Approaches to interrogating reaction mechanisms. a, Parameters used for free-energy correlations: Hammett electronic parameters based on the dissociation constant of substituted benzoic acids (top). Steric Sterimol parameters: substituent length (L), and minimal (B_1) and maximal (B_5) widths perpendicular to the length (bottom). **b**, Energetic considerations in free-energy relationships (top) and in isotopologue vibrational energies (bottom). $\Delta\Delta G^{\ddagger}$ is the difference in Gibbs free energy between the two selectivity-determining transition states, E_A is the activation energy barrier and ZPE is the zero-point energy. **c**, Simulated infrared spectrum of benzoic acid and functional group spectral ranges.

[1]Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112, USA.



e

R	CMe_3	CCl_3	Ph	F_5Ph	*p*-*t*-BuPh	Range
ΔΔG‡	1.73	0.74	0.51	0.02	0.77	—
B_1	2.92	2.58	1.77	1.77	2.86	—
ν_1	1,563.6	1,567.3	1,564.6	1,565.5	1,565.1	4.0 cm^{-1}
ν_2	1,665.3	1,667.1	1,667.4	1,668.6	1,668.0	3.4 cm^{-1}
ν_3	1,688.9	1,690.4	1,688.8	1,689.3	1,689.3	2.2 cm^{-1}

Figure 2 | Using infrared vibrations and Sterimol values to correlate enantioselectivity. a, Reaction scheme for the desymmetrization of bisphenols. Reactant torsion is proposed to have a role in the mechanism. **b**, Correlation between Sterimol values and enantioselectivity (normalized model), including the training set, previous sterically modulated validations, and new simultaneously sterically and electronically modulated validations. **c**, Correlation between Sterimol B_1 (minimal width), vibrations and enantioselectivity (normalized model). **d**, Illustration of the vibrational frequencies used for the correlation of enantioselectivity: ν_1, antisymmetric ring stretch with secondary C–H bends and a C–O stretch; ν_2, symmetric ring stretch with a secondary O–H bend; ν_3, antisymmetric ring stretch with secondary C–C bends and a C–O stretch. **e**, Parameter values for the steric model outliers (CCl_3, F_5Ph and *p*-*t*-BuPh) and for two sterically homologous but electronically divergent substituents (CMe_3 and Ph).

Because the steric and electronic features of **R** modulate enantioselectivity, we speculated that stretches of the bisphenol ring would be sensitive to enantioselectivity trends. These stretches are influenced by the mass and charge of **R**, incorporate various secondary C–H and C–C bends and affect the O–H group. Thus, the frequencies selected for modelling were six computationally derived, distinct sp^2 C=C stretches in the 1,700–1,500 cm^{-1} spectral region, which involve either one or both rings[20–24] (Methods and Supplementary Information). The most predictive, statistically significant model developed (slope, 0.99; intercept, –0.01; $R^2 = 0.94$) contains four parameters (Fig. 2c). These include Sterimol parameter B_1, which describes the minimal radius of **R**, and three computationally derived infrared frequencies (Fig. 2c, d). The derived model is highly predictive for both isolated steric effects and substrates containing concomitant steric and electronic changes.

The extreme outliers in the Sterimol analysis are **R** = $-CCl_3$, $-F_5Ph$ and –4*t*-Bu-Ph (where *t*-Bu is *tert*-butyl) (Fig. 2b). If we consider the Sterimol parameters as descriptors of repulsive steric interactions within the catalytic reaction site, a geometry-based parameter would not suffice to define the first two substituents. Because vibrational analysis takes charge distribution into account, and adds a directional aspect to steric interactions, it is able to address electronically diverse substituents. These results are consistent with a mechanistic hypothesis asserting a direct interaction between the peptide catalyst and the bisphenol substituent at the selectivity-determining step of the reaction[14,19]; such an interaction would be sensitive to substituent electronegativity. Inspecting the model reveals that in the cases of **R** = $-CCl_3$ and $-F_5Ph$, all three vibrations are shifted to a higher frequency relative to their respective non-halogenated steric homologues, $-CMe_3$ and –Ph (Fig. 2e). This generalized trend could indicate that the vibrational parameters are functioning as an electronic correction to the steric description.

It has been proposed that the minimum radius B_1 (Fig. 1a) describes steric effects proximal to the phenol rings, wherein the substituent applies torsion on the rings in a propeller-like strain[14] (Fig. 2a). However, the inability to predict the enantioselectivity of –4*t*-Bu-Ph using Sterimol analysis points to a limitation in applying this parameter to substituted aromatic rings. Sterimol values define this group similarly to a $-CMe_3$ group in terms of B_1. However, on the basis of both the distal location of the $-CMe_3$ group in –4*t*-Bu-Ph and the empirically observed enantioselectivity (enantiomeric ratio of 82:18 for –4*t*-Bu-Ph versus 97.5:2.5 for $-CMe_3$), this substituent more closely resembles an unsubstituted phenyl ring (enantiomeric ratio, 75:25). This significantly restricts the use of Sterimol values for evaluating steric effects in aromatic systems. Because infrared ring-stretching vibrations are modulated in response to substituent steric effects, they represent an auxiliary, directional aspect of the substituent geometry and allow for the prediction of groups such as –4*t*-Bu-Ph.

To explore the potential of infrared vibrational analysis further, as a second case study we evaluated our recently reported enantioselective hydrogenation of 1,1-diarylalkenes[25] (Fig. 3a). The original scope included 1,1-diarylalkenes in which ring substitution patterns were modulated to explore the origin of enantioselectivity (Fig. 3d). An unusual observation was that 3,5-dimethoxy substitutions were required to achieve high enantioselectivity (Fig. 3b). We began examining the origin of this observation using our new technology by including twelve of the originally reported substrates in the training set. This set was selected on the basis of both the rings' substitution patterns, in terms of substituent position (that is, *meta* or *para*), and the diversity of steric and electronic effects, as well as a significant enantioselectivity range. For external validation, four of the original substrates were combined with eight new substrates, which were specifically designed to introduce additional structural patterns.

Mechanistically relevant infrared vibrations were proposed according to structural features of the various substrates. Potential parameters for modelling enantioselectivity were six ring vibrations analogous to those used to describe the bisphenol rings in the previous case study. Additionally, three alkene vibrations were included, because the alkene is directly engaged with the catalyst during the transformation. Evaluating the minimized structures of the diarylalkenes showed that aryl substitution affects ring torsion. Therefore, the measured distance between adjacent *ortho*-carbons on the geminal aryl rings was introduced into the parameter set (Fig. 3c). Finally, considering conjugation between the aryl rings and the alkene, we incorporated intensities and frequencies of two vibration modes that involve both groups (Fig. 3c).

A predictive model (slope, 0.95; intercept, 0.06; $R^2 = 0.88$) with three parameters was determined (Fig. 3b, Supplementary Information). Of particular note is that vibrational intensities were identified as relevant descriptors, in contrast to the bisphenol model, where frequencies are sufficient. Intensity and frequency measure different, but interrelated, effects; frequency is dependent on force constants, bond energy and molecular distances, whereas intensity is a derivative of the dipole moment influenced by molecular symmetry and electronic structure[7,16,17,26]. The three parameters included in the model are the torsion distance, two infrared intensities and a cross-term between these intensities. The

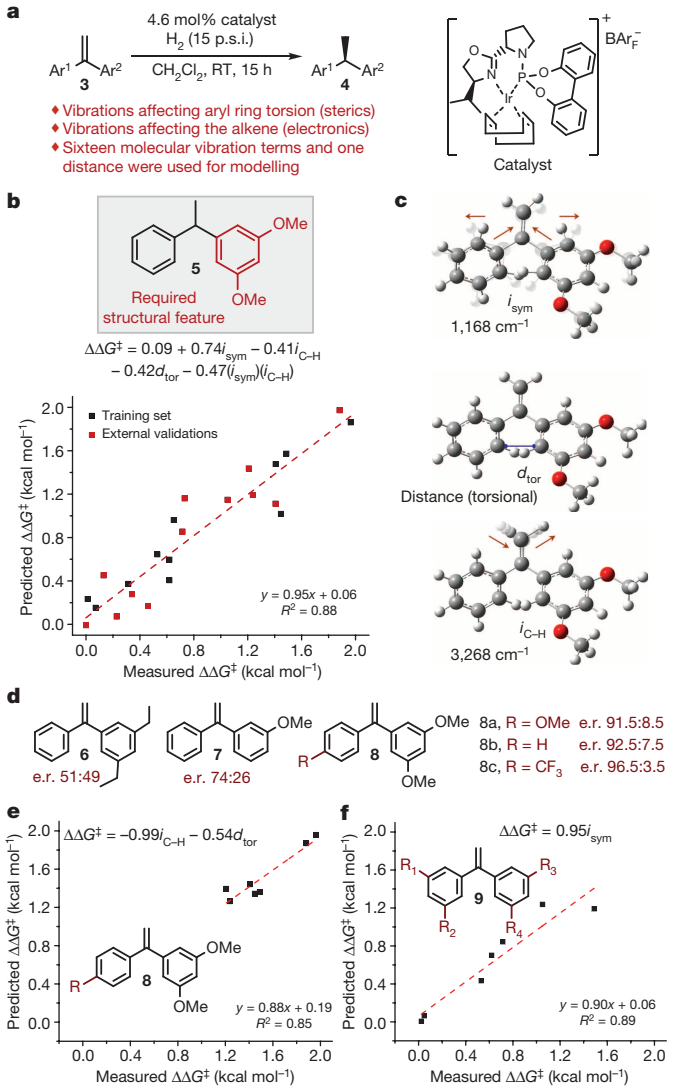


Figure 3 | Using infrared vibrations to correlate enantioselectivity. **a**, Reaction scheme for the enantioselective hydrogenation of 1,1-diarylalkenes. **b**, Mechanistically derived normalized model for the correlation of steric and electronic features of the entire set of substrates to enantioselectivity. **c**, Vibrational intensities and distance measurement used for the correlation of enantioselectivity: d_{tor}, tortional distance between *ortho*-positions; i_{sym}, symmetric central C–C stretch between rings and double bond; $i_{C–H}$, antisymmetric alkene C–H stretch. **d**, Selected substrates from the training set and their respective enantiomeric ratios (e.r.). **e**, Normalized model derived for a subset of all substrates with a 3,5-methoxy motif on one aryl ring and *para*-substitution on the geminal aryl. **f**, Normalized model derived for a subset of all substrates with various *meta*-substitution patterns on both rings.

first intensity, i_{sym}, belongs to the symmetric central C–C stretch between the geminal aryls and the alkene, which is proposed to describe conjugation, and the second, $i_{C–H}$, is the antisymmetric alkene C–H stretch (Fig. 3c).

To assess the effect of each parameter on the prediction of selectivity, two structurally distinct subsets of substrates were evaluated. Hammett σ-constants for *para*-substituents do not adequately describe the enantioselectivity in a first subset where one aryl ring contains a 3,5-dimethoxy motif and the *para*-position of the other is modulated (**8**; Fig. 3e). Deconstructing the original model reveals that this series is defined by only the torsion distance, d_{tor}, and the alkene stretch, $i_{C–H}$ (Fig. 3e). Although the source of the torsional effect may be electronic, these two observations imply that the role of *para*-substitution in determining enantioselectivity is not purely electronic. Interestingly, modulation of *meta*-substituents cannot be described by the same model, and a new model for this second substrate subset relies only on the i_{sym} vibration (**9**; Fig. 3f). It has been proposed[25] that *meta*-substituents could act as directing groups by pre-coordinating to the catalyst. On the basis of this mechanistic hypothesis, the identified intensity, i_{sym}, may denote a selectivity-imparting, direct interaction of *meta*-substituents with the catalyst. Together, these effects serve to define substrates in which both *meta*- and *para*-positions have been perturbed sterically and electronically. Notably, in the full model (Fig. 3b), a cross-term of the two intensities implies a relation between these components for the description of selectivity, suggestive of synergistic effects between the aryl substituents and the alkene.

In the final case study, we sought to expand this technique by exploring site selectivity and simultaneously interrogating two reactants. This is demonstrated in the context of our recently reported redox-relay oxidative Heck arylation, which is highly enantioselective for a broad range of reaction partners and uses a simple chiral ligand[27,28] (Fig. 4a). An interesting mechanistic aspect of this process is the site selectivity of the organometallic addition, leading to constitutional isomers (defined as γ and β). As observed empirically, the respective natures of the alkene and aryl sources both affect site selectivity. Specifically, a Hammett correlation has been observed as a function of the boronic-acid derivative (but only for *meta*- and *para*-substituted examples), with the most electron poor leading to high site selectivity for γ-addition[27]. Furthermore, as the distance between the alkene and the alcohol is increased, the selectivity is reduced. The latter result was initially correlated to the difference in ^{13}C NMR chemical shifts of the two alkene carbons, also suggesting an electronic origin to site selectivity. Finally, although we were unable to correlate the alkene substituent effect quantitatively, we observed a qualitative trend of larger substituents leading to higher site selectivity for γ-addition. These results suggest that both steric and electronic effects affect site selection, providing an exciting platform to utilize our new infrared-vibration-based parameters.

We elected to construct a model for this reaction in three stages. The first stage was aimed at determining whether infrared vibrations can be combined with Hammett analysis to model site selectivity as a function of both boronic-acid and alkene substituents. A specific hypothesis was that the alkene C=C stretching frequency would be sensitive to substituent changes at **R** (Fig. 4b). To test this, we designed[29] an eight-membered library by modulating the alkenol sterically (**R** = *i*-Pr (*iso*-propyl), Et (ethyl), Me) and the boronic acid electronically (**E** = CO_2Me, F, OMe). Using the C=C stretching frequency in combination with the Hammett σ-value provided an excellent model (slope, 0.98; intercept, 0.02; $R^2 = 0.98$; Fig. 4b). These results suggest that the C=C stretching frequency effectively describes the substituent effect on the alkene.

The second stage of this study was directed at the more profound question of describing simultaneous steric and electronic variation. Specifically, we wished to assess whether vibrational analysis could be a substitute for Hammett analysis, yet also allow for the modelling of multiple substituents on the ring or *ortho*-substitution—a significant deficiency of Hammett analysis. Inspired by this question and the classic Hammett analysis, we elected to use infrared vibrations of substituted benzoic acids as a generalized parameter system and as a mimic for arylboronic acids. We chose to examine a single homoallylic alcohol in combination with 12 distinctive aryl boronic acids, including *ortho*-substituted examples. An excellent model was derived, incorporating a ring stretch intensity, the carboxylic acid C=O stretch frequency and a cross-term between the two (slope, 0.95; intercept, 0.08; $R^2 = 0.94$; Fig. 4c). The cross-term is the most important term, with the largest regression coefficient, suggesting that, together, these two vibrations synergistically serve to describe the various effects induced by the boronic-acid component. Markedly, substrates with *ortho*-substitution were effectively predicted.

As the final stage, a model was desired to account for all the products with measured site-selectivity ratios, including those with different

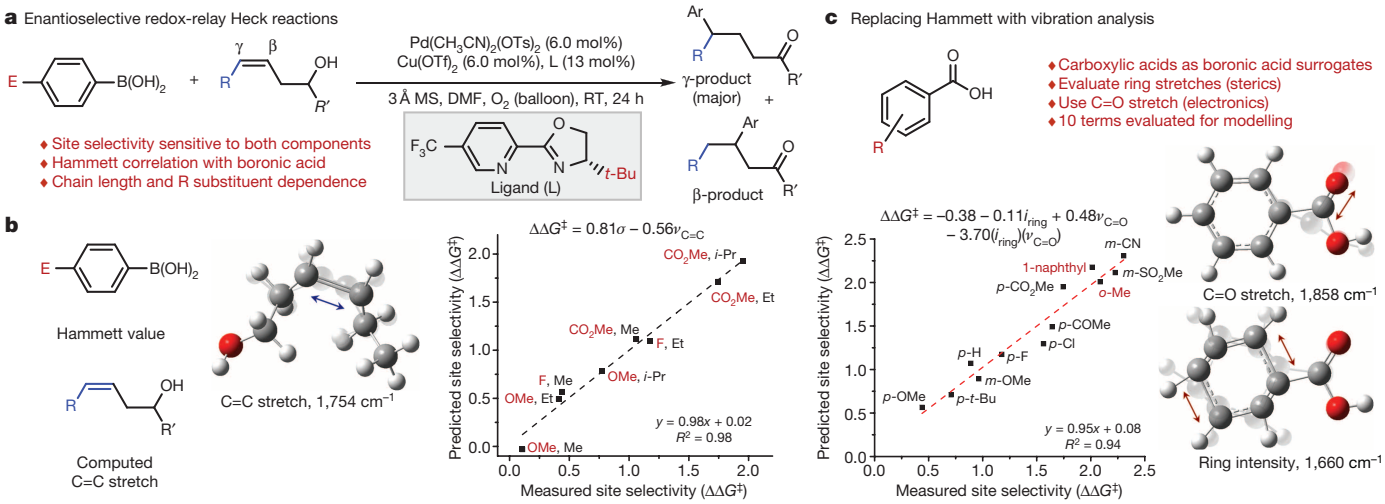


Figure 4 | Using infrared vibrations to correlate site selectivity. **a**, Enantioselective redox-relay Heck reaction scheme and mechanistic considerations. **b**, Correlation of site selectivity to the arene Hammett value and alkenol double-bond stretch (normalized model). **c**, Correlation between site selectivity with different aryl boronic acids and vibrational parameters that describe the aryl moiety (normalized model).

alkene geometries, chain lengths, **R** groups and arylboronic acids (Fig. 5a). A model was developed that effectively predicts 17 external validations (slope, 1.01; intercept, 0.03; $R^2 = 0.92$; Fig. 5b, c). The terms describing the alkene are the sp^2 C–H symmetric stretch and a cross-term between it and the C=C stretch. In this case, the difference relative to those above (Fig. 4b, c) is that this inclusive model must adequately describe the effect of chain length and alkene geometry. Impressively, only two examples of these variations are included in the training set. The arene is described by the same benzoic-acid ring stretch intensity as above, as well as its cross-term with the carbonyl stretch, but an additional ring frequency improves the model. Again, this is presumably due to the more complex interactions between the diverse alkenols and the boronic acids. An intriguing feature of this type of more comprehensive modelling is the potential ability to apply it in extrapolative fashion to predict the performance of new substrates. Because this is a reaction under further investigation, the present analysis will inform mechanistic hypotheses and reaction development.

In this work, we have introduced a method to model reaction selectivity in catalytic systems based on vibrational parameters. The value of this approach is that simultaneous changes to steric and electronic properties can be accounted for and successfully evaluated, overcoming a significant limitation of classical parameter systems.

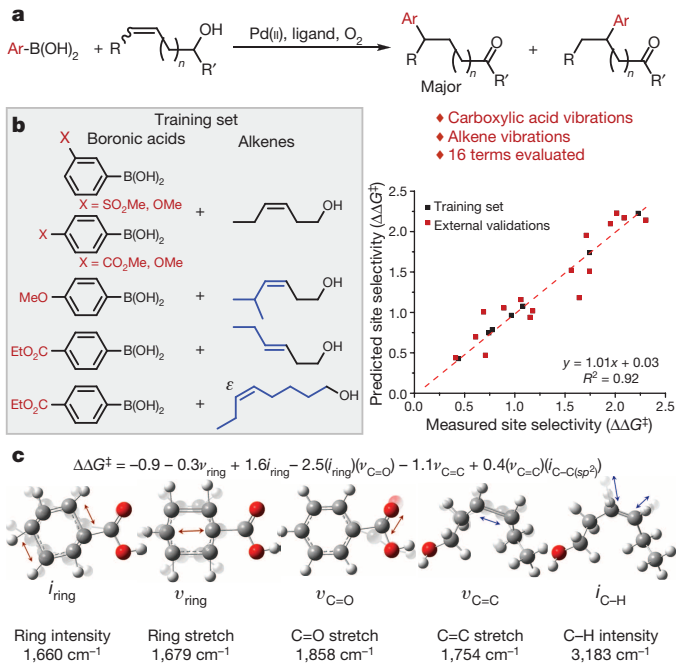


Figure 5 | Developing a comprehensive model. **a**, Enantioselective redox-relay Heck general reaction scheme. **b**, The training set used to develop a model describing site selectivity for products with different alkene geometries, chain lengths, **R** groups and aryl boronic acids. **c**, Normalized model and vibrational terms used for the correlation of enantioselectivity. The two terms on the right, C = C stretch and C–H intensity, are used to describe the alkenol, and the three on the left, C = O stretch, ring intensity and ring stretch, describe the aryl.

METHODS SUMMARY

To execute a strategy that involves computationally derived infrared vibrations as molecular parameters for the mathematical interpretation of selectivity in catalytic reactions, two steps were performed. The first step was identifying vibrations that are sensitive to substituent effects within a given reaction. Realizing this step requires an initial hypothesis for selecting mechanistically relevant vibrations, because complex molecules have an abundance of vibrational modes. Additionally, the selected vibration modes must be identified for all modulated substrates, thus representing the same effect throughout the data set. The abundance, superposition and coalescence of vibration modes lead to complex infrared spectra. This complexity stresses the utility of computed infrared spectra because individual vibrations can be visualized, allowing us to pick out the correct vibration terms for each set of molecules studied. Thus, to produce predictive linear regression models, in each case study we selected specific infrared vibrational modes with a mechanistic hypothesis in mind. Indeed, we found that a hypothesis-based interrogation of the experimental space is required to suggest parameters that can yield predictive, mechanistically relevant models. The second step requires design and evaluation of an empirical training set, which encompasses a systematic variation of substituent properties[29]. The experimental output, enantiomeric or site-selectivity ratios, were mathematically modelled through linear regression techniques to reveal which of the proposed parameters allow for the prediction of new outcomes[30] (Methods and Supplementary Information). The models produced were evaluated for their goodness of fit, and their robustness is demonstrated by external validations' goodness of fit. The nearer the R^2 and slope values are to one (indicating a tight, one-to-one correlation between predicted and measured outcomes) and the nearer the intercept is to zero (indicating minimal systematic error), the more robust the model. Of the potential models, those containing a minimal number of parameters were preferred, because this allows for a mechanistically informative interrogation.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 November 2013; accepted 7 January 2014.

1. Williams, A. *Free Energy Relationships in Organic and Bio-Organic Chemistry* (Royal Society of Chemistry, 2003).

2. Verloop, A., &. Tipker, J. Physical basis of Sterimol and related steric constants. *Pharmacochem. Libr.* **10,** 97–102 (1987).
3. Taft, R. W. Linear steric energy relationships. *J. Am. Chem. Soc.* **75,** 4538–4539 (1953).
4. Hammett, L. P. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **59,** 96–103 (1937).
5. Charton, M. Steric effects. I. Esterification and acid-catalyzed hydrolysis of esters. *J. Am. Chem. Soc.* **97,** 1552–1556 (1975).
6. Hansch, C., Leo, A. & Taft, R. W. A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **91,** 165–195 (1991).
7. Anslyn, E. V. & Dougherty, D. A. *Modern Physical Organic Chemistry* (University Science Books, 2006).
8. Jacobsen, E. N., Zhang, W. & Guler, M. L. Electronic tuning of asymmetric catalysts. *J. Am. Chem. Soc.* **113,** 6703–6704 (1991).
9. Harper, K. C. & Sigman, M. S. Three-dimensional correlation of steric and electronic free energy relationships guides asymmetric propargylation. *Science* **333,** 1875–1878 (2011).
10. Curtin, D. Y. Stereochemical control of organic reactions differences in behaviour of diastereoisomers. *Rec. Chem. Prog.* **15,** 110–128 (1954).
11. Jaffe, H. H. A reexamination of the Hammett equation. *Chem. Rev.* **53,** 191–261 (1953).
12. Hammett, L. P. Some relations between reaction rates and equilibrium constants. *Chem. Rev.* **17,** 125–136 (1935).
13. Charton, M. The application of the Hammett equation to ortho-substituted benzene reaction series. *Can. J. Chem.* **38,** 2493–2499 (1960).
14. Harper, K. C., Bess, E. N. & Sigman, M. S. Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nature Chem.* **4,** 366–374 (2012).
15. Peng, C. S., Baiz, C. R. & Tokmakoff, A. Direct observation of ground-state lactam-lactim tautomerization using temperature-jump transient 2D IR spectroscopy. *Proc. Natl Acad. Sci. USA* **110,** 9243–9248 (2013).
16. Coates, J. in *Encyclopedia of Analytical Chemistry* (ed. Meyers, R. A.) 10815–10837 (Wiley, 2000).
17. Meyer, M. P. New applications of isotope effects in the determination of organic reaction mechanisms. *Adv. Phys. Org. Chem.* **46,** 57–120 (2012).
18. Harper, K. C., Vilardi, S. C. & Sigman, M. S. Prediction of catalyst and substrate performance in the enantioselective propargylation of aliphatic ketones by a multidimensional model of steric effects. *J. Am. Chem. Soc.* **135,** 2482–2485 (2013).
19. Gustafson, J. L., Sigman, M. S. & Miller, S. J. Linear free-energy relationship analysis of a catalytic desymmetrization reaction of a diarylmethane-bis(phenol). *Org. Lett.* **12,** 2794–2797 (2010).
20. Valero, R., Gomes, J. R. B., Truhlar, D. G. & Illas, F. Good performance of the M06 family of hybrid meta generalized gradient approximation density functionals on a difficult case: CO adsorption on MgO(001). *J. Chem. Phys.* **129,** 124710 (2008).
21. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120,** 215–241 (2008).
22. Schäfer, A., Huber, C. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **100,** 5829 (1994).
23. Schäfer, A., Horn, H. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **97,** 2571 (1992).
24. Gaussian 09 rev. C.01 (Gaussian Inc., Wallingford, 2009).
25. Bess, E. N. & Sigman, M. S. Distinctive meta-directing group effect for iridium-catalyzed 1,1-diarylalkene enantioselective hydrogenation. *Org. Lett.* **15,** 646–649 (2013).
26. Dunitz, J. D. & Ibberson, R. M. Is deuterium always smaller than protium? *Angew. Chem. Int. Ed.* **47,** 4208–4210 (2008).
27. Mei, T.-S., Werner, E. W., Burckle, A. J. & Sigman, M. S. Enantioselective redox-relay oxidative heck arylations of acyclic alkenyl alcohols using boronic acids. *J. Am. Chem. Soc.* **135,** 6830–6833 (2013).
28. Werner, E. W., Mei, T. S., Burckle, A. J. & Sigman, M. S. Enantioselective Heck arylations of acyclic alkenyl alcohols using a redox-relay strategy. *Science* **338,** 1455–1458 (2012).
29. Livingstone, D. *Data Analysis for Chemists* (Oxford Univ. Press, 1995).
30. Draper, N. R. & Smith, H. *Applied Regression Analysis* (Wiley, 1998).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the US National Science Foundation (CHE-0749506). We thank S. J. Miller and S. Yoganathan for discussions and for providing the peptide catalyst used in these studies. The support and resources of the Center for High Performance Computing at the University of Utah are gratefully acknowledged.

Author Contributions A.M. and M.S.S. had the idea for the work; A.M. and E.N.B. performed the experiments; A.M. carried out computational and modelling analyses; A.M., E.N.B. and M.S.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S.S. (sigman@chem.utah.edu).

METHODS

In this work, we develop linear regression models using parameters that pertain to steric and electronic properties of the substrates in each case study. For the development of a model, mechanistically relevant parameters were equated to the enantioselectivity or site selectivity in terms of relative energy—$\Delta\Delta G^{\ddagger}$ (kcal mol^{-1}). The first step requires selection of an adequate library that could be used as a training set for the identification of a model that describes selectivity. An oversized, poorly designed training set could lead to the inclusion of parameters that describe idiosyncrasies of the training set rather than true reactivity or selectivity trends[31]. Design-of-experiment principles guide training set selection, that is, systematic variation of substituent properties[29].

For the case study on the desymmetrization of bisphenols, steric models were previously reported[14,19] and, hence, we were able to minimize the substrate training set by comparing the Sterimol B_1 and enantioselectivity values of these studies. It was assumed that because the B_1 values have the greatest effect on enantioselectivity, substrates with similar B_1 and enantioselectivity values are redundant in the training set. In the subsequent case studies, we applied chemical intuition concerning steric and electronic perturbations to select a training set. A general consideration in picking training sets is choosing a minimal number of substrates that exemplify many of the structural variations present in the entire data set. Having this in mind, it should be noted that we purposefully excluded some substrates from the training sets, with the intention of interrogating the ability of a model to externally predict these substrates. Such an external validation establishes the generality of a model, because it is able to predict variations that were not explicitly incorporated in the training set. One such deliberate omission of substrates was *ortho*-substituted boronic acids in the training set in the case study on Heck arylation.

The second step includes four stepwise regression algorithms (for details, see Supplementary Information) that assess the significance of each parameter by applying statistical criteria. To realize this assessment, each set of parameter values is normalized by subtracting its respective mean and dividing by its standard deviation. The four stepwise regression algorithms are built into the MATLAB statistical toolbox and add or remove normalized parameters from an initial model according to a P value threshold[32,33]. Additional suggestions for models are inputted manually on the basis of the results of the four preliminary algorithm runs or on mechanistic hypotheses. A linear fit is performed to probe each manually suggested model, and each suggestion is examined as an initial model for a subsequent stepwise regression iteration to seek a more effective model. After identifying several statistically probable models, an external validation is carried out to determine the predictive nature of the proposed models for substrates that are not included in the training set. This procedure allows us to propose statistically probable models that describe the training set, and then to assess the predictive efficacy of each model towards an external validation set.

The theoretical context of this work is epitomized by Hammett's seminal observation that the acidity of benzoic acid derivatives can be correlated to new equilibria and, ultimately, to reaction rates[4,6,12] (Fig. 1a). Such types of free-energy relationships have been an extraordinarily revealing tool for reaction study[1,8,9,14,18,34–38]. However, a significant limitation is the inability of a single parameter to describe a substituent that introduces perturbations to more than one molecular property. As an example, Hammett parameters can be used only to describe the *ortho*-position on a benzene ring in cases where there is no steric effect involved, limiting their applicability to data sets that contain solely electronic trends[13]. Conversely, steric molecular descriptors, which relate to the spatial arrangement of atoms in a molecule, such as Taft[3,39], Charton[5,40–43] or Sterimol[2,44–46] parameters (Fig. 1a), can be used only to interrogate a reaction trend in the absence of significant electronic effects.

To overcome the aforementioned limitations, we applied a parameter set of molecular vibrations for the quantitative analysis of concurrent steric and electronic structural perturbations. Therefore, in each of the disclosed case studies, we selected and applied several infrared frequencies and intensities, as well as additional mechanistically relevant parameters that are presumed to have a bearing on selectivity. This approach is reminiscent of developing quantitative structure activity relationship (QSAR) regression models for the exploration of chemical or biological systems[47,48]. Yet, so far, infrared vibrations have not been an established parameter of choice in QSAR[49–51]. As an example, the evaluation of the fingerprint region of experimental infrared spectra as a whole (between 1,500 and 600 cm^{-1}), underperformed relative to models derived from typical QSAR parameters[51]. Nevertheless, by proposing infrared vibrational-energy-derived parameters that represent mechanistically pertinent features of the substrates and reactions at hand, it is possible to produce predictive, meaningful correlations.

31. Baldi, P. & Brunak, S. *Bioinformatics: the Machine Learning Approach* 2nd edn, 5–7 (MIT Press, 2001).
32. Goodman, S. A dirty dozen: twelve *P*-value misconceptions. *Semin. Hematol.* **45,** 135–140 (2008).
33. MATLAB student version v.7.14.0.739 (R2012a) (MathWorks Inc., 2012).
34. Jensen, K. H. & Sigman, M. S. Evaluation of catalyst acidity and substrate electronic effects in a hydrogen bond-catalyzed enantioselective reaction. *J. Org. Chem.* **75,** 7194–7201 (2010).
35. Sigman, M. S. & Miller, J. J. Examination of the role of Taft-type steric parameters in asymmetric catalysis. *J. Org. Chem.* **74,** 7633–7643 (2009).
36. Miller, J. J. & Sigman, M. S. Quantitatively correlating the effect of ligand-substituent size in asymmetric catalysis using linear free energy relationships. *Angew. Chem. Int. Ed.* **47,** 771–774 (2008).
37. Harper, K. C. & Sigman, M. S. Using physical organic parameters to correlate asymmetric catalyst performance. *J. Org. Chem.* **78,** 2813–2818 (2013).
38. Bess, E. N. & Sigman, M. S. in *Asymmetric Synthesis II: More Methods and Applications* (eds Christmann, M. & Bräse, S.) 363–370 (Wiley-VCH, 2012).
39. Taft, R. W. Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters. *J. Am. Chem. Soc.* **74,** 3120–3128 (1952).
40. Charton, M. Steric effects. II. Base-catalyzed ester hydrolysis. *J. Am. Chem. Soc.* **97,** 3691–3693 (1975).
41. Charton, M. Steric effects. III. Bimolecular nucleophilic substitution. *J. Am. Chem. Soc.* **97,** 3694–3697 (1975).
42. Charton, M. Steric effects. 7. Additional V constants. *J. Org. Chem.* **41,** 2217–2220 (1976).
43. Charton, M. Steric effects. 8. Racemization of chiral biphenyls. *J. Org. Chem.* **42,** 2528–2529 (1977).
44. Verloop, A. in *Drug Design* Vol. III (ed. Ariens, E. J.) 133 (Academic, 1976).
45. Verloop, A. & Tipker, J. A comparative study of new steric parameters in drug design. *Pharmacochem. Libr.* **2,** 63–81 (1977).
46. Verloop, A. in *Pesticide Chemistry, Human Welfare and Environment* Vol. 1 (eds Miyamoto, J. & Kearney, P. C.) 339–344 (Pergamon, 1983).
47. Shahlaei, M. Descriptor selection methods in quantitative structure–activity relationship studies: a review study. *Chem. Rev.* **113,** 8093–8103 (2013).
48. Dearden, J. C. & Cronin, M. T. D. in *Smith and Williams' Introduction to the Principles of Drug Design and Action* (ed. Smith, H. J. & Williams, H.) 185–209 (2006).
49. Pino, A., Giuliani, A. & Benigni, R. Toxicity mode-of-action: discrimination via infrared spectra and eigenvalues of the modified adjacency matrix. *QSAR Comb. Sci.* **22,** 191–195 (2003).
50. Benigni, R., Giuliani, A. & Passerini, L. Infrared spectra as chemical descriptors for QSAR models. *J. Chem. Inf. Model.* **41,** 727–730 (2001).
51. Benigni, R., Passerini, L., Livingstone, D. J., Johnson, M. A. & Giuliani, A. Infrared spectra information and their correlation with QSAR descriptors. *J. Chem. Inf. Model.* **39,** 558–562 (1999).

LETTER

doi:10.1038/nature12963

Conformation-induced remote *meta*-C–H activation of amines

Ri-Yuan Tang[1], Gang Li[1] & Jin-Quan Yu[1]

Achieving site selectivity in carbon–hydrogen (C–H) functionalization reactions is a long-standing challenge in organic chemistry. The small differences in intrinsic reactivity of C–H bonds in any given organic molecule can lead to the activation of undesired C–H bonds by a non-selective catalyst. One solution to this problem is to distinguish C–H bonds on the basis of their location in the molecule relative to a specific functional group. In this context, the activation of C–H bonds five or six bonds away from a functional group by cyclometallation has been extensively studied[1–13]. However, the directed activation of C–H bonds that are distal to (more than six bonds away) functional groups has remained challenging, especially when the target C–H bond is geometrically inaccessible to directed metallation owing to the ring strain encountered in cyclometallation[14,15]. Here we report a recyclable template that directs the olefination and acetoxylation of distal *meta*-C–H bonds—as far as 11 bonds away—of anilines and benzylic amines. This template is able to direct the *meta*-selective C–H functionalization of bicyclic heterocycles via a highly strained, tricyclic-cyclophane-like palladated intermediate. X-ray and nuclear magnetic resonance studies reveal that the conformational biases induced by a single fluorine substitution in the template can be enhanced by using a ligand to switch from *ortho*- to *meta*-selectivity.

The selective functionalization of inert C–H bonds at different sites of organic molecules provides an opportunity for the introduction of diverse structural modifications and the development of novel retrosynthetic disconnections. However, the widespread application of C–H functionalization in organic synthesis is hampered by a lack of catalysts, reagents and methodologies that enable the site-selective functionalization of C–H bonds, which often have very subtle differences in intrinsic reactivity. We have broadly focused on the development of metal-catalysed C–H activation reactions that are directed by weakly coordinating functional groups[1]. In analogy to the principles of proximity-driven metallation[2], this type of methodology enables the selective functionalization of C–H bonds that are five or six bonds away from the directing atom, through cyclometallation[3–13]. Although this approach has enabled the discovery of numerous transformations over the past decade, the functionalization of C–H bonds that are located farther away from the coordinating functional group remains a largely unsolved problem in organic synthesis, especially when their locations do not permit cyclometallation owing to geometric strain[14–18].

Recently we developed an end-on-coordinating, nitrile-based template that is able to direct Pd(II)-catalysed *meta*-selective olefination and arylation of hydrocinnamic acids[14,15]. This discovery led us to explore three key questions: first, whether this new end-on template approach can be applied to other substrate classes; second, whether other types of transformation using different catalytic manifolds can be achieved using end-on templates; and, third, whether there are critical and general underlying principles for the design of an effective template to direct remote C–H activation. In this context, we recognized that the selective activation of C–H bonds at C7 of tetrahydroquinolines is a conceptually intriguing and synthetically important challenge. A novel template will be required to accommodate a highly strained intermediate with a tricyclic cyclophane structure (Fig. 1a).

Here we report the rational design of a nitrile-containing template that directs C7-selective C–H activation of tetrahydroquinolines (Fig. 1). By systematically modifying the structure of the template, we identified the template conformation as a critical factor in favouring remote *meta*- or proximate *ortho*-selectivity (Fig. 1b). Remarkably, by tuning the properties of the Pd(II) catalyst through use of *N*-acetyl-glycine (Ac-Gly-OH) as a ligand, the pre-existing conformational bias in the template can be further amplified to achieve remote *meta*-C–H olefination in excellent yield and with high levels of site selectivity (Fig. 1c). This optimized template is broadly applicable to the remote C–H activation of 2-phenylpyrrolidines, 2-phenylpiperidines and other aniline-type substrates, despite the intrinsic electronic biases in these substrates that favour *ortho*-functionalization (Fig. 1d). In addition to *meta*-C–H olefination via a Pd(II)/Pd(0) redox cycle, we were also able to demonstrate *meta*-C–H acetoxylation via Pd(II)/Pd(IV) catalysis using this template. This template can be easily installed and later recycled, similar to chiral auxiliaries that are widely used in organic synthesis, such as the well-known Evans oxazolidinone. This work paves the way for practical applications of remote C–H activation through template control.

A procedure exists for the *meta*-selective olefination of hydrocinnamic acids using a novel end-on 2-aminobenzonitrile template[14]. Crucial for the *meta*-selectivity, the directing nitrile group is in extended conjugation with the carbonyl moiety of the substrate, which positions the nitrile group in close proximity to and coplanar with the target *meta*-C–H bond. However, in translating this insight to the *meta*-selective olefination of tetrahydroquinolines, we needed to design an entirely novel end-on template and we faced several considerable challenges in determining the optimal structural design. First, the hypothetical palladation intermediate would involve a highly strained intermediate with a tricyclic cyclophane structure (Fig. 1a). Second, although a simple amide linkage is desirable for practical attachment of the template, we realised that the amide group could potentially favour the activation of the *ortho*-(C8) position, an established mode of reactivity for anilide-type substrates[19]. Moreover, amine substituents are well-known *ortho*/*para* directors in electrophilic aromatic substitution reactions, including electrophilic palladation. Third, to avoid the pitfall of over-engineering, we hoped to develop a simple amide template with an sp^3-hybridized backbone without having to build in an extended conjugation. We recognized that such a template could lead our substrates to exist in multiple interconverting conformations, each with a low equilibrium population. This would translate into a high entropic barrier for formation of the highly organized transition state required for cyclopalladation. We therefore aimed to acquire an improved understanding of how the conformation of non-constrained atoms in a template can be manipulated to favour remote C–H activation. In addition, we hoped to develop a catalyst to recognize and harness subtle conformational biases and amplify pre-existing template-induced preferences for *meta*-selectivity.

To begin our investigation, we attached the simple nitrile templates $\mathbf{T_1}$–$\mathbf{T_3}$ to tetrahydroquinoline and tested these substrates in a model reaction, the Pd(II)-catalysed C–H olefination (Fig. 2a, b). Although the reaction of **1** did not provide any olefinated products, **2** and **3** afforded mixtures of positional isomers that are difficult to separate. The addition

[1]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

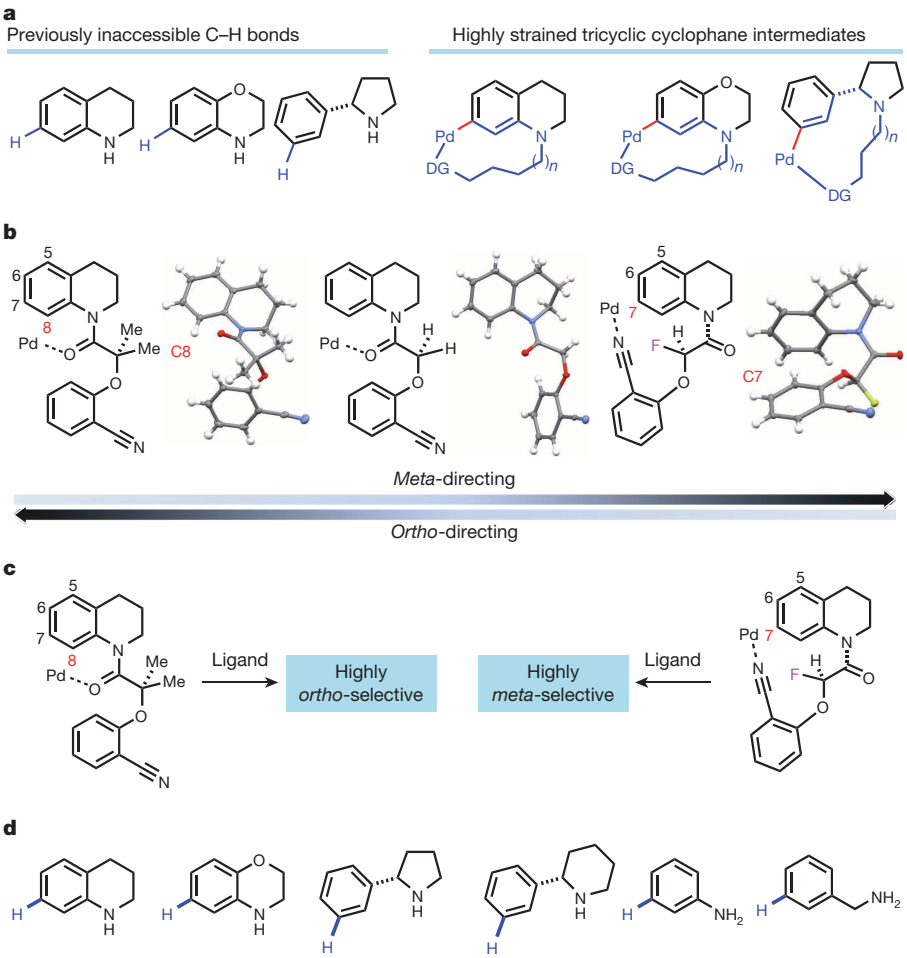


Figure 1 | Design of a versatile template to direct *meta*-C–H activation. **a**, The challenge of remote *meta*-C–H activation of bicyclic heterocycles, illustrated by previously inaccessible C–H bonds and highly strained tricyclic cyclophane intermediates. DG, directing group. **b**, Proposed conformation-controlled *meta*-C–H activation. Me, methyl. **c**, Amplification of *meta*-selectivity by use of a ligand. **d**, Scope of *meta*-C–H activation: olefination and acetoxylation. We highlight the C–H bonds and the palladation intermediate (blue), the newly formed bonds (red) and the importance of fluorine substitution to this reactivity (magenta).

of our previously identified ligand, Ac-Gly-OH, did not improve the level of selectivity with these templates. We attributed the poor site selectivity to either a lack of conformational rigidity in the template backbones ($\mathbf{T_1}$ and $\mathbf{T_2}$) or the template being too short to reach the target C–H bond ($\mathbf{T_3}$). We subsequently explored the α-hydroxy template structure ($\mathbf{T_4}$, $\mathbf{T_5}$ and $\mathbf{T_6}$) and observed an encouraging trend favouring C7 selectivity in C–H olefination (Fig. 2b). Although exclusive C8-olefination was observed using template $\mathbf{T_4}$ (substrate **4**), the use of templates $\mathbf{T_5}$ (substrate **5**) and $\mathbf{T_6}$ (substrate **6a**) afforded C7-olefinated product with selectivities of 11% and 20%, respectively. Although the level of *meta*-selectivity was still poor with these templates, the observation that a single fluorine substituent, in $\mathbf{T_6}$, nearly doubles the *meta*-selectivity prompted us to investigate the origin of this phenomenon.

Because fluorine substituents can lead to pronounced changes in molecular conformation[20,21], we studied the conformations of **4**–**6** in the solid state and in solution. X-ray crystal structures of **4** and **5** showed that the carbonyl group is perfectly oriented to perform *ortho*-C–H activation. However, in the X-ray crystal structure of **6a** the carbonyl group is oriented away from C7, which presumably makes it better suited for the nitrile moiety to approach the *meta*-C–H bond (C8). Studies of the nuclear Overhauser effect showed that a similar conformational trend is present in solution (Supplementary Information). Steric hindrance from the *gem*-dimethyl substituents in **4** probably raises the activation energy for amide N–C bond rotation, leading to highly restricted interconversion between the pro-*ortho* (ground state) and pro-*meta* conformations. In contrast, the amide C–N bond in **5** can more freely rotate, leading to some of the *meta*-C–H olefination product. Notably, introduction of the α-fluorine substitution in **6a** leads to a conformational switch wherein the carbonyl group is directed away from the *ortho*-C–H bond. Increasing the proportion of this conformation seems to improve the rate of *meta*-C–H activation; however, a low barrier for interconversion between the pro-*meta* and pro-*ortho* conformers still results in substantial levels of *ortho*-C–H activation.

We reasoned that the *meta*-selectivity of **6a** could be amplified by the use of a bulkier and more electron-rich catalyst because the C7 position is less sterically hindered and more electron poor than C8 (ref. 22). In fact, we found that the use of Ac-Gly-OH as a ligand in conjunction with template $\mathbf{T_6}$ improves the level of *meta*-selectivity from 20% to 92% (Fig. 2b). After deprotection, the *meta*-olefinated product **7a** was isolated in 75% yield. Under these ligand-enhanced reaction conditions, the *meta*-selectivity of substrate **5** is also markedly increased (11% to 84%). In contrast, olefination of substrate **4** with Ac-Gly-OH as a ligand provides the olefinated product with 91% *ortho*-selectivity (92% yield). These results suggest that, although the Ac-Gly-OH ligand can promote *meta*-selectivity, the appropriate conformation of the template is a prerequisite for achieving high levels of *meta*-selectivity. We have also tested the analogue of $\mathbf{T_6}$ that contains α-difluoro substituents, in an attempt to improve the *meta*-selectivity further. We obtained a yield and *meta*-selectivity similar to those gained using $\mathbf{T_6}$, indicating that electronic effect does not have a major role in controlling *meta*-selectivity (Supplementary Information).

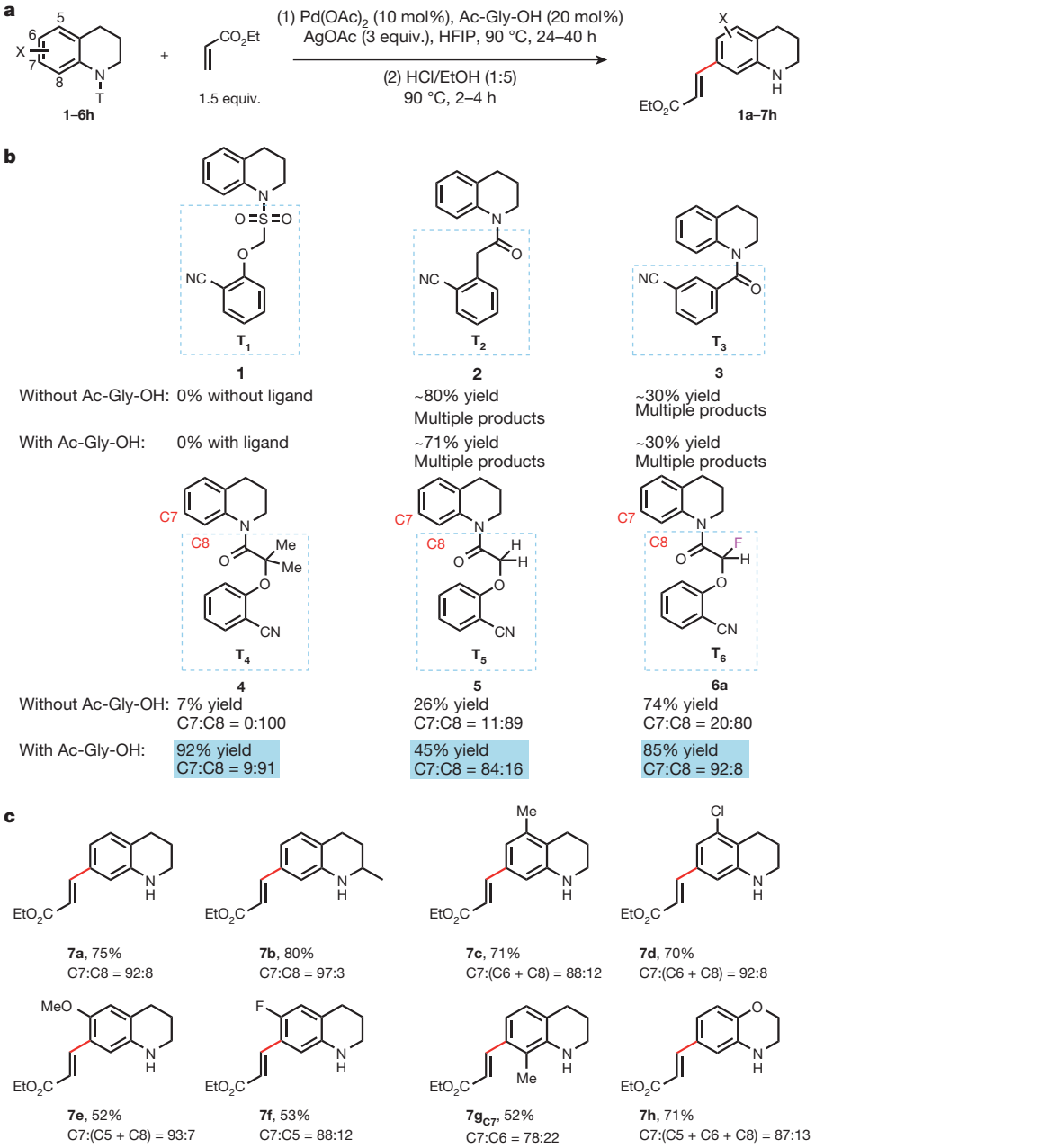


Figure 2 | Development of templates to direct *meta*-C–H olefination. **a**, Olefination of tetrahydroquinoline derivatives. **b**, Six representative templates designed to screen the *meta*-C–H olefination. **c**, Tetrahydroquinolines with a variety of substitution patterns appended with template $\mathbf{T_6}$ (**6a**–**6h**) undergo facile *meta*-C–H olefination. The yields of **2** and **3** are NMR yields with CH_2Br_2 as the internal standard. The selectivity is not determined, owing to multiple olefinated products detected. The isolated yields of other olefinated products are shown along with the selectivity (combined yields are shown in **b**). See Supplementary Information for experimental details. Selectivity of the olefinated products was determined by ^{1}H NMR analysis or gas chromatography mass spectrometry (GCMS) using a flame ionization detector. The variance is estimated to be within 5%. HFIP, hexafluoroisopropanol.

Having established an optimal system for *meta*-selective C–H olefination, we proceeded to investigate the scope of tetrahydroquinolines (Fig. 2c). Substitution at C2 improves the *meta*-selectivity (**7b**, 97:3). Substitutions at the C5 and C6 positions were well tolerated (**7c**–**7f**). C8 substitution decreases the level of *meta*-selectivity to 78:22, presumably owing to the steric hindrance of the C7 position (**7g**). Dihydrobenzoxazine **6h** was also selectively *meta*-olefinated to give the desired product in 71% isolated yield (**7h**). It is worth noting that the C7 positions of these heterocyclic skeletons are very difficult to functionalize, and that this methodology could enable the synthesis of a new subclass of these medically active heterocycles.

We subsequently explored the use of our template system in the *meta*-selective olefination of anilines. The directed *ortho*-C–H olefination of acetanilides demonstrates the high reactivity of the *ortho*-position of these substrates towards palladation[19]. A *meta*-selective C–H olefination of anilines would offer a complementary retrosynthetic disconnection for the synthesis of substituted anilines. Thus, aniline was attached to the optimized template $\mathbf{T_6}$ and subjected to the established olefination conditions (Fig. 3). A mixture of mono- and di-olefinated products ($\mathbf{9a_{mono}}$ and $\mathbf{9a_{di}}$) was obtained in 88% combined yield with 99% *meta*-selectivity, suggesting that this template can successfully override an electronic bias towards *ortho*-palladation. A variety of electron-rich and electron-poor *meta*-substituted anilines gave the mono-olefinated products in good to high yields with the *meta*-selectivity ranging from 94% to 99% (**9b**–**9g**). *Ortho*-substituted anilines were selectively olefinated at the less hindered position to give mono-olefinated products ($\mathbf{9h_{mono}}$

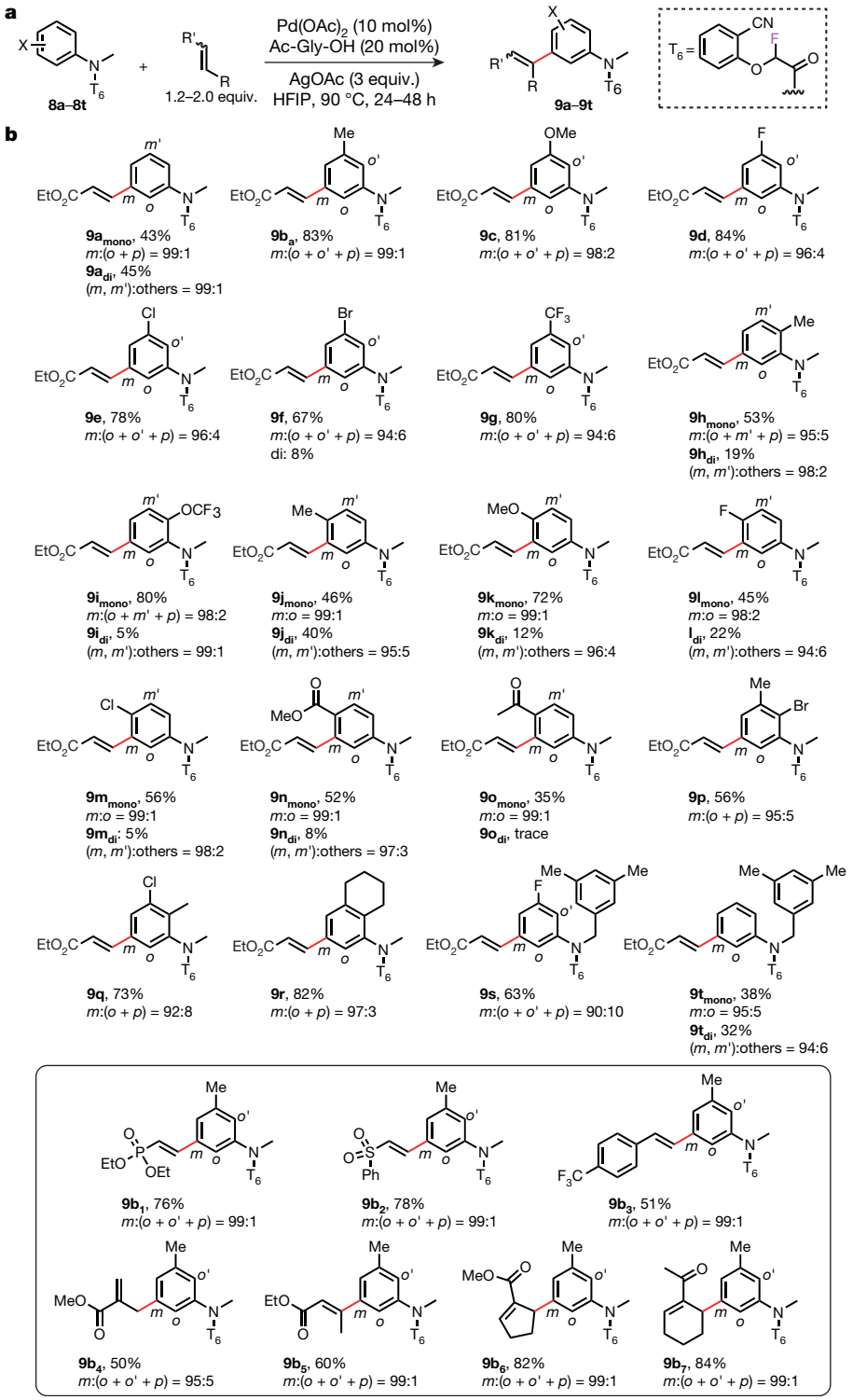


Figure 3 | Template-directed remote C–H olefination of *N*-methylanilines. **a**, Anilines with a variety of substitution patterns were found to undergo facile *meta*-C–H olefination. **b**, In the box, electron-deficient olefins and various di-and tri-substituted olefins were compatible with the transformation in **a**. The isolated yields of the mono-olefinated product (and also the isolated yields of the di-olefinated product, when applicable) are shown along with the selectivity. See Supplementary Information for experimental details. Selectivity of the mono- and di-olefinated products was determined by ^{1}H NMR analysis and GCMS analysis using a flame ionization detector. The variance is estimated to be less than 5%. Ph, phenyl.

and **9i$_{mono}$**) accompanied by small amounts of the di-olefinated byproducts. Despite the steric hindrance, excellent *meta*-selectivity was also obtained for a number of *para*-substituted anilines, albeit with varied mono- or di-selectivity (**9j**–**9o**). Polysubstituted anilines were also smoothly olefinated at the remaining *meta*-positions to afford anilines **9p**–**9r** containing complex 1,2,3,5-substitution patterns. *Meta*-olefination of **8d** was also carried out using 5 mol% $Pd(OAc)_2$ to give the desired product, **9d**, in 60% isolated yield (Supplementary Information). The replacement of the *N*-methyl group by a readily removable benzyl-type protecting group was also tolerated, thus improving the utility of this reaction (**9s** and **9t**). The hydrogenolysis of the benzyl-type protecting group also reduced the installed olefin unit to corresponding alkyl (Supplementary

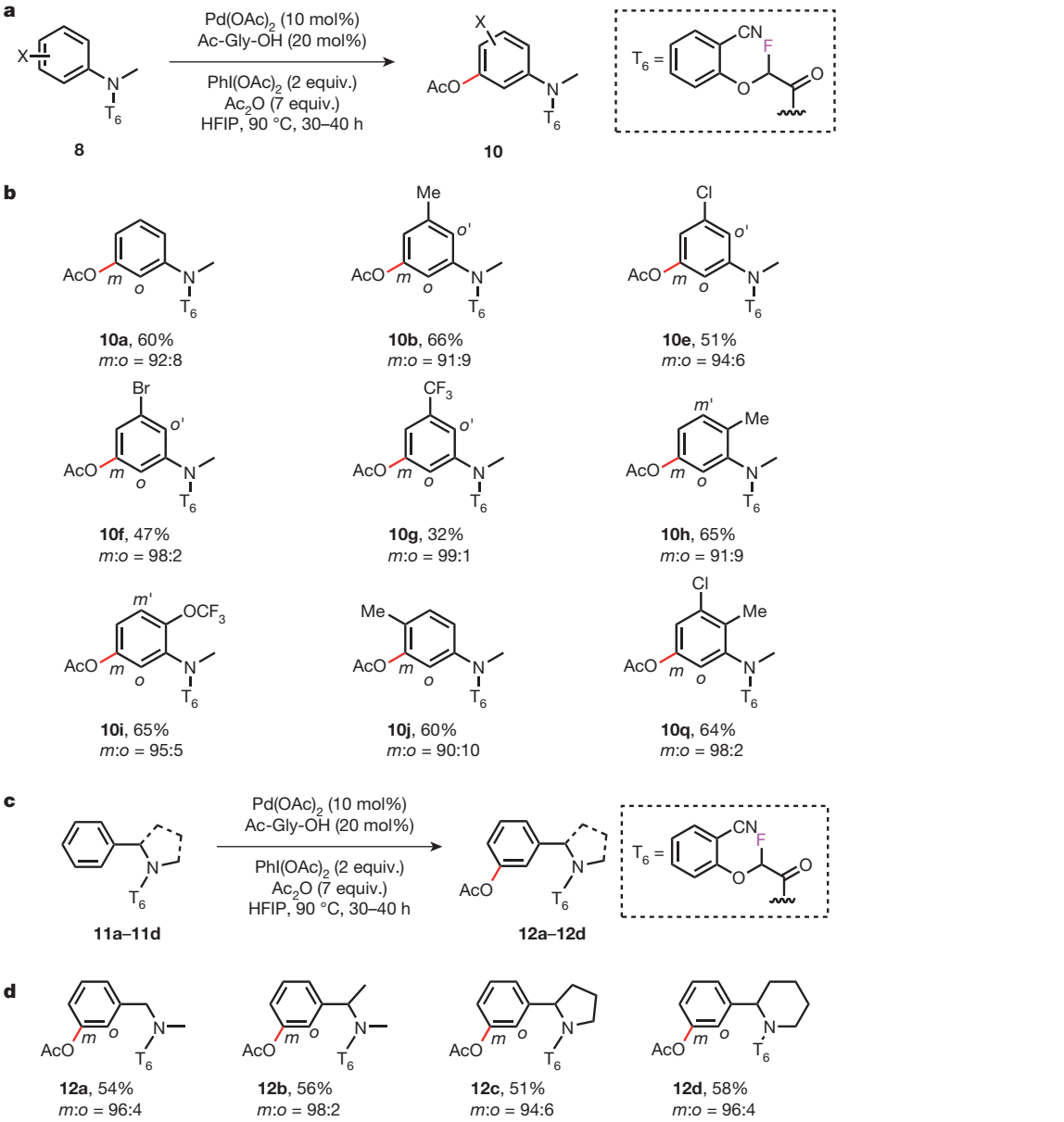


Figure 4 | *Meta*-C–H acetoxylation of *N*-methylanilines and benzylamine derivatives. a, Anilines (**8**) with a variety of substitution patterns undergo facile *meta*-C–H acetoxylation. **b**, The isolated yield of the acetoxylated aniline is shown along with the selectivity. **c**, Benzylamines derivatives (**11a**–**11d**) were found to undergo facile *meta*-C–H acetoxylation. **d**, The isolated yield of the acetoxylated benzylamine derivative is shown along with the selectivity. See Supplementary Information for experimental details. The selectivities of the acetoxylated products were determined by ^{1}H NMR analysis and GCMS analysis using a flame ionization detector. The variance is estimated to be less than 5%.

Information). Finally, a range of olefin coupling partners, including 1,2-di-substituted olefins, were shown to be compatible with this transformation, demonstrating broader scope than the majority of directed *ortho*-C–H olefination reactions (**9b$_1$**–**9b$_7$**).

The utility of our template-directed remote C–H activation approach in developing other types of C–C and C–heteroatom bond-forming reactions via different catalytic manifolds remained to be demonstrated. Thus, amine substrate **8a** was subjected to various C–H oxidation reaction conditions (Fig. 4a, b). Notably, these transformations proceed via a Pd(II)/Pd(IV) redox chemistry as opposed to the Pd(0)/Pd(II) catalytic cycle of C–H olefination. We found that the use of $PhI(OAc)_2$ oxidant[8] affords *meta*-acetoxylated amine **10a** as the major product in 60% isolated yield. Excellent levels of *meta*-selectivity (90%–98%) were obtained with various substituted anilines. A variety of *ortho*-, *meta*- and *para*-substituents were tolerated in this transformation (**10b**–**10j**) although more electron-withdrawing substituents led to a depreciation in yield (**10f** and **10g**). *Meta*-selective acetoxylation of an *ortho*,*meta*-di-substituted aniline was also successful (**10q**). The versatility of our newly developed methodology is further demonstrated by the *meta*-selective acetoxylation of acyclic and cyclic benzylamines (Fig. 4c, d). Notably, the C–H bonds that are cleaved in these in benzylamine substrates are 11 bonds away from the directing nitrogen atom, which is an unprecedentedly long distance for direct C–H activation. *Meta*-selective acetoxylation of 2-phenylpyrrolidine **11c** and 2-phenylpiperidine **11d** is a potentially powerful methodology for accessing diverse structures of medicinally important heterocycles. The hydrolytic removal of the template also converted the acetate to hydroxyl group in one pot (Supplementary Information).

Preliminary mechanistic studies of the acetoxylation of aniline also revealed that conformation of templates, again, has a decisive role in controlling the site selectivity. Control experiments showed that Ac-Gly-OH had only a minor beneficial effect on the yield of acetoxylation and

a negligible influence on the site selectivity. Remarkably, we found that the levels of *meta*-selectivity of the acetoxylation of aniline were respectively 46%, 66% and 92% when templates $\mathbf{T_4}$, $\mathbf{T_5}$ and $\mathbf{T_6}$ were used in the absence of an amino-acid ligand, reflecting the intrinsic conformational biases of these templates very clearly (Supplementary Information).

We have developed a versatile template approach to direct the remote *meta*-C–H bond activation of tetrahydroquinoline, benzoxazines, anilines, benzylamines, 2-phenylpyrrolidines and 2-phenylpiperidines, all of which are commonly used as building blocks in drug discovery. Template $\mathbf{T_6}$ can be readily installed through acylation of the amine substrates with the commercially available 2-(2-cyanophenoxy)-2-fluoroacetic acid (Sigma-Aldrich catalogue number: 791369). Owing to their electronic biases, these amine substrates are incompatible with other known approaches for *meta*-C–H activation[23–30]. We demonstrate that small conformational biases can be amplified by the judicious use of an amino-acid ligand to enhance *meta*-selectivity drastically, although the precise role of the α-fluoro group on the template conformation remains hypothetical at this stage.

METHODS SUMMARY

The general procedure for template-directed *meta*-selective C–H olefination is as follows. A 35-ml sealed tube (with a Teflon cap) equipped with a magnetic stir bar was charged with amide substrate (0.10 mmol, 1.0 equiv.), $Pd(OAc)_2$ (2.3 mg, 0.01 mmol, 10 mol%), Ac-Gly-OH (2.4 mg, 0.02 mmol, 20 mol%) and AgOAc (50 mg, 0.30 mmol, 3.0 equiv.). HFIP (0.5 ml) was added to the mixture, followed by ethyl acrylate (1.2–2.0 equiv.) and, finally, another portion of HFIP (0.5 ml). The tube was then capped and submerged into an oil bath pre-heated to 90 °C. The reaction was stirred for 24–48 h and cooled to room temperature (~25 °C). The crude reaction mixture was diluted with EtOAc (5 ml) and filtered through a short pad of Celite. The sealed tube and Celite pad were washed with an additional 20 ml of EtOAc. The filtrate was concentrated *in vacuo*, and the resulting residue was purified by preparative thin-layer chromatography using hexanes, EtOAc and dichloromethane as the eluent. The positional selectivity was determined by GCMS with a flame ionization detector, and by ^{1}H NMR analysis of the unpurified reaction mixture. Full experimental details and characterization of new compounds can be found in Supplementary Information.

Received 31 August; accepted 18 December 2013.

1. Engle, K. M., Mei, T.-S., Wasa, M. & Yu, J.-Q. Weak coordination as a powerful means for developing broadly useful C–H functionalization reactions. *Acc. Chem. Res.* **45,** 788–802 (2012).
2. Hartung, C. G. & Snieckus, V. in *Modern Arene Chemistry* (ed. Astruc, D.) 330–367 (Wiley-VCH, 2004).
3. Flemming, J. P., Berry, M. B. & Brown, J. M. Sequential *ortho*-lithiations; the sulfoxide group as a relay to enable *meta*-substitution. *Org. Biomol. Chem.* **6,** 1215–1221 (2008).
4. Kakiuchi, F. *et al.* Catalytic addition of aromatic carbon–hydrogen bonds to olefins with the aid of ruthenium complexes. *Bull. Chem. Soc. Jpn* **68,** 62–83 (1995).
5. Jun, C.-H., Hong, J.-B. & Lee, D.-Y. Chelation-assisted hydroacylation. *Synlett*1–12 (1999).
6. Colby, D. A., Bergman, R. G. & Ellman, J. A. Rhodium-catalyzed C–C bond formation via heteroatom-directed C–H bond activation. *Chem. Rev.* **110,** 624–655 (2010).
7. Daugulis, O., Do, H.-Q. & Shabashov, D. Palladium- and copper-catalyzed arylation of carbon–hydrogen bonds. *Acc. Chem. Res.* **42,** 1074–1086 (2009).
8. Desai, L. V., Hull, K. L. & Sanford, M. S. Palladium-catalyzed oxygenation of unactivated sp^3 C–H bonds. *J. Am. Chem. Soc.* **126,** 9542–9543 (2004).
9. Satoh, T. & Miura, M. Oxidative coupling of aromatic substrates with alkynes and alkenes under rhodium catalysis. *Chemistry* **16,** 11212–11222 (2010).
10. Guimond, N., Gorelsky, S. I. & Fagnou, K. Rhodium(III)-catalyzed heterocycle synthesis using an internal oxidant: improved reactivity and mechanistic studies. *J. Am. Chem. Soc.* **133,** 6449–6457 (2011).
11. Rakshit, S., Grohmann, C., Besset, T. & Glorius, F. Rh(III)-catalyzed directed C–H olefination using an oxidizing directing group: mild, efficient, and versatile. *J. Am. Chem. Soc.* **133,** 2350–2353 (2011).
12. Park, S. H., Kim, J. Y. & Chang, S. Rhodium-catalyzed selective olefination of arene esters via C–H bond activation. *Org. Lett.* **13,** 2372–2375 (2011).
13. Ackermann, L. & Pospech, J. Ruthenium-catalyzed oxidative C–H bond alkenylations in water: expedient synthesis of annulated lactones. *Org. Lett.* **13,** 4153–4155 (2011).
14. Leow, D., Li, G., Mei, T.-S. & Yu, J.-Q. Activation of remote *meta*-C–H bonds assisted by an end-on template. *Nature* **486,** 518–522 (2012).
15. Wan, L., Dastbaravardeh, N., Li, G. & Yu, J.-Q. Cross-coupling of remote *meta*-C–H bonds directed by a U-shaped template. *J. Am. Chem. Soc.* **135,** 18056–18059 (2013).
16. Schwarz, H. Remote functionalization of C–H and C–C bonds by naked transition-metal ions (*cosi fan tutte*). *Acc. Chem. Res.* **22,** 282–287 (1989).
17. Breslow, R. Biomimetic control of chemical selectivity. *Acc. Chem. Res.* **13,** 170–177 (1980).
18. Das, S., Incarvito, C. D., Crabtree, R. H. & Brudvig, G. W. Molecular recognition in the selective oxygenation of saturated C–H bonds by a dimanganese catalyst. *Science* **312,** 1941–1943 (2006).
19. Boele, M. D. K. *et al.* Selective Pd-catalyzed oxidative coupling of anilides with olefins through C−H bond activation at room temperature. *J. Am. Chem. Soc.* **124,** 1586–1587 (2002).
20. Paul, S., Schweizer, W. B., Ebert, M.-O. & Gilmour, R. A novel fluorinated gold(I) *N*-heterocyclic carbene complex: exploiting fluorine stereoelectronic effects to control molecular topology. *Organometallics* **29,** 4424–4427 (2010).
21. Banks, J. W. *et al.* The preferred conformation of α-fluoroamides. *J. Chem. Soc. Perkin Trans. 2* 2409–2411 (1999).
22. Wang, D.-H., Engle, K. M., Shi, B.-F. & Yu, J.-Q. Ligand-enabled reactivity and selectivity in a synthetically versatile aryl C–H olefination. *Science* **327,** 315–319 (2010).
23. Cho, J.-Y., Tse, M. K., Holmes, D., Maleczka, R. E. Jr & Smith, M. R. III. Remarkably selective iridium catalysts for the elaboration of aromatic C–H bonds. *Science* **295,** 305–308 (2002).
24. Ishiyama, T. *et al.* Mild iridium-catalyzed borylation of arenes. High turnover numbers, room temperature reactions, and isolation of a potential intermediate. *J. Am. Chem. Soc.* **124,** 390–391 (2002).
25. Zhang, Y.-H., Shi, B.-F. & Yu, J.-Q. Pd(II)-catalyzed olefination of electron-deficient arenes using 2,6-dialkylpyridine ligands. *J. Am. Chem. Soc.* **131,** 5072–5074 (2009).
26. Ye, M., Gao, G.-L. & Yu, J.-Q. Ligand-promoted C-3 selective C–H olefination of pyridines with Pd catalysts. *J. Am. Chem. Soc.* **133,** 6964–6967 (2011).
27. Saidi, O. *et al.* Ruthenium-catalyzed *meta*-sulfonation of 2-phenylpyridines. *J. Am. Chem. Soc.* **133,** 19298–19301 (2011).
28. Duong, H. A., Gilligan, R. E., Cooke, M. L., Phipps, R. J. & Gaunt, M. J. Copper(II)-catalyzed *meta*-selective direct arylation of α-aryl carbonyl compounds. *Angew. Chem. Int. Ed.* **50,** 463–466 (2011).
29. Hull, K. L. & Sanford, M. S. Catalytic and highly regioselective cross-coupling of aromatic C−H substrates. *J. Am. Chem. Soc.* **129,** 11904–11905 (2007).
30. Hofmann, N. & Ackermann, L. *meta*-Selective C–H bond alkylation with secondary alkyl halides. *J. Am. Chem. Soc.* **135,** 5877–5884 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We gratefully acknowledge The Scripps Research Institute and the NIH (NIGMS, 1R01 GM102265) for their financial support. R.-Y.T. is a visiting scholar from Wenzhou University and is sponsored by the National Natural Science Foundation of China (21202121).

Author Contributions R.-Y.T. and G.L. performed the experiments and developed the reactions. R.-Y.T. and J.-Q.Y. designed the templates. J.-Q.Y. had the idea for this work and prepared this manuscript with feedback from R.-Y.T. and G.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-Q.Y. (yu200@scripps.edu).

LETTER

doi:10.1038/nature13080

Hydrous mantle transition zone indicated by ringwoodite included within diamond

D. G. Pearson[1], F. E. Brenker[2], F. Nestola[3], J. McNeill[4], L. Nasdala[5], M. T. Hutchison[6], S. Matveev[1], K. Mather[4], G. Silversmit[7], S. Schmitz[2], B. Vekemans[7] & L. Vincze[7]

The ultimate origin of water in the Earth's hydrosphere is in the deep Earth—the mantle. Theory[1] and experiments[2–4] have shown that although the water storage capacity of olivine-dominated shallow mantle is limited, the Earth's transition zone, at depths between 410 and 660 kilometres, could be a major repository for water, owing to the ability of the higher-pressure polymorphs of olivine—wadsleyite and ringwoodite—to host enough water to comprise up to around 2.5 per cent of their weight. A hydrous transition zone may have a key role in terrestrial magmatism and plate tectonics[5–7], yet despite experimental demonstration of the water-bearing capacity of these phases, geophysical probes such as electrical conductivity have provided conflicting results[8–10], and the issue of whether the transition zone contains abundant water remains highly controversial[11]. Here we report X-ray diffraction, Raman and infrared spectroscopic data that provide, to our knowledge, the first evidence for the terrestrial occurrence of any higher-pressure polymorph of olivine: we find ringwoodite included in a diamond from Juína, Brazil. The water-rich nature of this inclusion, indicated by infrared absorption, along with the preservation of the ringwoodite, is direct evidence that, at least locally, the transition zone is hydrous, to about 1 weight per cent. The finding also indicates that some kimberlites must have their primary sources in this deep mantle region.

Samples of mantle-derived peridotites show that olivine (Mg_2SiO_4) is the dominant phase in the Earth's shallow upper mantle, to a depth of ~400 km (ref. 12). At greater depths, between approximately 410 and 660 km, within the transition zone, the high-pressure olivine polymorphs wadsleyite and ringwoodite are thought to dominate mantle mineralogy owing to the fit of seismic discontinuity data to predictions from phase equilibria[12,13]. No unretrogressed samples of any high-pressure olivine polymorph have been sampled from the mantle, and, hence, this inference is highly likely, but is unconfirmed by sampling. Sampling the transition zone is important because it is thought to be the main region of water storage in the solid Earth, sandwiched between relatively anhydrous shallow upper mantle and lower mantle[4–7]. The potential presence of significant water in this part of the Earth has been invoked to explain key aspects of global volcanism[5] and has significant implications for the physical properties and rheology of the transition zone[3,11,14]. Finding confirmatory evidence of the presence of ringwoodite in Earth's mantle, and determining its water content, is an important step in understanding deep Earth processes.

The discovery of ultradeep diamonds, originating below the lithospheric mantle[15–22], allows a unique window into the material constituting the Earth's transition zone. As such, these diamonds should provide the best opportunity for finding both wadsleyite and ringwoodite. Moreover, several studies have reported olivine that may have originated as a higher-pressure polymorph[21–25].

In this study, we focused on diamonds from the Juína district of Mato Grosso, Brazil, in a search for ultrahigh-pressure inclusions. Alluvial deposits centred on tributaries East of the Rio Aripuanã, Juína District, contain abundant diamonds that originate in the Earth's transition zone and lower mantle[15–19,24,26].

Diamond JUc29 is a 0.09 g, colourless/light-brown, irregular crystal (Extended Data Fig. 1) from deposits of the Rio Vinte e Um de Abril, downstream from kimberlite pipe Aripuanã-01. It exhibits a high degree of surface resorption, is moderately plastically deformed and its nitrogen content is below detection by infrared spectrometry; that is, the diamond is type IIa. These are all characteristics of most ultradeep diamonds from Juína[18]. A crystal of greenish appearance and ~40 μm in its maximum dimension was located optically in the diamond (Extended Data Fig. 1). Synchrotron X-ray tomography shows the inclusion to form part of a pair, with a Ca-rich and a Fe-bearing phase immediately adjacent (Extended Data Fig. 2). Single-crystal X-ray diffraction of the Fe-bearing phase revealed the main four diffraction peaks of ringwoodite, in their relative order of expected intensity[4], that is, in descending order of intensity, the (113) plane at 2.44 Å, the (440) plane at 1.40 Å, the (220) plane at 2.81 Å and the (115) plane at 1.51 Å (Extended Data Fig. 3). The expected fifth peak at about 2.02 Å was not found, being covered by the very intense diamond peak, which occurs at the same *d* spacing (the single distance between two atomic lattice planes belonging to a family of infinite lattice planes all equidistant and parallel). The positions of these peaks (that is, the *d* spacing) and, in particular, the precisely measured relative order of intensities, detected by charge-coupled device (CCD), confirm the identity of the inclusion as ringwoodite but do not allow an accurate compositional estimate.

Micro-Raman spectra of the inclusion (Fig. 1, grey traces) allowed ringwoodite to be identified by the two intense Raman bands that form a doublet corresponding to the asymmetric (T_{2g}) and symmetric (A_{1g}) stretching vibrations of SiO_4 tetrahedra and which occur in the spectral regions ~807 and 860 cm^{-1}, respectively. We refer to these bands as DB1 and DB2, respectively. The spacing of these two bands is 30% wider than those present in olivine, and DB1 is displaced to significantly lower wavenumbers. Band DB1 in JUc29 is defined from peak fitting to be located between 807 and 809 cm^{-1}, with DB2 between 854 and 860 cm^{-1}. The increase in wavenumber of both DB1 and DB2 relative to the reference spectrum in Fig. 1 (red trace) and other synthetic ringwoodites is due largely to the influence of the compressive stress developed around the inclusion. This stress results from the difference in the volume expansion of the inclusion relative to the diamond that has helped to preserve the ringwoodite. All JUc29 Raman spectra show significant broadening of these SiO_4 stretching vibrations. This broadening is probably due to increased disordering resulting from a tendency for ringwoodite to revert to olivine at lower pressure, and hampers the use of the doublet band separation in estimating the composition of the ringwoodite. Nevertheless, an estimate of the composition can be attempted, on the basis of the shift in DB1 in response to pressure and increasing Fe in the structure, which have opposite effects (see Methods

[1]Department of Earth and Atmospheric Sciences, 1-26 Earth Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E3, Canada. [2]Geoscience Institute – Mineralogy, Goethe University, Altenhöferallee 1, 60438 Frankfurt, Germany. [3]Dipartimento di Geoscienze, Università di Padova, 35137 Padua, Italy. [4]Department of Earth Sciences, Durham University, Durham DH1 3LE, UK. [5]Institut für Mineralogie und Kristallographie, Universität Wien, Althanstrasse 14, 1090 Wien, Austria. [6]Trigon GeoServices Ltd, 2780 South Jones Boulevard, #35-15, Las Vegas, Nevada 89146, USA. [7]Department of Analytical Chemistry, Ghent University, Krijgslaan 281 S12, B-9000 Ghent, Belgium.

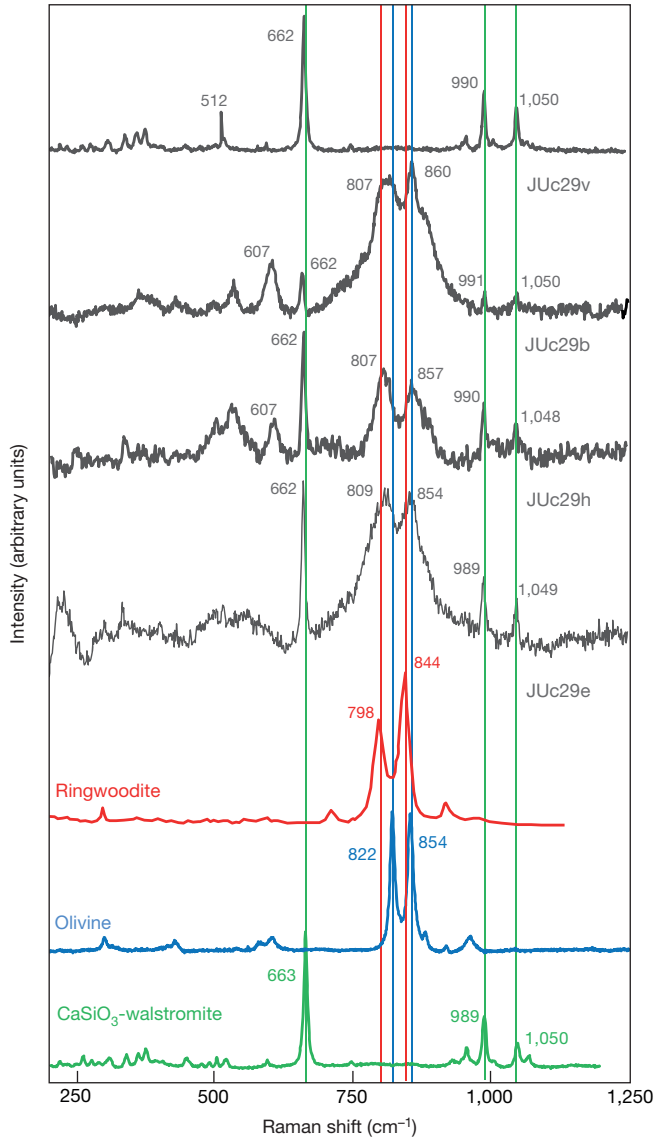


Figure 1 | Raman spectra of ringwoodite and walstromite inclusions in Juína diamond JUc29. Raman spectra (unsmoothed, background-subtracted, in relative intensity units, stacked for clarity, shown in grey) for two-phase inclusion within JUc29 diamond, Juína. Spectra are complex, displaying SiO_4^- stretching modes for ringwoodite ([Mg,Fe]SiO_4) that are broadened, probably by disordering induced by incipient retrogression, as well as the characteristic modes for Ca-walstromite ($CaSiO_3$). Reference spectra for olivine (blue), ringwoodite (red) and $CaSiO_3$-walstromite (green) are from refs 20, 33.

section on Raman spectroscopy). The compressive stress imposed on the inclusion was estimated by measuring the Raman shift of the main diamond band in the immediately adjacent diamond (1,337 cm^{-1}), which yields internal pressures of between 1.7 and 2.3 GPa depending on the pressure calibration of the Raman shift used (see Methods as above). Our estimate for the resulting phase composition yields a Mg number, Mg# = 100Mg/(Mg+Fe), of 75^{+24}_{-21}, where the uncertainty is dominated by the uncertainty in the confining pressure, the exact position of DB1 and the calibration of DB1's position with composition (see Methods as above). Although the compositional uncertainty is large, the presence of significant Fe in the structure is consistent with the confocal X-ray fluorescence data (Extended Data Fig. 2).

Additional Raman-active bands at 662, 990 and 1,050 cm^{-1} are present in the JUc29 spectra and can be attributed to the presence of $CaSiO_3$-walstromite (Fig. 1) adjacent to ringwoodite. Spectrum JUc29v sampled only the Ca-rich phase and is spectrally very similar to reference $CaSiO_3$-walstromite (Fig. 1, green trace).

Fourier transform infrared (FTIR) spectra for the inclusion reveal a pronounced OH^- stretching vibration with the band centre between 3,150 and 3,200 cm^{-1} (Fig. 2). The broad band at 3,150 cm^{-1} and that at 3,680 cm^{-1} correspond to OH^- stretching modes reported in synthetic hydrous ringwoodite[27–31]. The correspondence between the general form of the JUc29 FTIR spectra and that of synthetic hydrous ringwoodite[4,27–31], together with the location of the main OH^- stretching band at considerably lower wavenumber than either hydrous olivine or wadsleyite strongly support the identification of our inclusion as not only ringwoodite, but ringwoodite containing significant water. The location of the main OH^- band at between 3,160 and 3,180 cm^{-1} seems to support a composition between Mg# 60 and Mg# 100 (see Methods section on FTIR spectroscopy), and is hence consistent with the Raman estimate.

The phase assemblage presented by the inclusion pair can be used to constrain their likely depth of origin. Two scenarios are possible, indicative of different depths of mantle sampling. Ca-walstromite is stable, along a mantle geotherm, at or below 10 GPa (refs 20, 26), where ringwoodite with Mg# ~75 must coexist with olivine in a two-phase loop[12]. Although the peak broadening of the main doublet in some of the ringwoodite Raman spectra (for example JUc29v; Fig. 1) indicates the possibility of partial retrogression to olivine in parts of the crystal, there is no indication of a highly crystalline olivine phase from the X-ray measurement. Hence, we interpret this phenomenon as disorder induced during the incipient breakdown of ringwoodite to olivine. Given this, the most likely interpretation of this two-phase assemblage is that it represents a partly retrogressed portion of a somewhat Fe-rich peridotitic mantle, in which hydrous ringwoodite and former $CaSiO_3$-perovskite coexisted above 15GPa[12], that is, in the transition zone, probably with majorite garnet. The ringwoodite has largely avoided retrogression, whereas the $CaSiO_3$-perovskite precursor reverted to Ca-walstromite. The slightly more Fe-rich composition of the ringwoodite may arise by reaction between the peridotitic and basaltic portions of a subducted slab[26] and may not be indicative of the bulk of the transition zone because of the resulting broadening of the 410-km seismic discontinuity that would be seen at such Fe-rich compositions[12].

It is important to constrain the amount of water in the ringwoodite inclusion because this has implications for the water content of the transition zone. From experiments, ringwoodite may incorporate up to 2.5 wt% H_2O under transition-zone conditions[2,4,30,31]. The difficulties in constraining sample thickness during FTIR measurement, especially in determining whether the beam was sampling part of the Ca-walstromite inclusion, plus any spectral absorption by the rather impure diamond host, make the estimation of the ringwoodite water content subject to large uncertainty. The main OH^- stretching band at ~3,150 cm^{-1} in hydrous ringwoodite becomes more pronounced with increasing H_2O content, up to ~0.8 wt% H_2O (ref. 4; Fig. 2). The JUc29 spectra show strong OH^- absorption, clearly indicative of significant H_2O content, and are consistent with a minimum estimate between 1.4 and 1.5 wt% H_2O, derived by integrating the spectra in Fig. 2 (see Methods section on FTIR spectroscopy). Although the uncertainty in these estimates may be as large as 50%, we note that in synthetic ringwoodites containing 2 wt% H_2O or more, the satellite OH^- stretching mode at 3,645–3,680 cm^{-1} transforms from a broad shoulder to a sharply defined vibrational band[30]. This stretching mode is well defined in the JUc29 inclusion, supporting our calculated water concentration as a minimum estimate.

Two main scenarios arise from the water-rich nature of the ringwoodite inclusion coming from transition-zone depths. In one, water within the ringwoodite reflects inheritance from a hydrous, diamond-forming fluid, from which the inclusion grew as a syngenetic phase. In this model, the hydrous fluid must originate locally, from the transition zone, because there is no evidence that the lower mantle contains a significant amount of water. Alternatively, the ringwoodite is 'protogenetic', that is, it was present before encapsulation by the diamond and its water content reflects that of the ambient transition zone. Both

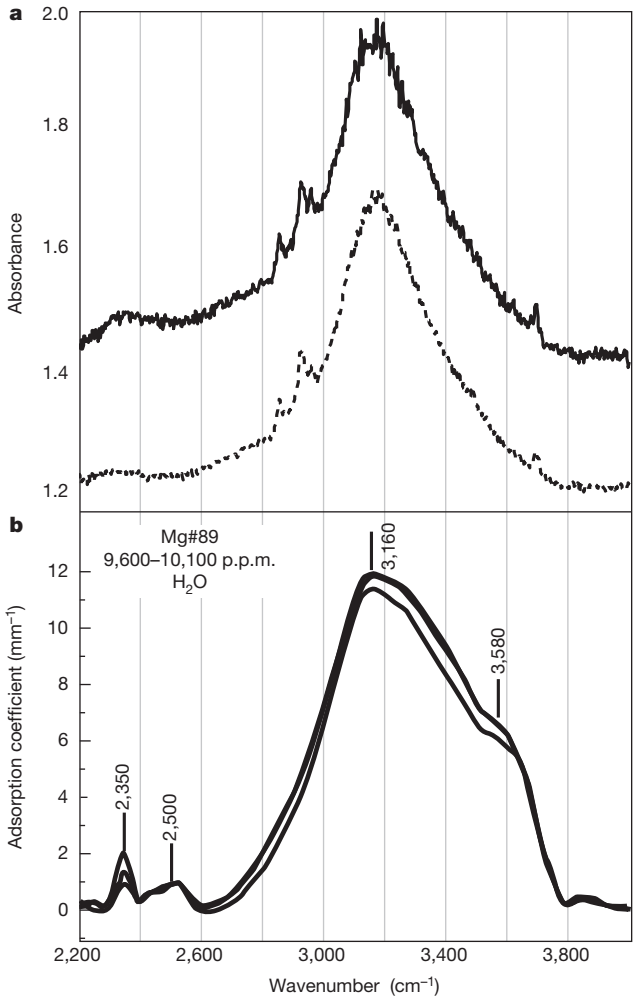


Figure 2 | FTIR spectra of ringwoodite inclusion in Juína diamond JUc29. **a**, Unpolarized FTIR spectra for ringwoodite inclusion in diamond JUc29 between 2,200 and 3,900 cm^{-1}. The two spectra were measured at ~90° degrees to each other and are unsmoothed, but were corrected for a background that includes the intrinsic response of the host diamond. Water contents calculated by integration of these two spectra are between 1.4 and 1.5 wt% (see Methods section on FTIR). **b**, Reference spectra for hydrous Fe-bearing ringwoodite (Mg# 89) containing ~1 wt% H_2O (ref. 27).

models implicate a transition zone that is at least locally water-rich. It is interesting to explore the protogenetic option further to see what bounds would be placed on the bulk transition-zone water content in the light of geophysical observations.

Using a conservative estimate of the H_2O content of JUc29 ringwoodite, of 1.4 wt%, combined with mineral mode estimates[12] and water solubilities for majorite and Ca-perovskite[6], results in a bulk water content of ~1.0 wt% for the transition zone sampled by our diamond. This value is broadly aligned with the highest transition-zone water contents estimated from electromagnetic data[11,32]. Other studies of ultradeep diamonds have indicated the transition zone could contain stagnated subducted slabs[17–26] that may transport water to this mantle region. The presence of hydrous ringwoodite in a diamond from transition-zone depths supports the view that high fluid activity, notably that of water, has a key role in the genesis of ultradeep diamonds[26] and is consistent with the proposal of regionally localized 'wet-spots' in the transition zone[32] that may host thin melt layers above the 410-km discontinuity[7]. Our observations provide clear support for experimental measurements[27] showing that the *P*- and *S*-wave velocities of the lower transition zone are consistent with a hydrated ringwoodite-rich composition. The preservation of ringwoodite within diamond also provides a strong indication that some kimberlites must come from at least transition-zone depths.

METHODS SUMMARY

Full descriptions of all analytical methods and calculations of the compositional estimates are provided in Methods. Micro-X-ray fluorescence measurements were performed at beamline L of the DORIS-III synchrotron facility at HASYLAB (DESY, Germany). Measurements were made using confocal detection of an internal microscopic volume element of approximately 22 μm × 22 μm × 16 μm (full-width at half-maximum). Single-crystal X-ray diffraction was performed at the Dipartimento di Geoscienze, Università di Padova, Italy using a CCD detector coupled to a STOE STADI IV single-crystal diffractometer, via monochromatized $Mo_{K\alpha}$ radiation ($\lambda = 0.71073$ Å), working at 50 kV and 40 mA and with an exposure time of 60 s. We obtained the main four diffraction peaks of ringwoodite (RINGW: Extended Data Fig. 3), that is, the planes (113) at 2.44 Å, (440) at 1.40 Å, (220) at 2.81 Å and (115) at 1.51 Å, in the expected order of relative intensity. Raman spectroscopy was carried out at the Geoscience Institute, Goethe University, Germany, using a Renishaw micro-Raman spectrometer (RM-1000) equipped with a Leica DMLM optical microscope and CCD detector. Spectra were excited with the He–Ne 632.8-nm line (max 50 mW). The wavenumber accuracy was 0.5 cm^{-1} and the spectral resolution was ~1 cm^{-1}. The lateral resolution at the sampling depth was several micrometres and the depth resolution was several tens of micrometres. Details of the calculation of the ringwoodite composition from the Raman spectra are given in Methods. FTIR spectra were obtained with a Nicolet Continuμm infrared microscope attached to a Thermo Nicolet Nexus 470 FTIR Spectrometer at the De Beers Laboratory of Diamond Research, University of Alberta, Canada. All measurements were performed in transmitted mode, with an unpolarized beam of aperture size 70 μm. Two hundred scans were acquired with a spectral resolution 4 cm^{-1}. Details of the calculation of water content from the spectra are given in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 September 2013; accepted 21 January 2014.

1. Smyth, J. R. β-Mg_2SiO_4: a potential host for water in the mantle? *Am. Mineral.* **72,** 1051–1055 (1987).
2. Chen, J., Inoue, T., Yurimoto, H. & Weidner, D. J. Effect of water on olivine- wadsleyite phase boundary in the $(Mg,Fe)_2SiO_4$ system. *J. Geophys. Res. Lett.* **29,** 1875 (2002).
3. Kohlstedt, D. L., Keppler, H. & Rubie, D. C. Solubility of water in the α, β and γ phases of $(Mg,Fe)_2SiO_4$. *Contrib. Mineral. Petrol.* **123,** 345–357 (1996).
4. Smyth, J. R. *et al.* Structural systematics of hydrous ringwoodite and water in Earth's interior. *Am. Mineral.* **88,** 1402–1407 (2003).
5. Bercovici, D. & Karato, S. Whole-mantle convection and the transition zone water filter. *Nature* **425,** 39–44 (2003).
6. Bolfan-Casanova, N. Water in the Earth's mantle. *Mineral. Mag.* **69,** 229–258 (2005).
7. Hirschmann, M. Water, melting and the deep Earth H_2O cycle. *Annu. Rev. Earth Planet. Sci.* **34,** 629–653 (2006).
8. Huang, X., Xu, Y. & Karato, S. Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite. *Nature* **434,** 746–749 (2005).
9. Xu, Y., Shankland, T. J. & Rubie, D. C. Electrical conductivity of olivine, wadsleyite and ringwoodite under upper-mantle conditions. *Science* **280,** 1415–1418 (1998).
10. Yoshino, T., Manthilake, G., Matsuzaki, T. & Katsura, T. Dry mantle transition zone inferred from the conductivity of wadsleyite and ringwoodite. *Nature* **451,** 326–329 (2008).
11. Khan, A. & Shankland, T. J. A geological perspective on mantle water content and melting: inverting electromagnetic sounding data using laboratory-based electrical conductivity profiles. *Earth Planet. Sci. Lett.* **317–318,** 27–43 (2012).
12. Agee, C. B. in *Ultrahigh-Pressure Mineralogy* (ed. Hemley, R. J.) 165–203 (Rev. Mineral. 37, Mineralogical Society of America, 1998).
13. Ringwood, A. E. & Major, A. The system Mg_2SiO_4-Fe_2SiO_4 at high pressures and temperatures. *Phys. Earth Planet. Inter.* **3,** 89–108 (1970).
14. Ye, Y. *et al.* Compressibility and thermal expansion of hydrous ringwoodite with 2.5 wt% H_2O. *Am. Mineral.* **97,** 573–582 (2012).
15. Harris, J. W., Hutchison, M. T., Hursthouse, M., Light, M. & Harte, B. A new tetragonal silicate mineral occurring as inclusions in lower mantle diamonds. *Nature* **387,** 486–488 (1997).
16. Harte, B. & Harris, J. W. Lower mantle associations preserved in diamonds. *Mineral. Mag.* **58A,** 384–385 (1994).
17. Harte, B., Harris, J. W., Hutchison, M. T., Watt, G. R. & Wilding, M. C. in *Mantle Petrology: Field Observations and High Pressure Experimentation* Vol. 6 (eds Fei, Y. & Bertka, C. M.) 125–153 (Geochem. Soc. Spec. Publ., The Geochemical Society, 1999).

18. Hutchison, M. T., Cartigny, P. & Harris, J. W. in *Proc. 7th Int. Kimberlite Conf.* (eds Gurney, J. J., Gurney, J. L., Pascoe, M. D. & Richardson, S. H.) 372–382 (Red Roof Design, 1999).
19. Hutchison, M. T., Hursthouse, M. B. & Light, M. E. Mineral inclusions in diamonds: associations and chemical distinctions around the 670-km discontinuity. *Contrib. Mineral. Petrol.* **142,** 119–126 (2001).
20. Brenker, F. *et al.* Detection of a Ca-rich lithology in the deep (>300 km) convecting mantle. *Earth Planet. Sci. Lett.* **236,** 579–587 (2005).
21. Stachel, T., Brey, G. P. & Harris, J. W. Inclusions in sublithospheric diamonds: glimpses of deep Earth. *Elements* **1,** 73–78 (2005).
22. Stachel, T., Harris, J. W., Brey, G. P. & Joswig, W. Kankan diamonds (Guinea) II: lower mantle inclusion paragenesis. *Contrib. Mineral. Petrol.* **140,** 16–27 (2000).
23. Bulanova, G. *et al.* Mineral inclusions in sublithospheric diamonds from Collier 4 kimberlite pipe, Juina, Brazil: subducted protoliths, carbonated melts and primary kimberlite magmatism. *Contrib. Mineral. Petrol.* **160,** 489–510 (2010).
24. Hayman, P. C., Kopylova, M. G. & Kaminsky, F. V. Lower mantle diamonds from Rio Soriso (Juina area, Mato Grosso, Brazil). *Contrib. Mineral. Petrol.* **149,** 430–445 (2005).
25. Tappert, R., Stachel, T., Harris, J. W., Shimizu, N. & Brey, G. P. Mineral inclusions in diamonds from the Panda kimberlite, Slave Province, Canada. *Eur. J. Mineral.* **17,** 423–440 (2005).
26. Harte, B. Diamond formation in the deep mantle: the record of mineral inclusions and their distribution in relation to mantle dehydration zones. *Mineral. Mag.* **74,** 189–215 (2010).
27. Jacobsen, S. D., Smyth, J. R., Spetzler, H., Holl, C. M. & Frost, D. J. Sound velocities and elastic constants of iron-bearing hydrous ringwoodite. *Phys. Earth Planet. Inter.* **143–144,** 47–56 (2004).
28. Blanchard, M., Balan, E. & Wright, K. Incorporation of water in iron-free ringwoodite: a first principles study. *Am. Mineral.* **94,** 83–89 (2009).
29. Bolfan-Casanova, N., Keppler, H. & Rubie, D. C. Water partitioning between nominally anhydrous minerals in the MgO-SiO_2-H_2O system up to 24 GPa: implications for the distribution of water in the Earth's mantle. *Earth Planet. Sci. Lett.* **182,** 209–221 (2000).
30. Keppler, H. & Bolfan-Casanova, N. in *Water in Nominally Anhydrous Minerals* (eds Keppler, H. & Smyth, J. R.) 193–230 (Rev. Mineral. 62, Mineralogical Society of America, 2006).
31. Kleppe, A. K., Jephcoat, A. P. & Smyth, J. R. Raman spectroscopic study of hydrous γ-Mg_2SiO_4 to 56.5 GPa. *Phys. Chem. Miner.* **29,** 473–476 (2002).
32. Kelbert, A., Schultz, A. & Egbert, G. Global electromagnetic induction constraints on transition-zone water content variations. *Nature* **460,** 1003–1006 (2009).
33. Chen, M., El Goresy, A. & Gillet, P. Ringwoodite lamellae in olivine: clues to olivine–ringwoodite phase transition mechanisms in shocked meteorites and subducting slabs. *Proc. Natl Acad. Sci. USA* **101,** 15033–15037 (2004).

Acknowledgements D.G.P. acknowledges CERC funding for this study. F.N. is supported by ERC Starting Grant 307322. Support from the Alfred P. Sloan Foundation's Deep Carbon Observatory project created this research partnership. We thank T. Stachel for comments on the manuscript plus access to the FTIR instrument at the De Beers Laboratory of Diamond Research at the University of Alberta, and we thank J. Harris for discussions. Sample JUc29 was provided by Trigon GeoServices Ltd.

Author Contributions D.G.P. had the idea for the study, wrote the manuscript and helped perform the Raman and FTIR measurements. F.E.B. performed the Raman measurements and ion-milling and made compositional estimates. F.N. performed X-ray measurements. J.M. and L.N. first identified the inclusion as ringwoodite. M.T.H. selected the diamond for this study and assisted with manuscript preparation and geological background. S.M. performed the FTIR measurements and the water content estimate. K.M. assisted with manuscript preparation. G.S., S.S., B.V. and L.V. performed the synchrotron X-ray mapping measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.G.P. (gdpearso@ualberta.ca).

LETTER

doi:10.1038/nature12960

Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European

Iñigo Olalde[1]*, Morten E. Allentoft[2]*, Federico Sánchez-Quinto[1], Gabriel Santpere[1], Charleston W. K. Chiang[3], Michael DeGiorgio[4,5], Javier Prado-Martinez[1], Juan Antonio Rodríguez[1], Simon Rasmussen[6], Javier Quilez[1], Oscar Ramírez[1], Urko M. Marigorta[1], Marcos Fernández-Callejo[1], María Encina Prada[7], Julio Manuel Vidal Encinas[8], Rasmus Nielsen[9], Mihai G. Netea[10], John Novembre[11], Richard A. Sturm[12], Pardis Sabeti[13,14], Tomàs Marquès-Bonet[1,15], Arcadi Navarro[1,15,16,17], Eske Willerslev[2] & Carles Lalueza-Fox[1]

Ancient genomic sequences have started to reveal the origin and the demographic impact of farmers from the Neolithic period spreading into Europe[1–3]. The adoption of farming, stock breeding and sedentary societies during the Neolithic may have resulted in adaptive changes in genes associated with immunity and diet[4]. However, the limited data available from earlier hunter-gatherers preclude an understanding of the selective processes associated with this crucial transition to agriculture in recent human evolution. Here we sequence an approximately 7,000-year-old Mesolithic skeleton discovered at the La Braña-Arintero site in León, Spain, to retrieve a complete pre-agricultural European human genome. Analysis of this genome in the context of other ancient samples suggests the existence of a common ancient genomic signature across western and central Eurasia from the Upper Paleolithic to the Mesolithic. The La Braña individual carries ancestral alleles in several skin pigmentation genes, suggesting that the light skin of modern Europeans was not yet ubiquitous in Mesolithic times. Moreover, we provide evidence that a significant number of derived, putatively adaptive variants associated with pathogen resistance in modern Europeans were already present in this hunter-gatherer.

Next-generation sequencing (NGS) technologies are revolutionizing the field of ancient DNA (aDNA), and have enabled the sequencing of complete ancient genomes[5,6], such as that of Ötzi, a Neolithic human body found in the Alps[1]. However, very little is known of the genetic composition of earlier hunter-gatherer populations from the Mesolithic period (circa 10,000–5,000 years before present, BP; immediately preceding the Neolithic period).

The Iberian site called La Braña-Arintero was discovered in 2006 when two male skeletons (named La Braña 1 and 2) were found in a deep cave system, 1,500 m above sea level in the Cantabrian mountain range (León, Northwestern Spain) (Fig. 1a). The skeletons were dated to approximately 7,000 years BP (7,940–7,690 calibrated BP)[7]. Because of the cold environment and stable thermal conditions in the cave, the preservation of these specimens proved to be exceptional (Fig. 1b). We identified a tooth from La Braña 1 with high human DNA content (48.4%) and sequenced this specimen to a final effective genomic depth-of-coverage of 3.40× (Extended Data Fig. 1).

We used several tests to assess the authenticity of the genome sequence and to determine the amount of potential modern human contamination. First, we observed that sequence reads from both the mitochondrial

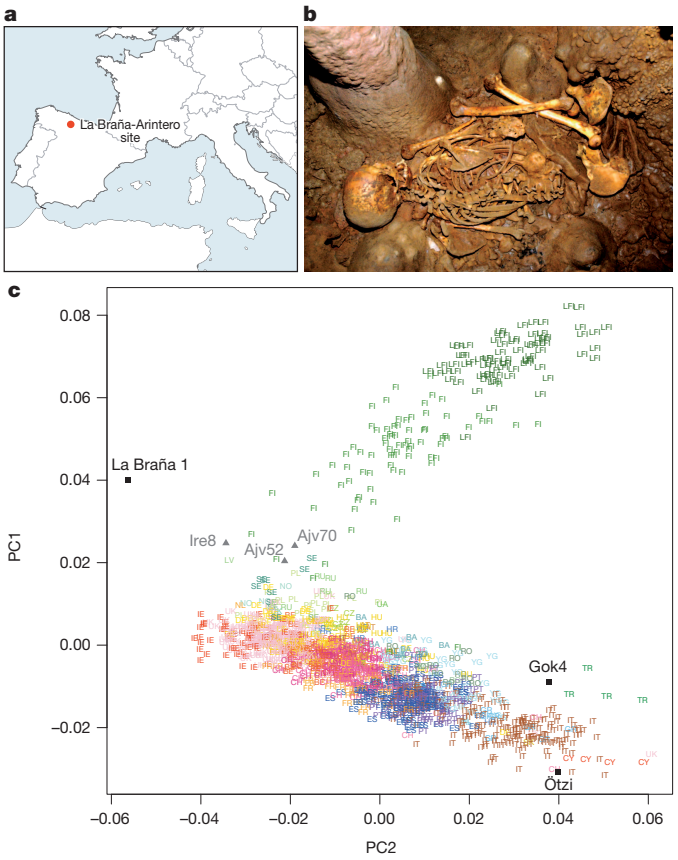


Figure 1 | Geographic location and genetic affinities of the La Braña 1 individual. **a**, Location of the La Braña-Arintero site (Spain). **b**, The La Braña 1 skeleton as discovered in 2006. **c**, PCA based on the average of the Procrustes transformations of individual PCAs with La Braña 1 and each of the five Neolithic samples[1,3]. The reference populations are the Finnish HapMap, FINHM and POPRES. Population labels with labelling of ref. 12 with the addition of FI (Finns) or LFI (late-settlement Finns). Ajv70, Ajv52, Ire8 and Gok4 are Scandinavian Neolithic hunter-gatherers and a farmer, respectively[3]. Ötzi is the Tyrolean Ice Man[1].

[1]Institut de Biologia Evolutiva, CSIC-UPF, Barcelona 08003, Spain. [2]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen K, Denmark. [3]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095, USA. [4]Department of Integrative Biology, University of California, Berkeley, California 94720, USA. [5]Department of Biology, Pennsylvania State University, 502 Wartik Laboratory, University Park, Pennsylvania 16802, USA. [6]Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. [7]I.E.S.O. 'Los Salados', Junta de Castilla y León, E-49600 Benavente, Spain. [8]Junta de Castilla y León, Servicio de Cultura de León, E-24071 León, Spain. [9]Center for Theoretical Evolutionary Genomics, University of California, Berkeley, California 94720, USA. [10]Department of Medicine and Nijmegen Institute for Infection, Inflammation and Immunity, Radboud University Nijmegen Medical Centre, 6500 Nijmegen, The Netherlands. [11]Department of Human Genetics, University of Chicago, Illinois 60637, USA. [12]Institute for Molecular Bioscience, Melanogenix Group, The University of Queensland, Brisbane, Queensland 4072, Australia. [13]Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [14]Broad Institute of the Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, USA. [15]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain. [16]Centre de Regulació Genòmica (CRG), Barcelona 08003, Catalonia, Spain. [17]National Institute for Bioinformatics (INB), Barcelona 08003, Catalonia, Spain.
*These authors contributed equally to this work.

DNA (mtDNA) and the nuclear DNA of La Braña 1 showed the typical ancient DNA misincorporation patterns that arise from degradation of DNA over time[8] (Extended Data Fig. 2a, b). Second, we showed that the observed number of human DNA fragments was negatively correlated with the fragment length ($R^2 > 0.92$), as expected for ancient degraded DNA, and that the estimated rate of DNA decay was low and in agreement with predicted values[9] (Extended Data Fig. 2c, d). We then estimated the contamination rate in the mtDNA genome, assembled to a high depth-of-coverage (91×), by checking for positions differing from the mtDNA genome (haplogroup U5b2c1) that was previously retrieved with a capture method[2]. We obtained an upper contamination limit of 1.69% (0.75–2.6%, 95% confidence interval, CI) (Supplementary Information). Finally, to generate a direct estimate of nuclear DNA contamination, we screened for heterozygous positions (among reads with >4× coverage) in known polymorphic sites (Single Nucleotide Polymorphism Database (dbSNP) build 137) at uniquely mapped sections on the X chromosome[6] (Supplementary Information). We found that the proportion of false heterozygous sites was 0.31%. Together these results suggest low levels of contamination in the La Braña 1 sequence data.

To investigate the relationship to extant European samples, we conducted a principal component analysis (PCA)[10] and found that the approximately 7,000-year-old Mesolithic sample was divergent from extant European populations (Extended Data Fig. 3a, b), but was placed in proximity to northern Europeans (for example, samples from Sweden and Finland)[11–14]. Additional PCAs and allele-sharing analyses with ancient Scandinavian specimens[3] supported the genetic similarity of the La Braña 1 genome to Neolithic hunter-gatherers (Ajv70, Ajv52, Ire8) relative to Neolithic farmers (Gok4, Ötzi) (Fig. 1c, Extended Data Figs 3c and 4). Thus, this Mesolithic individual from southwestern Europe represents a formerly widespread gene pool that seems to be partially preserved in some modern-day northern European populations, as suggested previously with limited genetic data[2,3]. We subsequently explored the La Braña affinities to an ancient Upper Palaeolithic genome from the Mal'ta site near Lake Baikal in Siberia[15]. Outgroup f_3 and D statistics[16,17], using different modern reference populations, support that Mal'ta is significantly closer to La Braña 1 than to Asians or modern Europeans (Extended Data Fig. 5 and Supplementary Information). These results suggest that despite the vast geographical distance and temporal span, La Braña 1 and Mal'ta share common genetic ancestry, indicating a genetic continuity in ancient western and central Eurasia. This observation matches findings of similar cultural artefacts across time and space in Upper Paleolithic western Eurasia and Siberia, particularly the presence of anthropomorphic 'Venus' figurines that have been recovered from several sites in Europe and Russia, including the Mal'ta site[15]. We also compared the genome-wide heterozygosity of the La Braña 1 genome to a data set of modern humans with similar coverage (3–4×). The overall genomic heterozygosity was 0.042%, lower than the values observed in present day Asians (0.046–0.047%), Europeans (0.051–0.054%) and Africans (0.066–0.069%) (Extended Data Fig. 6a). The effective population size, estimated from heterozygosity patterns, suggests a global reduction in population size of approximately 20% relative to extant Europeans (Supplementary Information). Moreover, no evidence of tracts of autozygosity suggestive of inbreeding was observed (Extended Data Fig. 6b).

To investigate systematically the timing of selection events in the recent history of modern Europeans, we compared the La Braña genome to modern populations at loci that have been categorized as of interest for their role in recent adaptive evolution. With respect to two recent well-studied adaptations to changes in diet, we found the ancient genome to carry the ancestral allele for lactose intolerance[4] and approximately five copies of the salivary amylase (*AMY1*) gene (Extended Data Fig. 7 and Supplementary Information), a copy number compatible with a low-starch diet[18]. These results suggest the La Braña hunter-gatherer was poor at digesting milk and starch, supporting the hypotheses that these abilities were selected for during the later transition to agriculture.

To expand the survey, we analysed a catalogue of candidate signals for recent positive selection based on whole-genome sequence variation from the 1000 Genomes Project[13], which included 35 candidate non-synonymous variants, ten of which were detected uniquely in the CEU (Utah residents with northern and western European ancestry) sample [19]. For each variant we assessed whether the Mesolithic genome carried the ancestral or derived (putatively adaptive) allele.

Of the ten variants, the Mesolithic genome carried the ancestral and non-selected allele as a homozygote in three regions: *C12orf29* (a gene with unknown function), *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) (Table 1). The latter two variants are the two strongest known loci affecting light skin pigmentation in Europeans[20–22] and their ancestral alleles and associated haplotypes are either absent or segregate at very low frequencies in extant Europeans (3% and 0% for *SLC45A2* and *SLC24A5*, respectively) (Fig. 2). We subsequently examined all genes known to be associated with pigmentation in Europeans[22], and found ancestral alleles in *MC1R*, *TYR* and *KITLG*, and derived alleles in *TYRP1*, *ASIP* and *IRF4* (Supplementary Information). Although the precise phenotypic effects cannot currently be ascertained in a European genetic background, results from functional experiments[20] indicate that the allelic combination in this Mesolithic individual is likely to have resulted in dark skin pigmentation and dark or brown hair. Further examination revealed that this individual carried the *HERC2* rs12913832*C single nucleotide polymorphism (SNP) and the associated homozygous haplotype spanning the *HERC2–OCA2* locus that is strongly associated

Table 1 | Mesolithic genome allelic state at 10 nonsynonymous variants recently selected in Europeans

Allelic state	Gene	Name	SNP	Amino-acid change	Function
La Braña 1 carries the derived allele	*PTX4*	Pentraxin 4	rs2745098	Arg281Lys	May be involved in innate immunity
	UHRF1BP1	UHRF1 binding protein 1	rs11755393	Gln454Arg	Risk locus for systemic lupus erythematosus
	GPATCH1	G patch domain containing 1	rs10421769	Leu520Ser	Receptor for OmpA expressed by *E. coli*
	WWOX	WW domain-containing oxidoreductase	rs12918952	Ala179Thr	Acts as a tumour suppressor and has a role in apoptosis
	CCDC14	Coiled-coil domain-containing protein 14	rs17310144	Thr365Pro	Unknown
La Braña 1 carries both the ancestral and the derived allele	*SETX*	Senataxin	rs1056899	Val2587Ile	Involved in spinocerebellar ataxia and amyotrophic lateral sclerosis
	TDRD12	Tudor domain containing 12	rs11881633	Glu413Lys	Unknown
La Braña 1 retains the ancestral allele	*C12orf29*	Chromosome 12 open reading frame 29	rs9262	Val238Leu	Unknown
	SLC45A2	Solute carrier family 45, member 2	rs16891982	Leu374Phe	Associated with skin pigmentation
	SLC24A5	Solute carrier family 24, member 5	rs1426654	Ala111Thr	Associated with skin pigmentation

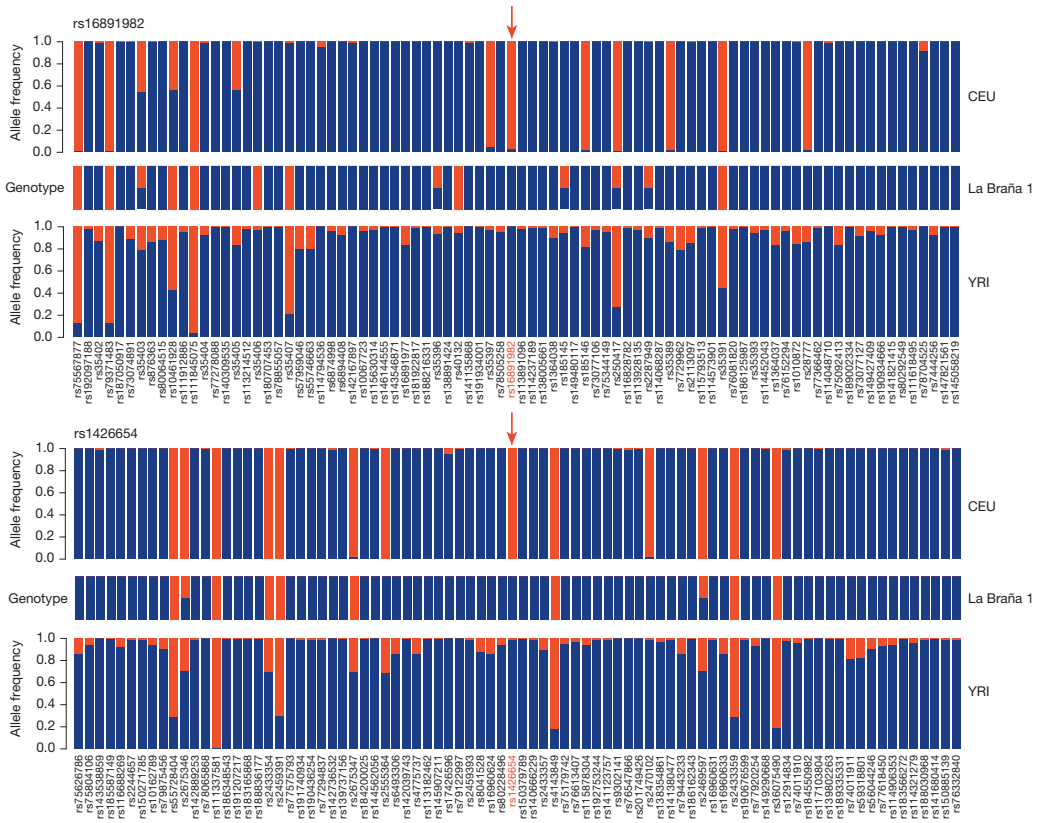


Figure 2 | Ancestral variants around the *SLC45A2* (rs16891982, above) and *SLC24A5* (rs1426654, below) pigmentation genes in the Mesolithic genome. The SNPs around the two diagnostic variants (red arrows) in these two genes were analysed. The resulting haplotype comprises neighbouring SNPs that are also absent in modern Europeans (CEU) (n=112) but present in Yorubans (YRI) (n=113). This pattern confirms that the La Braña 1 sample is older than the positive-selection event in these regions. Blue, ancestral; red, derived.

with blue eye colour[23]. Moreover, a prediction of eye colour based on genotypes at additional loci using HIrisPlex[24] produced a 0.823 maximal and 0.672 minimal probability for being non-brown-eyed (Supplementary Information). The genotypic combination leading to a predicted phenotype of dark skin and non-brown eyes is unique and no longer present in contemporary European populations. Our results indicate that the adaptive spread of light skin pigmentation alleles was not complete in some European populations by the Mesolithic, and that the spread of alleles associated with light/blue eye colour may have preceded changes in skin pigmentation.

For the remaining loci, La Braña 1 displayed the derived, putatively adaptive variants in five cases, including three genes, *PTX4, UHRF1BP1* and *GPATCH1 (*ref. 19), involved in the immune system (Table 1 and Extended Data Fig. 8). *GPATCH1* is associated with the risk of bacterial infection. We subsequently determined the allelic states in 63 SNPs from 40 immunity genes with previous evidence for positive selection and for carrying polymorphisms shown to influence susceptibility to infections in modern Europeans (Supplementary Information). La Braña 1 carries derived alleles in 24 genes (60%) that have a wide range of functions in the immune system: pattern recognition receptors, intracellular adaptor molecules, intracellular modulators, cytokines and cytokine receptors, chemokines and chemokine receptors and effector molecules. Interestingly, four out of six SNPs from the first category are intracellular receptors of viral nucleic acids (*TLR3*, *TLR8*, *IFIH1* (also known as *MDA5*) and *LGP2*)[25].

Finally, to explore the functional regulation of the genome, we also assessed the La Braña 1 genotype at all expression quantitative trait loci (eQTL) regions associated to positive selection in Europeans (Supplementary Information). The most interesting finding is arguably the predicted overexpression of eight immunity genes (36% of those with described eQTLs), including three Toll-like receptor genes (*TLR1*, *TLR2* and *TLR4*) involved in pathogen recognition[26].

These observations suggest that the Neolithic transition did not drive all cases of adaptive innovation on immunity genes found in modern Europeans. Several of the derived haplotypes seen at high frequency today in extant Europeans were already present during the Mesolithic, as neutral standing variation or due to selection predating the Neolithic. *De novo* mutations that increased in frequency rapidly in response to zoonotic infections during the transition to farming should be identified among those genes where La Braña 1 carries ancestral alleles.

To confirm whether the genomic traits seen at La Braña 1 can be generalized to other Mesolithic populations, analyses of additional ancient genomes from central and northern Europe will be needed. Nevertheless, this genome sequence provides the first insight as to how these hunter-gatherers are related to contemporary Europeans and other ancient peoples in both Europe and Asia, and shows how ancient DNA can shed light on the timing and nature of recent positive selection.

METHODS SUMMARY

DNA was extracted from the La Braña 1 tooth specimen with a previously published protocol[2]. Indexed libraries were built from the ancient extract and sequenced on the Illumina HiSeq platform. Reads generated were mapped with BWA[27] to the human reference genome (NCBI 37, hg19) after primer trimming. A metagenomic analysis and taxonomic identification was generated with the remaining reads using BLAST 2.2.27+ and MEGAN4 (ref. 28) (Extended Data Fig. 9). SNP calling was undertaken using a specific bioinformatic pipeline designed to account for ancient DNA errors. Specifically, the quality of misincorporations likely caused by ancient DNA damage was rescaled using the mapDamage2.0 software[29], and a set of variants with a minimum read depth of 4 was produced with GATK[30]. Analyses including PCA[10], Outgroup f_3[16] and D statistics[17] were performed to determine the population affinities of this Mesolithic individual (Supplementary Information).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 October; accepted 17 December 2013.

Published online 26 January 2014.

1. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Commun.* **3,** 698 (2012).
2. Sánchez-Quinto, F. *et al.* Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22,** 1494–1499 (2012).
3. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336,** 466–469 (2012).
4. Laland, K. N., Odling-Smee, J. & Myles, S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Rev. Genet.* **11,** 137–148 (2010).
5. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757–762 (2010).
6. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334,** 94–98 (2011).
7. Vidal Encinas, J. M. & Prada Marcos, M. E. *Los hombres mesolíticos de La Braña-Arintero (Valdelugueros, León)* (León: Junta de Castilla y León, 2010).
8. Overballe-Petersen, S., Orlando, L. & Willerslev, E. Next-generation sequencing offers new insights into DNA degradation. *Trends Biotechnol.* **30,** 364–368 (2012).
9. Allentoft, M. E. *et al.* The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B Biol. Sci.* **279,** 4824–4733 (2012).
10. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
11. Nelson, M. R. *et al.* The population reference sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* **83,** 347–358 (2008).
12. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456,** 98–101 (2008).
13. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
14. Surakka, I. *et al.* Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging. *Genome Res.* **20,** 1344–1351 (2010).
15. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
16. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489–494 (2009).
17. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
18. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39,** 1256–1260 (2007).
19. Grossman, S. R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152,** 703–713 (2013).
20. Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310,** 1782–1786 (2005).
21. Norton, H. L. *et al.* Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24,** 710–722 (2007).
22. Sturm, R. A. & Duffy, D. L. Human pigmentation genes under environmental selection. *Genome Biol.* **13,** 248 (2012).
23. Sturm, R. A. *et al.* A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82,** 424–431 (2008).
24. Walsh, S. *et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7,** 98–115 (2013).
25. Aoshi, T., Koyama, S., Kobiyama, K., Akira, S. & Ishii, K. J. Innate and adaptive immune responses to viral infection and vaccination. *Curr. Opin. Virol.* **1,** 226–232 (2011).
26. Moresco, E. M. Y., LaVine, D. & Beutler, B. Toll-like receptors. *Curr. Biol.* **21,** R488–R493 (2011).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
28. Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21,** 1552–1560 (2011).
29. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29,** 1682–1684 (2013).
30. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank L. A. Grau Lobo (Museo de León) for access to the La Braña specimen, M. Rasmussen and H. Schroeder for valid input into the experimental work, and M. Raghavan for early access to Mal'ta genome data. Sequencing was performed at the Danish National High-Throughput DNA-Sequencing Centre, University of Copenhagen. The POPRES data were obtained from dbGaP (accession number 2038). The authors are grateful for financial support from the Danish National Research Foundation, ERC Starting Grant (260372) to TM-B, and (310372) to M.G.N., FEDER and Spanish Government Grants BFU2012-38236, the Spanish Multiple Sclerosis Netowrk (REEM) of the Instituto de Salud Carlos III (RD12/0032/0011) to A.N., BFU2011-28549 to T.M.-B., BFU2012-34157 to C.L.-F., ERC (Marie Curie Actions 300554) to M.E.A., NIH NRSA postdoctoral fellowship (F32GM106656) to C.W.K.C., NIH (R01-HG007089) to J.N., NSF postdoctoral fellowship (DBI-1103639) to M.D., the Australian NHMRC to R.A.S. and a predoctoral fellowship from the Basque Government (DEUI) to I.O.

Author Contributions C.L.-F. and E.W. conceived and lead the project. M.E.P. and J.M.V.E. provided anthropological and archaeological information. O.R. and M.E.A. performed the ancient extractions and library construction, respectively. I.O., M.E.A., F.S.-Q., J.P.-M., S.R., O.R., M.F.-C. and T.M.-B. performed mapping, SNP calling, mtDNA assembly, contamination estimates and different genomic analyses on the ancient genome. I.O., F.S.-Q., G.S., C.W.K.C., M.D., J.A.R., J.Q., O.R., U.M.M. and A.N. performed functional, ancestry and population genetic analyses. R.N. and J.N. coordinated the ancestry analyses. M.G.N., R.A.S. and P.S. coordinated the immunological, pigmentation and selection analyses, respectively. I.O., M.E.A., T.M.-B., E.W. and C.L.-F. wrote the majority of the manuscript with critical input from R.N., M.G.N., J.N., R.A.S., P.S. and A.N.

Author Information Alignment data are available through the Sequence Read Archive (SRA) under accession numbers PRJNA230689 and SRP033596. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W. (ewillerslev@snm.ku.dk) or C.L.-F. (carles.lalueza@upf.edu).

LETTER

doi:10.1038/nature13112

doublesex is a mimicry supergene

K. Kunte[1]*, W. Zhang[2]*, A. Tenger-Trolander[2], D. H. Palmer[3], A. Martin[4], R. D. Reed[4], S. P. Mullen[5] & M. R. Kronforst[2,3]

One of the most striking examples of sexual dimorphism is sex-limited mimicry in butterflies, a phenomenon in which one sex—usually the female—mimics a toxic model species, whereas the other sex displays a different wing pattern[1]. Sex-limited mimicry is phylogenetically widespread in the swallowtail butterfly genus *Papilio*, in which it is often associated with female mimetic polymorphism[1–3]. In multiple polymorphic species, the entire wing pattern phenotype is controlled by a single Mendelian 'supergene'[4]. Although theoretical work has explored the evolutionary dynamics of supergene mimicry[5–9], there are almost no empirical data that address the critical issue of what a mimicry supergene actually is at a functional level. Using an integrative approach combining genetic and association mapping, transcriptome and genome sequencing, and gene expression analyses, we show that a single gene, *doublesex*, controls supergene mimicry in *Papilio polytes*. This is in contrast to the long-held view that supergenes are likely to be controlled by a tightly linked cluster of loci[4]. Analysis of gene expression and DNA sequence variation indicates that isoform expression differences contribute to the functional differences between *dsx* mimicry alleles, and protein sequence evolution may also have a role. Our results combine elements from different hypotheses for the identity of supergenes, showing that a single gene can switch the entire wing pattern among mimicry phenotypes but may require multiple, tightly linked mutations to do so.

Wing pattern mimicry in butterflies, a phenomenon in which natural selection by predators causes unrelated species to evolve similar wing patterns, has served as an important model for studying adaptation since the earliest days of modern evolutionary theory[10]. Classical Batesian mimicry, in which an undefended mimic evolves to look like a toxic model, is a parasitic relationship in which the mimic gains an advantage at the expense of the model. Such systems have well-characterized frequency dependence[1,7], sometimes resulting in sexual dimorphism and mimetic polymorphism[1–3,8,11,12]. Swallowtail butterflies in the genus *Papilio* are well-known Batesian mimics, providing some of the most extreme examples of sexual dimorphism and polymorphism among living organisms[1,2,12]. For instance, in the species *Papilio polytes*, males all display the same non-mimetic wing pattern, whereas females display either a male-like pattern (form *cyrus*) or one of several different patterns that mimic toxic species in the genus *Pachliopta* (Fig. 1). Female wing pattern is polymorphic in local areas and there are no intermediate forms. The early crossing experiments of Clarke and Sheppard[13] revealed that variation in the entire wing pattern, as well as the presence versus absence of hindwing 'tails', is controlled by a single Mendelian locus, with female polymorphism resulting from multiple alleles, each with its place in a dominance hierarchy. Clarke and Sheppard also showed that the mimicry locus is autosomal, so sexual dimorphism is not directly mediated by sex linkage in this case[13].

This phenomenon, in which the entire wing pattern is controlled by a single Mendelian locus, is referred to as 'supergene' mimicry[4]. Because Clarke and Sheppard occasionally witnessed individuals with putatively recombinant wing patterns, they envisioned a supergene as a tightly linked cluster of loci, each controlling a distinct subset of the wing pattern. However, Clarke and Sheppard found virtually no evidence for recombination in *P. polytes*[13], although they did recover apparently recombinant phenotypes in other species, such as *P. memnon*[14]. Over the past few decades, supergene mimicry has received considerable theoretical attention[5–8], but there are almost no empirical data that address the molecular basis of a supergene. One example from *Heliconius* butterflies, which involves supergene mimicry but not sexual dimorphism, suggests that supergenes may be the result of chromosomal inversions that lock multiple adjacent genes into a single, non-recombining unit[15].

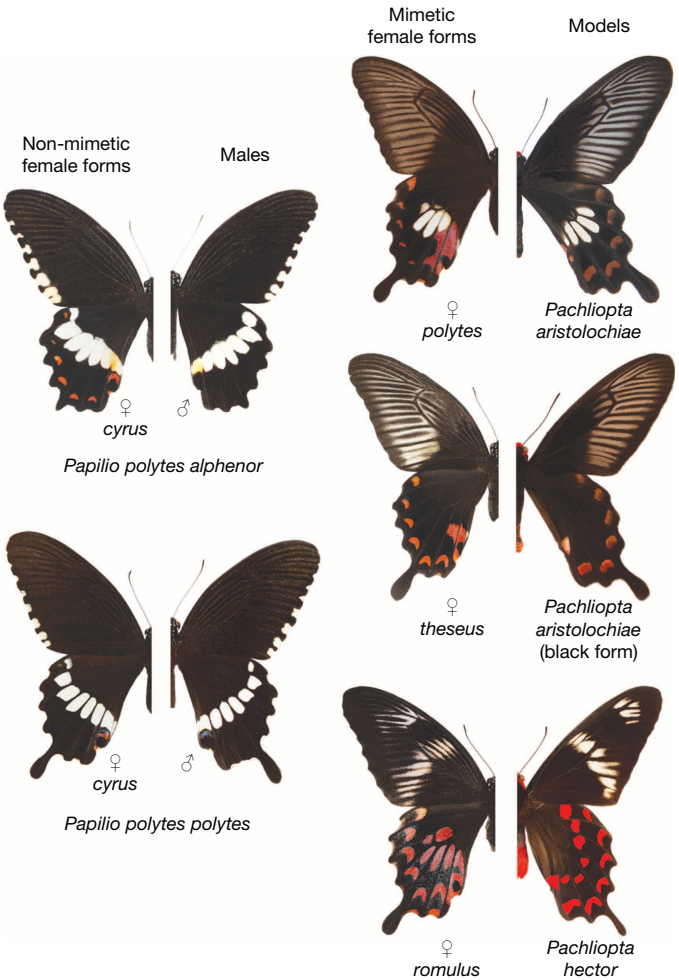


Figure 1 | Polymorphic, sex-limited mimicry in *Papilio polytes*. Non-mimetic (form *cyrus*) females look like males, whereas mimetic female morphs (forms *polytes*, *theseus* and *romulus*) mimic distantly related, toxic *Pachliopta* swallowtails. The presence of hindwing tails on males and *cyrus* females is variable among *P. polytes* populations. Our analyses focused on *P. polytes alphenor*, a group lacking tails on non-mimetic butterflies, and presence versus absence of tails segregated perfectly with female wing pattern in our crosses.

[1]National Center for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India. [2]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA. [3]Committee on Evolutionary Biology, University of Chicago, Chicago, Illinois 60637, USA. [4]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA. [5]Department of Biology, Boston University, Boston, Massachusetts 02215, USA.
*These authors contributed equally to this work.

Other evolutionary phenomena that involve supergene-like genetic architectures, such as self-incompatibility in plants and segregation distortion in *Drosophila melanogaster*, have also been traced back to multiple linked genes[16,17]. A second possibility is that a master regulator could gain control of the distinct networks that pattern various aspects of the wing, and hence control the entire phenotype from a single locus. Although this single-gene hypothesis has been discussed[6,9,18], there are no empirical data to support it.

Using a multi-step genetic mapping process that involved rearing nine F_2 backcross families (Fig. 2a), bulk segregant analysis with restriction-site associated DNA (RAD) markers, screening and sequencing bacterial artificial chromosome (BAC) clones, and fine-mapping, we mapped the mimicry locus in *P. polytes* back to a 300-kilobase (kb) region of the genome that contained five genes (Fig. 2b). We were intrigued to find that one of these genes was *doublesex* (*dsx*), a transcription factor in insects that controls somatic sex differentiation by alternative splicing[19,20]. In *Drosophila*, *dsx* is alternatively spliced into two isoforms: a male-specific form that leads to male sexual differentiation, and an alternative female form that causes female sexual differentiation[19–21]. In other insects, *dsx* functions the same way although there can be more than one male and female isoform[22,23].

On the basis of our mapping data and the known role of *dsx* in mediating sexual dimorphism[23–26], we proposed that *dsx* might control both the sex-limited and female polymorphism components of *P. polytes* mimicry. To test this hypothesis, we generated a reference genome sequence across our target interval and performed comprehensive association mapping by re-sequencing the genomes of 15 mimetic (form *polytes*) and 15 non-mimetic (form *cyrus*) butterflies (Extended Data Table 1). This yielded multiple perfect associations in *dsx* but only weak associations immediately outside of *dsx* (Fig. 2c). A separate genome-wide association study (GWAS) also yielded *dsx* as the top association hit (Extended Data Table 2). Long-term balancing selection, which maintains mimicry polymorphism[1,8,12], is expected to result in a localized excess of nucleotide variation driven by the accumulation of neutral substitutions on alternative alleles[27]. Analysis of DNA sequence variation revealed a highly significant excess of nucleotide polymorphism in *dsx*, relative to neighbouring genes (Table 1 and Extended Data Table 3), and comparisons between mimetic and non-mimetic individuals revealed over 1,000 nucleotide substitutions differentiating mimetic and non-mimetic *dsx* alleles (Table 1). This is in contrast to all neighbouring genes, which show little polymorphism and no fixed differences between mimicry forms.

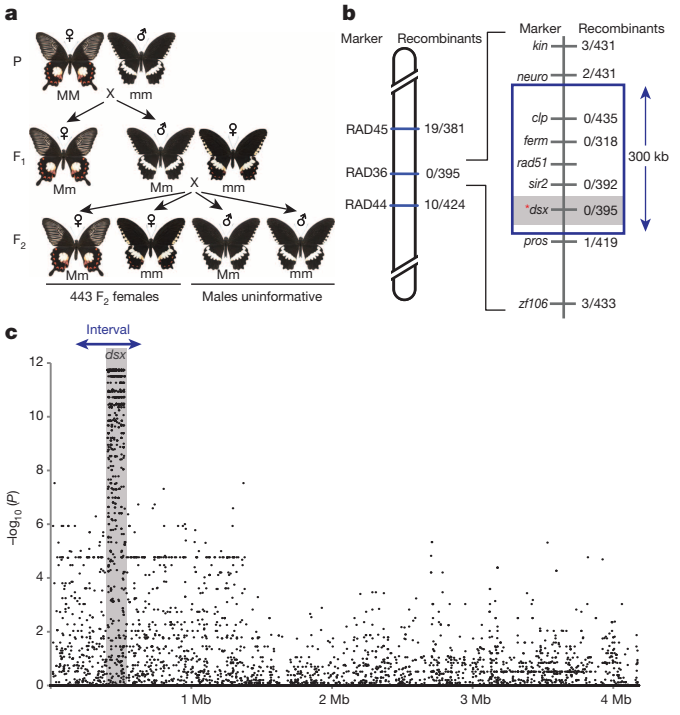


Figure 2 | Mapping the mimicry supergene. **a**, A series of nine backcross families yielded a total of 443 F_2 females that segregated 1:1 for female mimicry phenotype. **b**, Genome-wide mapping with RAD markers and subsequent fine-mapping localized the mimicry locus to a 300-kb interval containing five genes, one of which was *doublesex* (*dsx*). **c**, Association mapping, based on full genome sequences of 30 *P. polytes* butterflies, revealed multiple perfect associations inside *dsx* but none outside the gene. The positions of the 300-kb zero-recombinant interval and *dsx* are indicated. Data points represent false-discovery rate (FDR)-adjusted *P* values for a total of 94,776 SNPs.

The involvement of *dsx* as the mimicry supergene indicates a potential role for alternative splicing in the control of wing pattern. Transcriptome assembly based on wing-disc-derived RNA yielded three distinct female *dsx* isoforms and one male isoform (Fig. 3a). However, cloning and sequencing *dsx* isoforms from mimetic and non-mimetic males and females yielded the same repertoire of isoforms in butterflies with alternative mimicry alleles. Comparisons of isoform expression using quantitative reverse transcription PCR (qRT–PCR) revealed that all three female isoforms show strong female-biased expression (Fig. 3b–d). Two of these, isoforms 1 and 2, further showed pronounced wing-biased expression, whereas the third female isoform had body-biased expression (Fig. 3b–d). Comparisons between mimetic and non-mimetic females for wing-biased isoforms 1 and 2 revealed marked upregulation in mimetic females relative to non-mimetic females (Fig. 3e, f). This biased expression probably contributes to the functional difference between mimicry alleles. Notably, expression of isoforms 1 and 2 seems to increase at day 5 after pupation (Fig. 3g), a stage at which immunodetection of Dsx spatial expression on mimetic forewings revealed a marked spatial correspondence with adult wing pattern (Fig. 3h).

Overall, our results indicate a surprising mode of action for *dsx* as a mimicry supergene. As a classic example of alternative splicing, our initial hypothesis was that alternative splicing would also underlie the phenotypic switch between female wing patterns. Although we do find clear evidence of alternative splicing, and different levels of isoform expression between female wing patterns, the set of female isoforms does not differ between groups. Rather, gene expression variation seems to have a central role in controlling mimicry polymorphism. Another striking feature of *dsx* in *P. polytes* is the large number of nucleotide substitutions that differ between mimicry alleles. The accumulation of neutral substitutions that is expected from balancing selection makes it difficult to infer which of these changes might be functionally related to mimicry polymorphism. However, we note that the proportion of fixed differences between *cyrus* and *polytes* haplotypes is over seven times greater in coding regions (72 out of 1,068 differences) compared to non-coding regions (972 out of 108,036 differences), and these coding region changes include 25 amino acid substitutions located primarily in the first exon (Table 1). The amino acid changes in exon 1 are clustered in two regions: the 5′ end of the protein, in front of the DNA binding (DM) domain, and the region between the DM domain and the dimerization domain; there are no amino acid changes in either domain (Extended Data Fig. 1). To explore the potential impact of these amino acid substitutions, we predicted secondary and tertiary structures for both the *cyrus* and *polytes* Dsx proteins and found that they differ markedly—the non-mimetic *cyrus* protein folds much like other insects, such as *Bombyx mori*, whereas the mimetic *polytes* protein structure is highly divergent (Extended Data Fig. 2). In addition to the differential expression of female isoforms, we speculate that distinct Dsx protein structures may also contribute to female polymorphism, with alternative alleles differentially regulating different downstream targets as a result of divergent DNA or coactivator binding properties.

How are a large number of nucleotide substitutions maintained in complete linkage disequilibrium over the approximately 100-kb length of *dsx*? Recombination between mimicry alleles in heterozygotes should break up *dsx* haplotypes, and the fact that we see many differences

Table 1 | DNA sequence variation in *Papilio polytes* near the mimicry supergene

Gene	Section	Length (bp)	Fixed synonymous/silent substitutions	Fixed non-synonymous substitutions	Total SNPs
neuro	ORF	519	0	0	8
clp	ORF	1,443	0	0	26
ferm	ORF	2,181	0	0	17
rad51	ORF	1,017	0	0	2
sir2	ORF	1,224	0	0	3
dsx	Exon 1	588	31	21	59
	Exon 2	144	5	0	8
	Exon 3	84	0	0	1
	Exon 4	69	2	1	3
	Exon 5	183	9	3	13
	Non-coding	108,036	972	NA	6,781
pros	ORF	2,895	0	0	21

Counts of synonymous/silent and non-synonymous nucleotide substitutions fixed between mimetic (*polytes*) and non-mimetic (*cyrus*) *P. polytes* butterflies in genes located near the mimicry supergene, as well as the total number of SNPs in each gene. Counts for *dsx* are separated by gene section (exons, non-coding) whereas counts for other genes represent predicted open reading frame (ORF).

between mimicry alleles suggests that something is reducing recombination immediately around *dsx*. Chromosomal inversions are well known to reduce recombination in heterozygotes[28], making this a likely explanation. We first verified that the *dsx* region does indeed exhibit elevated linkage disequilibrium relative to adjacent regions (Extended Data Fig. 3), and then we searched for evidence of structural variation around *dsx* using our genome re-sequencing data. As predicted, we found support for an inversion polymorphism associated with mimicry alleles, the breakpoints of which flank *dsx* (Extended Data Table 4 and Extended Data Fig. 4). Given the long history of speculation about the molecular identity of supergenes, it is interesting that we have uncovered a scenario that unites both possible explanations: reduced recombination among presumably different functional elements and single gene control. In essence, the multiple, tightly linked loci proposed by Clarke and Sheppard[13] may, in this case, actually be multiple, tightly linked mutations in the same gene.

It is perhaps unexpected that a gene so intimately connected to an essential developmental process could be co-opted to also control intraspecific polymorphism. Somehow, *dsx* has retained its highly conserved sex-differentiation properties[19–21] while also evolving new phenotype-switching properties in just one sex. Our results suggest two complementary mechanisms that may underlie the ability of *dsx* to have two distinct roles in *P. polytes*. First, although we found many mutations in the Dsx protein, none of these occurs in the DM or dimerization domains, which are essential components for its ancestral function in sexual differentiation. Second, we also found that different *dsx* isoforms are expressed on the wings and in the body of females, which may also allow this one gene to carry out a novel function on the wings.

R. A. Fisher called mimicry the "greatest post-Darwinian application of Natural Selection"[7] and supergene mimicry stands out as a particularly extreme adaptive endpoint. Although little is known about the molecular and developmental basis of supergene mimicry, previous evidence suggests that multiple, tightly linked genes probably underlie this phenomenon. Here we have integrated multiple approaches to reveal that a single gene acts as the mimicry supergene in *P. polytes*. In so doing, we have greatly expanded the known role of *doublesex* and the sexual differentiation pathway generally. Female-limited mimetic polymorphism has evolved independently multiple times in the genus *Papilio*[2], making this a useful system in which to investigate the generality of our results. One might predict that the sex determination pathway, and *dsx* in particular, may have been co-opted repeatedly to control this phenomenon because this pathway is preconfigured to mediate the most widespread polymorphism in the animal kingdom—sex. Interestingly, available data, although limited, suggest that this is not the case. For instance, female mimetic polymorphism in *Papilio dardanus* has previously been mapped to a genomic region containing the genes *engrailed* and *invected*[29], which is not linked to *dsx*. Furthermore, female mimetic

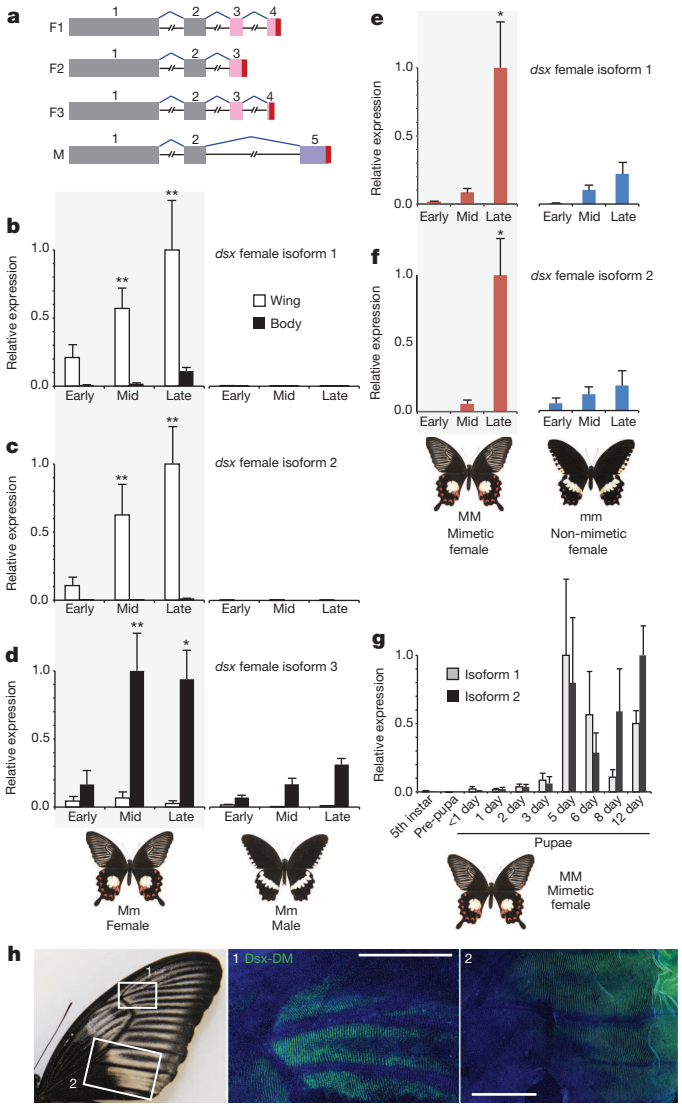


Figure 3 | Expression of *doublesex* in *P. polytes*. **a**, *dsx* is alternatively spliced into three female isoforms and one male isoform. **b**–**d**, Expression of female isoforms is strongly female-biased and isoform 1 (**b**) and isoform 2 (**c**) show wing-biased expression whereas isoform 3 expression (**d**) is body-biased; $n = 6$ (female early), 6 (female mid), 5 (female late), 6 (male early), 6 (male mid), 7 (male late). **e**, **f**, Female isoforms 1 and 2 also show elevated expression in mimetic females (*polytes*) relative to non-mimetic females (*cyrus*); $n = 9$ (*polytes* early), 9 (*polytes* mid), 12 (*polytes* late), 3 (*cyrus* early), 3 (*cyrus* mid), 3 (*cyrus* late). **g**, Finer scale temporal data for isoforms 1 and 2 on mimetic female wings suggests expression of both increases at 5 days after pupation; $n = 3$ for each time point. **h**, Immunodetection of Dsx protein on mimetic female wings 5 days after pupation reveals strong correlation with adult wing pattern. Scale bars, 1 mm. Data represented as mean ± s.e.m. All *n* values indicate number of biological replicates. $^*P < 0.05$; $^{**}P < 0.01$, ANOVA and Tukey's HSD test.

polymorphism in *Papilio glaucus* is sex-linked, with the primary switch locus on the W chromosome and a modifier on the Z chromosome[30]. Future work will determine whether other instances of sex-limited polymorphism, in butterflies and beyond, involve the sex differentiation pathway, but evolution, it seems, can take many paths to the extreme supergene genetic architecture, even among members of the same genus.

METHODS SUMMARY

Using one backcross mapping family (94 females: 48 *cyrus* and 46 *polytes*), we performed bulk segregant analysis with RAD markers. Subsequent fine mapping, using a total of nine mapping families (443 females: 229 *cyrus* and 214 *polytes*), and BAC sequencing isolated the mimicry locus to a 300-kb interval containing five genes, one of which was *dsx*. We sequenced the genomes of 30 laboratory-reared individuals (15 *polytes* and 15 *cyrus*) with an Illumina HiSeq 2000 and generated a reference genome sequence for *P. polytes* using both *de novo* and reference-guided assembly. Single nucleotide polymorphism (SNP) calling of the 30 sequenced genomes yielded 675,526 genome-wide SNPs and 94,776 SNPs across a 4-megabase (Mb) scaffold containing *dsx*. GWAS was performed by calculating genetic differentiation (F_{ST}) between *polytes* and *cyrus* individuals for *de novo* assembly scaffolds. Association tests across the 4-Mb *dsx* scaffold were performed using a false-discovery rate correction. We used Hudson–Kreitman–Aguadé (HKA) tests to compare nucleotide polymorphism among genes in the mimicry supergene region. Pairwise linkage disequilibrium was calculated among biallelic SNPs in two different portions of the *dsx* scaffold, and we used the short read sequence data to perform structural variant detection. We then used BLAST to identify scaffolds from a *de novo* assembly of *polytes* samples that appear to span an inversion containing *dsx*. Subsequent PCR tests isolated the 3′ breakpoint to a 2-kb interval. RNA-seq data, generated from wing-disc-derived *P. polytes* RNA, were used to perform transcriptome assembly and qRT–PCR was used to measure *dsx* isoform expression in males and females across development. We used a protein homology web server to infer secondary and tertiary structures of *polytes* and *cyrus* Dsx proteins, as well Dsx from *Bombyx mori*. Immunodetection of Dsx was carried out using a monoclonal anti-*Drosophila* Dsx DM domain antibody.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 November 2013; accepted 30 January 2014.
Published online 5 March 2014.

1. Joron, M. & Mallet, J. L. Diversity in mimicry: paradox or paradigm? *Trends Ecol. Evol.* **13,** 461–466 (1998).
2. Kunte, K. The diversity and evolution of batesian mimicry in *Papilio* swallowtail butterflies. *Evolution* **63,** 2707–2716 (2009).
3. Kunte, K. Female-limited mimetic polymorphism: a review of theories and a critique of sexual selection as balancing selection. *Anim. Behav.* **78,** 1029–1036 (2009).
4. Clarke, C. A. & Sheppard, P. M. Super-genes and mimicry. *Heredity* **14,** 175–185 (1960).
5. Charlesworth, D. & Charlesworth, B. Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *J. Theor. Biol.* **55,** 305–324 (1975).
6. Charlesworth, D. & Charlesworth, B. Mimicry: the hunting of the supergene. *Curr. Biol.* **21,** R846–R848 (2011).
7. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon Press, 1930).
8. Sheppard, P. M. The evolution of mimicry: a problem in ecology and genetics. *Cold Spring Harb. Symp. Quant. Biol.* **24,** 131–140 (1959).
9. Turner, J. R. G. in *The Biology of Butterflies* (eds Vane-Wright, R. I. & Ackery, P. R.) 141–161 (Academic, 1984).
10. Bates, H. W. Contributions to an insect fauna of the Amazon valley (Lepidoptera: Heliconidae). *Trans. Linn. Soc. (Lond.)* **23,** 495–566 (1862).
11. Ford, E. B. The genetics of polymorphism in the Lepidoptera. *Adv. Genet.* **5,** 43–87 (1953).
12. Mallet, J. & Joron, M. Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance, and speciation. *Annu. Rev. Ecol. Syst.* **30,** 201–233 (1999).
13. Clarke, C. A. & Sheppard, P. M. The genetics of the mimetic butterfly *Papilio polytes* L. *Phil. Trans. R. Soc. Lond. B* **263,** 431–458 (1972).
14. Clarke, C. A., Sheppard, P. M. & Thornton, I. W. B. The genetics of the mimetic butterfly *Papilio memnon* L. *Phil. Trans. R. Soc. Lond. B* **254,** 37–89 (1968).
15. Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477,** 203–206 (2011).
16. Larracuente, A. M. & Presgraves, D. C. The selfish *Segregation Distorter* gene complex of *Drosophila melanogaster*. *Genetics* **192,** 33–53 (2012).
17. Takayama, S. & Isogai, A. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56,** 467–489 (2005).
18. Nijhout, H. F. Developmental perspectives on evolution of butterfly mimicry. *Bioscience* **44,** 148–157 (1994).
19. Burtis, K. C. & Baker, B. S. *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56,** 997–1010 (1989).
20. Williams, T. M. & Carroll, S. B. Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature Rev. Genet.* **10,** 797–804 (2009).
21. Kopp, A. *Dmrt* genes in the development and evolution of sexual dimorphism. *Trends Genet.* **28,** 175–184 (2012).
22. Cho, S., Huang, Z. Y. & Zhang, J. Z. Sex-specific splicing of the honeybee *doublesex* gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. *Genetics* **177,** 1733–1741 (2007).
23. Kijimoto, T., Moczek, A. P. & Andrews, J. Diversification of *doublesex* function underlies morph-, sex-, and species-specific development of beetle horns. *Proc. Natl Acad. Sci. USA* **109,** 20526–20531 (2012).
24. Tanaka, K., Barmina, O., Sanders, L. E., Arbeitman, M. N. & Kopp, A. Evolution of sex-specific traits through changes in HOX-dependent *doublesex* expression. *PLoS Biol.* **9,** e1001131 (2011).
25. Williams, T. M. *et al.* The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134,** 610–623 (2008).
26. Loehlin, D. W. *et al.* Non-coding changes cause sex-specific wing size differences between closely related species of *Nasonia*. *PLoS Genet.* **6,** e1000821 (2010).
27. Charlesworth, B. & Charlesworth, D. *Elements of Evolutionary Genetics* (Roberts & Co., 2010).
28. Hoffmann, A. A., Sgro, C. M. & Weeks, A. R. Chromosomal inversion polymorphisms and adaptation. *Trends Ecol. Evol.* **19,** 482–488 (2004).
29. Clark, R. *et al.* Colour pattern specification in the Mocker swallowtail *Papilio dardanus*: the transcription factor *invected* is a candidate for the mimicry locus H. *Proc. R. Soc. Lond. B* **275,** 1181–1188 (2008).
30. Scriber, J. M., Hagen, R. H. & Lederhouse, R. C. Genetics of mimicry in the tiger swallowtail butterflies, *Papilio glaucus* and P. *canadensis* (Lepidoptera: Papilionidae). *Evolution* **50,** 222–236 (1996).

Acknowledgements We thank W. Wang for sharing genome sequence data, C. Robinett for providing the Dsx-DM monoclonal antibody, and E. Westerman, S. Nallu, M. Zhang, G. Garcia and N. Pierce for assistance and discussion. This project was funded by National Science Foundation grant DEB-1316037 to M.R.K.

Author Contributions K.K. conceived the project and helped design the study, reared mapping families and samples for gene expression analysis and genome sequencing, performed bulk-segregant analysis and RAD mapping, and contributed to drafting the manuscript. W.Z. generated the reference genome sequences and transcriptome assemblies, performed association mapping, GWAS analysis, HKA tests, structural variant detection and linkage disequilibrium analyses, analysis of protein structure and synonymous/non-synonymous calculations, and contributed to drafting the manuscript. A.T.-T. assisted with butterfly husbandry, performed fine mapping, cDNA sequencing and qRT–PCR analyses. D.H.P. performed qRT–PCR analyses. A.M. and R.D.R. performed Dsx immunohistochemistry. S.P.M. helped design the project and contributed to drafting the manuscript. M.R.K. designed and directed the project, analysed data and wrote the manuscript.

Author Information Sequence data are available from NCBI SRA (SRP035394) and GenBank (KJ150616–KJ150623). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.K. (mkronforst@uchicago.edu) or K.K. (krushnamegh@ncbs.res.in).

LETTER

doi:10.1038/nature13131

Dynamic sensory cues shape song structure in *Drosophila*

Philip Coen[1,2], Jan Clemens[1,2], Andrew J. Weinstein[1,2], Diego A. Pacheco[1,2], Yi Deng[3]† & Mala Murthy[1,2]

The generation of acoustic communication signals is widespread across the animal kingdom[1,2], and males of many species, including Drosophilidae, produce patterned courtship songs to increase their chance of success with a female. For some animals, song structure can vary considerably from one rendition to the next[3]; neural noise within pattern generating circuits is widely assumed to be the primary source of such variability, and statistical models that incorporate neural noise are successful at reproducing the full variation present in natural songs[4]. In direct contrast, here we demonstrate that much of the pattern variability in *Drosophila* courtship song can be explained by taking into account the dynamic sensory experience of the male. In particular, using a quantitative behavioural assay combined with computational modelling, we find that males use fast modulations in visual and self-motion signals to pattern their songs, a relationship that we show is evolutionarily conserved. Using neural circuit manipulations, we also identify the pathways involved in song patterning choices and show that females are sensitive to song features. Our data not only demonstrate that *Drosophila* song production is not a fixed action pattern[5,6], but establish *Drosophila* as a valuable new model for studies of rapid decision-making under both social and naturalistic conditions.

Drosophila melanogaster males chase females during courtship and produce song by wing vibration; females, meanwhile, arbitrate mating decisions. We developed a behavioural chamber to record acoustic signals and fly movements simultaneously (Fig. 1a and Supplementary Video 1); fly movements provide information on the sensory cues that may influence song production. We collected a large data set (>100,000 song bouts) to model the relationship between sensory cues and song patterning. Most experiments involve females that are pheromone-insensitive[7] and blind (termed PIBL) to facilitate auditory response measurements. All fly types used are described in Extended Data Table 1.

For one wild-type strain (WT1), we show that using arista-cut (deaf) females or wing-cut (mute) males increased the time to copulation and decreased the percentage of mated pairs (Fig. 1c). This corroborates prior work[8,9] demonstrating the importance of song for courtship success. Pairing WT1 males with wild-type, rather than PIBL, females did not alter these results (Fig. 1c), nor any of the results described below (not shown). All wild-type strains showed similar success with PIBL females (Extended Data Fig. 1b). Courtship songs comprise two modes (sine and pulse; Fig. 1b) and are part of a genetically hardwired mating ritual, thought to be stereotyped[6,10]. However, we find frequent mode transitions and variable mode durations individualize song bouts (Fig. 1d and Extended Data Fig. 2).

Males spend approximately 20% of courtship time singing (Extended Data Fig. 1c), and bouts can begin with either song mode. Using reverse correlation, we found that all tracking parameters correlate with bout initiations (Extended Data Fig. 3). We therefore turned to the generalized linear model (GLM) (Fig. 2a), widely used to analyse binary response data with several explanatory variables[11–13]. Unlike reverse correlation[14], the GLM we use includes a sparsity prior, which disentangles the contributions of correlated parameters to song patterning (see Methods)—this represents a major difference between our approach and previous studies[15].

Given similarities across fly strains (Extended Data Fig. 3), we combined data from all wild-type flies (315 pairs, 84,904 song bouts) for GLM analyses. We selected the most predictive features (≤ 600 ms of tracking parameter history) for each model based on deviance reduction (Extended Data Fig. 4a, b). For pulse song starts, combining two features: male forward velocity (mFV) and male lateral speed (mLS) strongly improved model fit (Fig. 2b). When tested on separate data, the fraction of correctly classified song starts (*P*Cor) was 0.67 (Fig. 2c), representing a 34% improvement over the null model (*P*Cor = 0.5). This compares favourably with fMRI-based predictions of human behaviour[16]

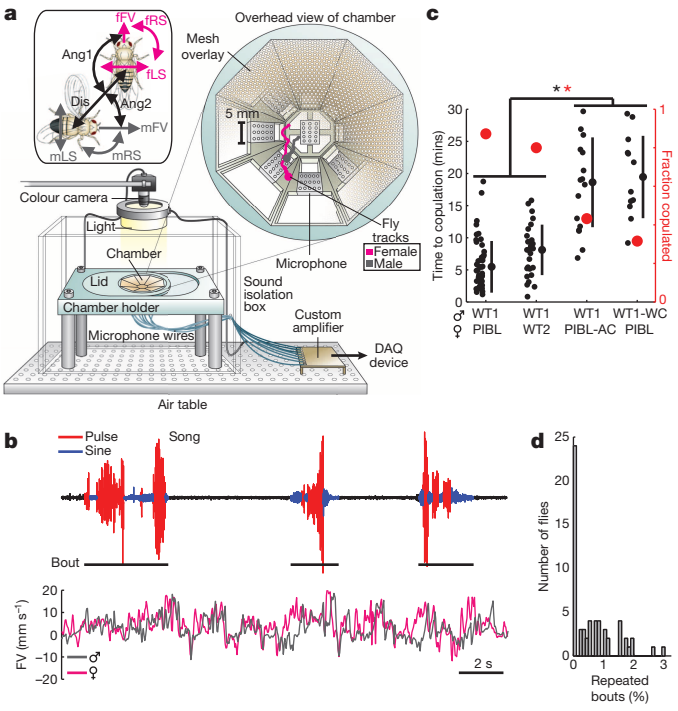


Figure 1 | A novel assay to study *Drosophila* song behaviour. a, Behavioural chamber with tracked fly movements (see Methods). Fly movements are divided into: male/female forward velocity (mFV/fFV), male/female lateral and rotational speeds (mLS/fLS and mRS/fRS), the distance between fly centres (Dis), the absolute angle from female/male heading to male/female centre (Ang1/Ang2). **b**, Segmentation of song bouts into pulse (red) and sine (blue) elements (top). Corresponding traces for mFV and fFV (bottom). **c**, Song is important for mating. Time to copulation increases (black, $*P < 0.001$) and fraction of copulated pairs decreases (red, $*P < 0.01$) when females are deaf or males are mute. Individual points, mean, and s.d. are given for each genotype ($n = 35$–48 pairs). AC, arista cut; WC, wing cut. **d**, Song is variable. The number of repeated bouts (containing pulse and sine) per fly (see Methods). $n = 60$ wild-type males.

[1]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA. [2]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [3]Department of Physics, Princeton University, Princeton, New Jersey 08544, USA. †Present address: Department of Biophysics, University of Washington School of Medicine, Seattle, Washington 98195, USA.

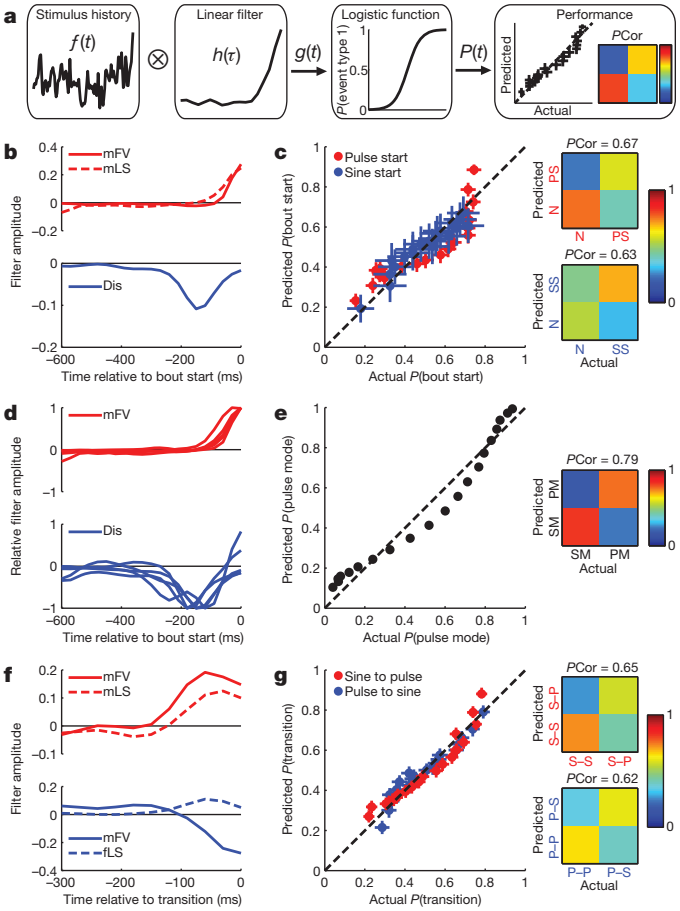


Figure 2 | Song bout patterning is predictable and based on few features. **a**, Schematic of the GLM (see Methods). Inputs—stimulus histories (features; $f(t)$) for each movement parameter—are used to predict binary event probabilities. Significant features are convolved with a linear filter $h(\tau)$, and the result, $g(t)$, is transformed into a probability $P(t)$, via a logistic function. Performance plots show the predicted and actual event probability relationships. Confusion matrices, from which we derive *P*Cor values, quantify model performance. **b**, Filters for pulse and sine song initiation GLMs. Unlike male lateral speed (mLS) or male forward velocity (mFV), the Dis filter indicates a time lag between distance estimation and sine song initiation. **c**, GLM performance for identifying pulse song starts (PS) using male forward velocity and male lateral speed filters ($n = 11{,}020$ test events from 315 males) and sine song starts (SS) using the Dis filter ($n = 2{,}476$ test events from 315 males). N = no song start. **d**, Male forward velocity pulse song start filters and Dis sine song start filters are similar for data from pheromone-insensitive or arista-cut males or males paired with arista-cut or sex-peptide-injected females; filters from wild-type males are also plotted. **e**, GLM performance for classifying current song mode (PM, pulse mode; SM, sine mode) using mean male forward velocity and male lateral speed ($n = 55{,}464$ test events from 315 males). **f**, Filters for sine to pulse (S–P) transitions (top) and the pulse to sine (P–S) transitions (bottom). **g**, GLM performance for identifying S–P transitions (versus continued sine song (S–S)) using male forward velocity and male lateral speed filters ($n = 17{,}118$ test events from 315 males) and P–S transitions (versus continued pulse song (P–P)) using male forward velocity and female lateral speed filters ($n = 11{,}748$ test events from 315 males). Error bars (most too small to visualize) indicate 95% confidence intervals (**c**, **e**, **g**).

and with two-alternative forced choice behavioural performance in *Drosophila*[17]. *P*Cor values are equivalent ($r^2 = 0.98$) to area under the curve values, an alternative performance measure (Extended Data Fig. 4c, d). We used a similar GLM framework to identify female forward velocity (fFV) as the best predictor of changes in male motion (male forward velocity (Extended Data Fig. 5a–c) and male lateral speed (data not shown)): that is, when the female speeds up, the male accelerates to follow her. Therefore, any correlation between male motion and song mode choice ultimately establishes a link between a sensory cue (for example, female motion) and song patterning. We address this point further below.

For songs that start in sine mode, the optimal model included only the distance between fly centres (Dis) (Fig. 2b, c and Extended Data Fig. 6a). Song start filters derived from PI (pheromone-insensitive) or arista-cut males paired with PIBL females, or from wild-type males paired with arista-cut or unreceptive/sex-peptide-injected[18] (SP) females, were indistinguishable from wild-type filters (Fig. 2d), even though males take longer to copulate with arista-cut females (Fig. 1c), and never copulate with unreceptive/sex-peptide-injected females (Extended Data Fig. 7a). A model designed to distinguish song bouts beginning in sine versus pulse mode retains male forward velocity and the distance between fly centres as the most predictive features, but with significantly increased performance (Extended Data Fig. 6b, c). Therefore, we focus hereafter on song patterning decisions, rather than the male's decision to sing versus perform another courtship behaviour. Here a decision refers to a behavioural choice biased by sensory information[19].

During song bouts, males typically alternate between sine and pulse modes, with each mode lasting tens to hundreds of milliseconds. We next investigated whether GLMs could also predict the current mode of song within bouts. Model performance was optimal using only two features: male forward velocity and male lateral speed (a 58% improvement over the null model; Fig. 2e). The absence of the distance between fly centres (Dis) feature in this model is probably due to its reduced variance during song (Extended Data Fig. 6d, e). Using different, male-centric, features only decreased model performance (Extended Data Fig. 8). We then went on to predict all mode transitions within a bout: increases in male forward velocity and male lateral speed predict transitions to pulse mode, whereas decreases in male forward velocity and increases in female lateral speed predict transitions to sine mode (Fig. 2f). Mode transitions represent a subtle change in behaviour (for example, whether 300 ms of pulse song is followed by 30 ms of sine song or 30 ms of continued pulse song); nonetheless, our model predictions produced a combined *P*Cor of 0.64 (Fig. 2g). Thus, taking into account male motion and inter-fly distance can largely explain variability in song patterning. Although studies in birds have shown that auditory cues, either produced by the singer itself[20] or by a duetting partner[21], affect acoustic sequence generation, to our knowledge, ours is the first demonstration of a role for non-auditory sensory inputs in dynamically patterning the structure of individual song sequences.

Next, we considered which sensory pathways mediated the male's decision-making during song production. Although male motion is the primary contributor to song patterning in our models, we observed a strong correlation ($r^2 = 0.95$) during song bouts between inter-fly distance (beyond the tactile range of ~5 mm; the tail of the distribution in Extended Data Fig. 6d) and the pulse/sine ratio (Fig. 3a; correlation is independent of male movement). We conclude that flies use vision to measure distance over this range, because blind males or wild-type males placed in the dark, sing significantly more pulse song (Fig. 3b, c); this is not true for any other sensory deficit and cannot be explained by changes in male speed (Extended Data Fig. 5e).

Previous studies have demonstrated that separate neurons control pulse and sine song production[22,23]. This indicates that song patterning is neurally controlled, and does not arise simply from mechanical coupling with male locomotion changes. In support of this, males sing both song modes at all velocities (Fig. 3d). We further conclude that visual measurements of optic flow are not used to convey male motion signals to song patterning networks, because a model based on only male forward velocity and male lateral speed predicts current song mode for blind males (Fig. 3e). This left two likely possibilities (Fig. 3f): either (mechanism 1) a cue from the female induces males to change speed and concomitantly affects song patterning or (mechanism 2) neural circuits that carry information about male motion (via either a copy of the motor commands or proprioceptive feedback from the legs) modulate song patterning circuits. Because female forward velocity predicts

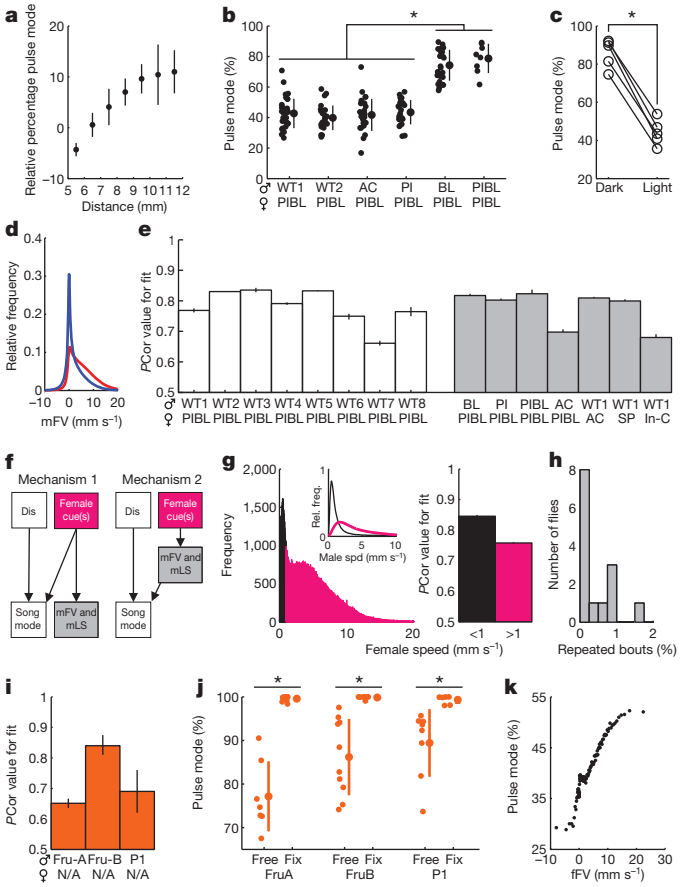


Figure 3 | Neural pathways that modulate song patterning. **a**, Percentage of song in pulse mode (mean and s.d.) increases with inter-fly distance ($r^2 = 0.95$; see Methods). **b**, Pulse song percentage increases in blind males ($n = 11$–48 flies, $^*P < 0.001$). **c**, Individual WT2 males (paired with PIBL females) produce more pulse song in dark versus light ($n = 5$ flies, $^*P < 0.0001$). **d**, Normalized event frequency for pulse or sine song (red or blue, $n = 57{,}0225$ or $n = 95{,}2541$, from 315 males) at each male forward velocity, across all wild-type strains. **e**, *P*Cor values from classifying current song mode (using mean male forward velocity and male lateral speed) for wild-type strains (white bars) and various sensory manipulations ($n = 924$–16,256 test events from 11–48 flies for each model). **f**, Two potential neural circuit mechanisms underlying the correlation between male motion and song. Female cue(s) directly modulate both song patterning and locomotor circuits (left) or circuits carrying information about male motion (right) modulate song patterning circuits. **g**, Wild-type data were split into songs produced when females were not moving or moving (left, black or magenta, $n = 9{,}454$ or $n = 46{,}204$, test events from 315 males; see Methods). Inset shows corresponding male speeds. Corresponding *P*Cor values from classifying current song mode using mean male forward velocity and male lateral speed (right). **h**, Song variability with TrpA1-activated flies ($n = 14$ males from 3 genotypes) is similar to wild type (Fig. 1d). **i**, *P*Cor values from classifying current song mode using mean mFV and mLS for Fru-A, Fru-B, and P1-activated males ($n = 1{,}987$ and 200 and 100 test events from 7 and 10 and 8 males, respectively). **j**, For each genotype, the percentage of pulse song increases ($^*P < 0.01$) when flies are fixed. **k**, Correlation between female forward velocity and the percentage of pulse song ($n = 16{,}092$ from 315 wild-type males paired with PIBL females, binned by percentile) at a 60 ms lag ($r^2 = 0.91$). Error bars indicate 95% confidence intervals (**e, g, i**). **b** and **j** show individual data points, mean and s.d. for each group.

male motion (Extended Data Fig. 5a–c), song patterning would ultimately be dependent on sensory cues for both mechanisms.

To distinguish between these mechanisms, we removed the link between female cues and male motion. Male lateral speed and male forward velocity still predicted current song mode when males are pheromone-insensitive and/or blind (Fig. 3e). Given that blind males do not follow females (Extended Data Fig. 5d), and produce song over a wide range of inter-fly distances/orientations (Extended Data Fig. 8d–g), it seems unlikely that a non-visual, long-range cue from the female patterns song in these males. Further, model performance remains high when females provide no motion cues (Fig. 3g and Extended Data Fig. 9). Finally, we examined the link between male motion and song patterning without a female present. We artificially activated song and targeted three neural subsets (see Methods). For all genotypes, levels of inter-bout variability were similar to wild-type levels despite constitutive thermal activation (Fig. 3h). Again, male lateral speed and male forward velocity predicted current song mode (Fig. 3i) within the performance range for wild-type strains (Fig. 3e). By preventing these males from moving, we observed a marked reduction in sine song (Fig. 3j); because this song mode is typically produced at lower male speeds, our results indicate that a copy of the locomotor commands (presumably still active in 'fixed' males) is more likely (than proprioceptive feedback) to pattern song. Thus, our data support the conclusion that activity in locomotor circuits influences song patterning, favouring mechanism 2 (Fig. 3f). Consistent with this, we observed the strongest correlation ($r^2 = 0.91$) between female forward velocity and the pulse/sine ratio in wild-type flies at the delay at which males follow females (Fig. 3k and Extended Data Fig. 5b).

As a final test of the importance of sensory cues in song patterning, we considered song bout terminations. The exponential distribution of syllable durations in songbirds has been proposed to support a stochastic mechanism for syllable termination[24]. Bout durations in *Drosophila* are also well fit by an exponential function ($r^2 = 0.98$; Extended Data Fig. 2). However, we identified female lateral speed as a significant predictor of bout ends (Fig. 4a, b). We posit that when males sense changes in female lateral speed they either transition to sine song (Fig. 2f) or they end song altogether.

On the basis of the data presented, we propose that, for *Drosophila*, detection of a female[25] gates the song production pathway. Once gated, sensory cues act directly (inter-fly distance and female lateral speed) or indirectly (via male forward velocity and male lateral speed) to pattern song on short timescales (Fig. 4c). Although trial-to-trial variability in acoustic signals may be useful for song learning in birds[26], *Drosophila* males do not learn their songs. Therefore, we considered alternative roles for patterning decisions in fly mating behaviours. Because female speed decreases before copulation (Fig. 4d), we reasoned that female slowing was a sign of receptivity. Indeed, using the GLM framework, we found a negative correlation between the amount of either song mode, in a given time window, and female speed (Fig. 4e). This relationship was reduced for deaf females, whereas unreceptive (SP) females showed a positive correlation between speed and song amount, a reversal of wild-type behaviour. In addition, females increased or decreased speed in response to song from *Drosophila simulans* (heterospecific) or WT1 (conspecific) males, respectively. This is particularly striking when considering that, for the same experiment, male motion predicts song mode choice for both *D. simulans* and *D. melanogaster* males (Fig. 4f). Therefore, male song patterning choices are unlikely to be used by females for mate selection. Indeed, successful males (those that copulated within 30 min) and unsuccessful males pattern song similarly (Extended Data Fig. 7b–d). We speculate that males bias towards their louder form of song (pulse) when far from the female (captured by the distance between fly centres feature) or when trying to catch up to, or locate, the female (captured by the male forward velocity and male lateral speed features); this would maximize the probability of the female hearing his song and decreasing her locomotion. It remains to be determined, however, over which specific distances and angles females can detect each song mode (and amplitude modulations therein).

In conclusion, instinctive behaviours, like acoustic signal production, have been generally considered to comprise a series of fixed action patterns, elicited and oriented by sensory information[27]. Courtship song production in *Drosophila* has long been regarded as an example of such a fixed action sequence[5,6], with the female serving as the trigger stimulus. In contrast to this view, we show that even the simple fly uses sensory

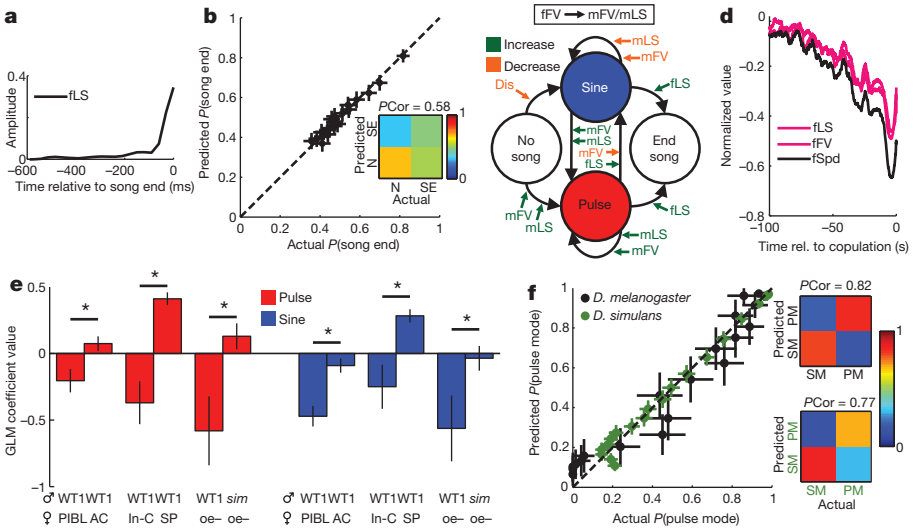


Figure 4 | Song patterning decisions and female responses. **a**, Filter from GLM for song ends. **b**, GLM performance for identifying song ends (SE) using female lateral speed ($n = 10{,}708$ events from 315 males). N, no song end. **c**, Summary of the influence of sensory inputs on song patterning, as revealed by GLM analysis. **d**, Normalized changes in female motion before copulation ($n = 233$ flies). **e**, GLM coefficient values between pulse or sine song density and female speed (*$P < 0.01$, $n = 40$ to $n = 1{,}429$ samples from 7–38 pairs; see Methods). **f**, GLM performance for classifying current song mode using mean male forward velocity and male lateral speed for WT2 or *sim* (black or green, $n = 1424$ or 9854 test events from 31 or 40 flies) males paired with oe- (oenocyte-less and PIBL) females. Error bars are 95% confidence intervals (**b**, **e** and **f**).

information to pattern his song sequences over short timescales. These data therefore offer a new window into the study of instinctive behaviours, and indicate that song production in flies may be more analogous to complex motor behaviours, such as prey capture, known to rely heavily on sensory feedback for patterning. More broadly, and consistent with recent studies of fly flight and human mobility[28,29], we suspect that seemingly stochastic behaviours may be more predictable than expected.

METHODS SUMMARY

Behavioural data (song and video recordings) were acquired by pairing virgin male and female flies in a custom-built chamber, designed to capture fly acoustic signals throughout the environment (~25 mm diameter, tiled with 9 microphones and connected to a specialized amplifier[30]) and to be compatible with our fly tracking software. Male wild-type strains came from diverse geographic locations; most females tested were genetically engineered to be both blind (*GMR-hid*) and pheromone-insensitive (*orco*). Other genetic and physical manipulations included arista cutting (deaf flies), wing cutting (mute males), sex-peptide-injection and oenocyte removal. Neural activation was achieved by expressing TrpA1 in three different subsets of Fru^{+} neurons and heating the entire chamber before introduction of male flies. All data processing and analysis was conducted in MATLAB. Song was segmented as previously described[30]. A modified generalized linear model[11], which uses a sparseness prior in order to penalize redundant features, was implemented to determine whether fly movements and positions could predict male song patterning choices, including bout initiation, song mode (pulse versus sine) within a bout, mode transitions and bout termination. When fitting or testing models over 1,000 iterations, data were randomly subsampled to equalize the frequency of each event type (a common method for dealing with uneven event frequencies). Predictive features for each model were selected using deviance reduction and model performance was tested using independent data sets. *P*Cor values were used to quantify model performance. To measure female responses to song, female speed and amount of male song were compared using a 1 min sliding window, with 50% overlap.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 June 2013; accepted 5 February 2014.
Published online 5 March 2014.

1. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22,** 567–631 (1999).
2. Bentley, D. & Hoy, R. R. The neurobiology of cricket song. *Sci. Am.* **231,** 34–50 (1974).
3. Sakata, J. T. & Brainard, M. S. Online contributions of auditory feedback to neural activity in avian song control circuitry. *J. Neurosci.* **28,** 11378–11390 (2008).
4. Jin, D. Z. & Kozhevnikov, A. A. A compact statistical model of the song syntax in Bengalese finch. *PLoS Comput. Biol.* **7,** e1001108 (2011).
5. Hall, J. C. The mating of a fly. *Science* **264,** 1702–1714 (1994).
6. Demir, E. & Dickson, B. J. *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell* **121,** 785–794 (2005).
7. Ziegler, A. B., Berthelot-Grosjean, M. & Grosjean, Y. The smell of love in *Drosophila*. *Front. Physiol.* **4,** 72 (2013).
8. von Schilcher, F. The role of auditory stimuli in the courtship of *Drosophila melanogaster*. *Anim. Behav.* **24,** 18–26 (1976).
9. Bennet-Clark, H. C. & Ewing, A. W. Pulse interval as a critical parameter in the courtship song of *Drosophila melanogaster*. *Anim. Behav.* **17,** 755–759 (1969).
10. Baker, B. S., Taylor, B. J. & Hall, J. C. Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell* **105,** 13–24 (2001).
11. Mineault, P. J., Barthelme, S. & Pack, C. C. Improved classification images with sparse priors in a smooth basis. *J. Vis.* **9,** http://dx.doi.org/10.1167/9.10.17 (2009).
12. Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L. & Woolley, S. M. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE* **6,** e16104 (2011).
13. Pillow, J. W. *et al.* Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454,** 995–999 (2008).
14. Sharpee, T., Rust, N. C. & Bialek, W. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput.* **16,** 223–250 (2004).
15. Trott, A. R., Donelson, N. C., Griffith, L. C. & Ejima, A. Song choice is modulated by female movement in *Drosophila* males. *PLoS ONE* **7,** e46025 (2012).
16. Soon, C. S., Brass, M., Heinze, H. J. & Haynes, J. D. Unconscious determinants of free decisions in the human brain. *Nature Neurosci.* **11,** 543–545 (2008).
17. Perisse, E. *et al.* Different Kenyon cell populations drive learned approach and avoidance in *Drosophila*. *Neuron* **79,** 945–956 (2013).
18. Yapici, N., Kim, Y. J., Ribeiro, C. & Dickson, B. J. A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature* **451,** 33–37 (2008).
19. McFarland, D. J. Decision making in animals. *Nature* **269,** 15–21 (1977).
20. Sakata, J. T. & Brainard, M. S. Real-time contributions of auditory feedback to avian vocal motor control. *J. Neurosci.* **26,** 9619–9628 (2006).
21. Fortune, E. S., Rodriguez, C., Li, D., Ball, G. F. & Coleman, M. J. Neural mechanisms for the coordination of duet singing in wrens. *Science* **334,** 666–670 (2011).
22. Ewing, A. W. The neuromuscular basis of courtship song in *Drosophila*: the role of the direct and axillary wing muscles. *J. Comp. Physiol.* **130,** 87–93 (1979).
23. Shirangi, T. R., Stern, D. L. & Truman, J. W. Motor control of *Drosophila* courtship song. *Cell Rep.* **5,** 678–686 (2013).
24. Aronov, D., Veit, L., Goldberg, J. H. & Fee, M. S. Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. *J. Neurosci.* **31,** 16353–16368 (2011).
25. Dickson, B. J. Wired for sex: the neurobiology of *Drosophila* mating decisions. *Science* **322,** 904–909 (2008).
26. Olveczky, B. P., Andalman, A. S. & Fee, M. S. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* **3,** e153 (2005).
27. Tinbergen, N. *The Study of Instinct* (Clarendon Press, 1951).

28. Censi, A., Straw, A. D., Sayaman, R. W., Murray, R. M. & Dickinson, M. H. Discriminating external and internal causes for heading changes in freely flying *Drosophila*. *PLoS Comput. Biol.* **9,** e1002891 (2013).
29. Song, C., Qu, Z., Blumm, N. & Barabasi, A. L. Limits of predictability in human mobility. *Science* **327,** 1018–1021 (2010).
30. Arthur, B. J., Sunayama-Morita, T., Coen, P., Murthy, M. & Stern, D. L. Multi-channel acoustic recording and automated analysis of *Drosophila* courtship songs. *BMC Biol.* **11,** 11 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Arthur and D. Stern for assistance in establishing the song recording system; P. Andolfatto for wild-type fly strains; S. Kamal and V. Cheng for assistance with selecting and maintaining fly strains; G. Guan for technical assistance; T. Tayler for help with injections; J. Shaevitz for help with the fly tracker; R. da Silveira for early discussions on reverse correlation; and G. Laurent, C. Brody, D. Aronov, I. Fiete, M. Ryan, and the entire Murthy lab for thoughtful feedback and comments on the manuscript. Figure 1a was illustrated by K. Ris-Vicari. P.C. is funded by an HHMI International Predoctoral Fellowship and M.M. is funded by the Alfred P. Sloan Foundation, the Human Frontiers Science Program, an NSF CAREER award, the McKnight Endowment Fund, and the Klingenstein Foundation.

Author Contributions P.C. and M.M. designed the study. P.C., A.J.W. and D.A.P. collected and processed the data. Y.D. developed the fly tracking algorithm. P.C. and J.C. analysed the data. P.C. and M.M. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (mmurthy@princeton.edu).

LETTER

doi:10.1038/nature12956

An excitatory paraventricular nucleus to AgRP neuron circuit that drives hunger

Michael J. Krashes[1]†*, Bhavik P. Shah[1]†*, Joseph C. Madara[1], David P. Olson[1]†, David E. Strochlic[2,3], Alastair S. Garfield[1,4], Linh Vong[1]†, Hongjuan Pei[5], Mitsuko Watabe-Uchida[6], Naoshige Uchida[3,6], Stephen D. Liberles[2,3] & Bradford B. Lowell[1,3]

Hunger is a hard-wired motivational state essential for survival. Agouti-related peptide (AgRP)-expressing neurons in the arcuate nucleus (ARC) at the base of the hypothalamus are crucial to the control of hunger. They are activated by caloric deficiency and, when naturally or artificially stimulated, they potently induce intense hunger and subsequent food intake[1–5]. Consistent with their obligatory role in regulating appetite, genetic ablation or chemogenetic inhibition of AgRP neurons decreases feeding[3,6,7]. Excitatory input to AgRP neurons is important in caloric-deficiency-induced activation, and is notable for its remarkable degree of caloric-state-dependent synaptic plasticity[8–10]. Despite the important role of excitatory input, its source(s) has been unknown. Here, through the use of Cre-recombinase-enabled, cell-specific neuron mapping techniques in mice, we have discovered strong excitatory drive that, unexpectedly, emanates from the hypothalamic paraventricular nucleus, specifically from subsets of neurons expressing thyrotropin-releasing hormone (TRH) and pituitary adenylate cyclase-activating polypeptide (PACAP, also known as ADCYAP1). Chemogenetic stimulation of these afferent neurons in sated mice markedly activates AgRP neurons and induces intense feeding. Conversely, acute inhibition in mice with caloric-deficiency-induced hunger decreases feeding. Discovery of these afferent neurons capable of triggering hunger advances understanding of how this intense motivational state is regulated.

To identify monosynaptic inputs to AgRP neurons, we used a modified rabies virus SADΔG–EGFP (EnvA)[11] in combination with Cre-dependent helper adeno-associated viruses (AAVs) expressing TVA (receptor for the avian sarcoma leucosis virus glycoprotein EnvA; AAV8-FLEX-TVA–mCherry) and RG (rabies envelope glycoprotein; AAV8-FLEX–RG). When used with *Agrp-IRES-Cre* mice, TVA and RG, respectively, allow for rabies infection of AgRP neurons and subsequent retrograde transynaptic spread[11,12] (Fig. 1a). AAV targeting of the helper viruses was specific to AgRP neurons (Supplementary Fig. 1). Three weeks post-AAV transduction, we injected SADΔG–EGFP (EnvA) into the same area and examined brains 7 days later for $EGFP^+$ signal. We detected the highest number of $EGFP^+$ cells in the ARC (38%), probably representing the initially infected AgRP neurons, and possibly local afferents (Fig. 1b; Supplementary Fig. 2). We next evaluated distant upstream anatomical areas for $EGFP^+$ neurons and found that the vast majority were located in two hypothalamic nuclei, the dorsal medial hypothalamus (DMH, 26%) which contains both glutamatergic and GABAergic neurons[13] and the paraventricular hypothalamus (PVH, 18%) consisting primarily of glutamatergic neurons[13] (Fig. 1b; Supplementary Fig. 2). Finally, we also observed a smaller number of $EGFP^+$ cells in other hypothalamic sites (Supplementary Fig. 2).

We next used channelrhodopsin (ChR2)-assisted circuit mapping (CRACM)[14,15] to both confirm and determine valence of functional monosynaptic connectivity between afferents and AgRP neurons. CRACM involves *in vivo* targeted expression of ChR2, a photoexcitable cation channel, in presumptive presynaptic upstream neurons (and their terminals), followed by *ex vivo* electrophysiologic assessment in acute brain slices of light-evoked postsynaptic currents in candidate downstream neurons. To investigate excitatory input to AgRP neurons, we stereotaxically injected Cre-dependent AAV expressing ChR2–mCherry (AAV8-DIO-ChR2–mCherry) (Supplementary Fig. 3a) into brain sites of *Vglut2-IRES-Cre; Npy-hrGFP* mice[13]. VGLUT2 (also known as SLC17A6) is the glutamate synaptic vesicle transporter expressed in the hypothalamus, hence *Vglut2-IRES-Cre* mice target relevant excitatory neurons[13]. As AgRP neurons co-express neuropeptide Y (NPY), *Npy-hrGFP* mice allow visualization of AgRP neurons[16,17]. Consistent with the rabies tracing, we detected light-evoked excitatory post-synaptic currents (EPSCs) in all $VGLUT2^{DMH} \rightarrow AgRP^{ARC}$ neurons tested (latency between onset of light and EPSC = 4.7 ± 0.2 ms; Fig. 1c; Supplementary Fig. 3f). These were blocked by CNQX (6-cyano-7-nitroquinoxaline-2,3-dione), an AMPA receptor antagonist, confirming their glutamatergic nature. Next, we examined monosynaptic connections between $VGLUT2^{PVH} \rightarrow AgRP^{ARC}$ neurons and again, consistent with the rabies mapping, we observed light-evoked EPSCs in all AgRP neurons tested (latency = 4.9 ± 0.4 ms; Fig. 1d; Supplementary Fig. 3g). These also were blocked by CNQX.

In addition, we selectively expressed ChR2 in the ventral medial hypothalamus (VMH) and lateral hypothalamus (LH), two sites with few $EGFP^+$ cells, and also the ARC, which could provide local afferents, and investigated possible connectivity to AgRP neurons. In agreement with the negative rabies data, no light-evoked EPSCs were detected in 36 out of 37 $VGLUT2^{VMH} \rightarrow AgRP^{ARC}$ neurons tested (Supplementary Fig. 3b, h) or in any $VGLUT2^{LH} \rightarrow AgRP^{ARC}$ neurons tested (Supplementary Fig. 3c, i). Likewise, we failed to detect light-evoked EPSCs in any $VGLUT2^{ARC} \rightarrow AgRP^{ARC}$ neurons tested (Supplementary Fig. 3d, j). However, and as previously noted[18], glutamatergic VMH neurons were monosynaptically connected to nearby pro-opiomelanocortin (POMC) neurons ($VGLUT2^{VMH} \rightarrow POMC^{ARC}$), as we observed light-evoked EPSCs in all POMC neurons tested (latency = 4.4 ± 0.2 ms; Supplementary Fig. 3e).

The CRACM studies suggest marked differences in the strength of $VGLUT2^{PVH} \rightarrow AgRP^{ARC}$ versus $VGLUT2^{DMH} \rightarrow AgRP^{ARC}$ inputs. First, the amplitude of light-evoked EPSCs generated from $VGLUT2^{PVH}$ inputs were approximately threefold greater (Fig. 1e). Second, the effectiveness of light pulses in evoking EPSCs differed, with DMH inputs showing a much higher failure rate (~32%; Fig. 1f; Supplementary Fig. 4) compared

[1]Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. [2]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Program in Neuroscience, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Center for Integrative Physiology, University of Edinburgh, Edinburgh EH8 9XD, UK. [5]Division of Pediatric Endocrinology, Departments of Pediatrics, University of Michigan, Ann Arbor, Michigan 48105, USA. [6]Center for Brain Science, Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA. †Present addresses: Diabetes, Endocrinology and Obesity Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA (M.J.K.); National Institute on Drug Abuse, National Institutes of Health, Baltimore, Maryland 21224, USA (M.J.K.); Cardiovascular and Metabolic Diseases, Pfizer, 610 Main Street, Cambridge, Massachusetts 02139, USA (B.P.S.); Division of Pediatric Endocrinology, Departments of Pediatrics, University of Michigan, Ann Arbor, Michigan 48105, USA (D.P.O.); Cardiovascular and Metabolic Diseases, Novartis Institutes for BioMedical Research, 100 Technology Square, Cambridge, Massachusetts 02139, USA (L.V.).

*These authors contributed equally to this work.

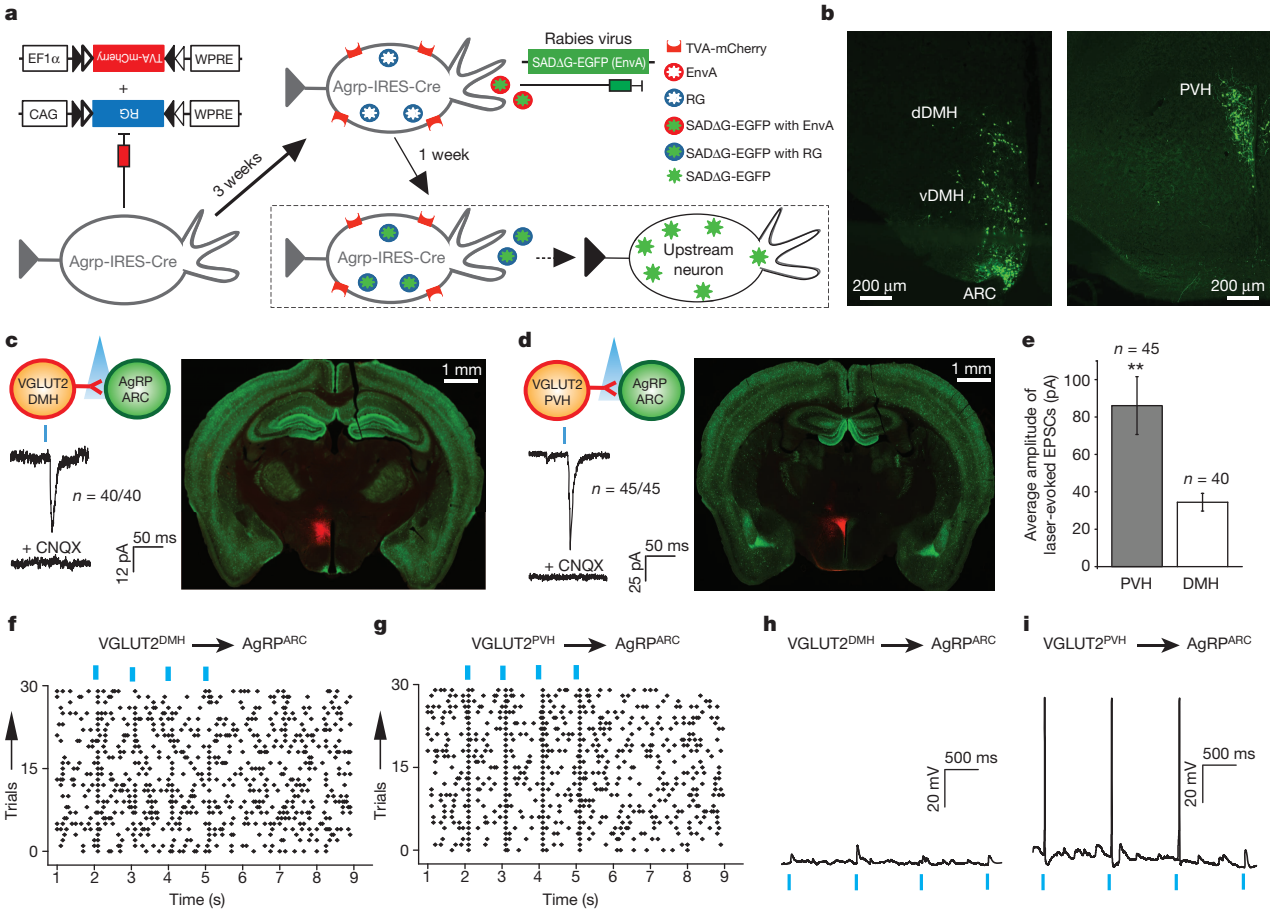


Figure 1 | Mapping and evaluating connectivity of inputs to AgRP^{ARC} neurons. **a**, Rabies schematic. **b**, EGFP detected in the ARC and DMH (left) and PVH (right) in *Agrp-IRES-Cre* mice. dDMH, dorsal DMH; vDMH, ventral DMH. **c**, **d**, Top, schematic shows connections being tested. Right, representative brains from mice stereotaxically injected with AAV8-DIO-ChR2–mCherry (red = ChR2–mCherry, green = hrGFP from *Npy-hrGFP*). Bottom, representative traces showing assessment of light-evoked excitatory postsynaptic currents (EPSCs) with a blue tic indicating the light pulse (473 nm wavelength, 2 msec). Mice used were *Vglut2-IRES-Cre; Npy-hrGFP*. ChR2 was targeted to DMH (**c**) and PVH (**d**). CNQX is an AMPA receptor blocker. **e**, Average amplitude of light-evoked EPSCs (pA) in AgRP neurons ($n = 40$ for VGLUT2^{DMH}→AgRP^{ARC} group; $n = 45$ for VGLUT2^{PVH}→AgRP^{ARC} group). Results are shown as mean ± s.e.m. *P* values for unpaired comparisons were calculated by two-tailed Student's *t*-test. $^{**}P < 0.001$. **f**, **g**, Representative raster plots of EPSCs for VGLUT2^{DMH}→AgRP^{ARC} (**f**) and VGLUT2^{PVH}→AgRP^{ARC} (**g**). **h**, **i**, Representative traces showing light-evoked changes in membrane potentials in AgRP neurons for VGLUT2^{DMH}→AgRP^{ARC} (**h**) and VGLUT2^{PVH}→AgRP^{ARC} (**i**).

with PVH inputs in which time-locked EPSCs followed every light pulse (0% failure rate; Fig. 1g; Supplementary Fig. 4). To test for consequences of these synaptic strength differences, we assessed the ability of VGLUT2^{DMH}→AgRP^{ARC} or VGLUT2^{PVH}→AgRP^{ARC} inputs to fire action potentials in AgRP neurons. Whereas we never detected VGLUT2^{DMH}→AgRP^{ARC} light-evoked action potentials (Fig. 1h), we observed VGLUT2^{PVH}→AgRP^{ARC} light-evoked action potentials approximately 75% of the time (Fig. 1i), demonstrating potent, excitatory input to AgRP neurons originating from the PVH.

To determine which PVH neurons monosynaptically drive AgRP neurons, we used many lines of Cre-expressing mice, each targeting subsets of PVH neurons, and then used CRACM to determine connectivity. *Sim1* encodes for the single-minded homologue 1 transcription factor required for developmental specification of the PVH, and is thus expressed in most PVH neurons[19], as verified in the *Sim1-Cre* transgenic mouse[20]. We observed light-evoked EPSCs in all SIM1^{PVH}→AgRP^{ARC} neurons tested (latency = 4.6 ± 0.4 ms; Fig. 2a). We next surveyed the Allen Brain Atlas for mRNAs highly enriched in the PVH, and then assembled a series of IRES-Cre knock-in mice targeting cells marked by prodynorphin (*Pdyn-IRES-Cre*; Supplementary Fig. 5), oxytocin (*Oxt-IRES-Cre*)[21], arginine vasopressin (*Avp-IRES-Cre*; Supplementary Fig. 6), corticotropin-releasing hormone (*Crh-IRES-Cre*; Supplementary Fig. 7), TRH (*Trh-IRES-Cre*; Supplementary Fig. 8) and PACAP (*Pacap-IRES-Cre*; Supplementary Fig. 9). Following ChR2 delivery into the PVH, we failed to detect light-evoked EPSCs in all PDYN^{PVH}→AgRP^{ARC} (Fig. 2b), OXT^{PVH}→AgRP^{ARC} (Supplementary Fig. 10a), AVP^{PVH}→AgRP^{ARC} (Supplementary Fig. 10b) or CRH^{PVH}→AgRP^{ARC} (Supplementary Fig. 10c) neurons tested. However, we detected light-evoked EPSCs in all TRH^{PVH}→AgRP^{ARC} neurons (latency = 4.7 ± 0.4 ms; Fig. 2c) and all PACAP^{PVH}→AgRP^{ARC} neurons tested (latency = 4.8 ± 0.3 ms; Fig. 2d). These findings demonstrate that TRH^{PVH} and PACAP^{PVH} neurons provide excitatory drive to AgRP neurons. It is not known if these excitatory inputs come from two distinct classes of PVH neurons or from one which co-expresses TRH and PACAP. The latter is possible given that 37% of TRH mRNA-expressing neurons are marked by *Pacap-IRES-Cre* (Supplementary Fig. 11).

Innervation by PACAP^{PVH} neurons suggests that PACAP, in addition to glutamate, could also activate AgRP neurons. To assess this, we exogenously added PACAP$_{1-38}$ (100 nM)[22] and determined effects on activity of synaptically isolated AgRP neurons (in the presence of picrotoxin (PTX) and kynurenate). Of note, PACAP$_{1-38}$ markedly depolarized and increased firing rate, and this was prevented by co-addition of the PAC1-receptor blocker, PACAP$_{6-38}$ (200 nM) (Fig. 2e–h). Thus, in addition to glutamate, PACAP working through PAC1 receptors also probably plays an important role in the PACAP^{PVH}→AgRP^{ARC} circuit.

POMC and AgRP neurons lie in close proximity within the ARC, but have opposing effects on feeding[2,23]. If fidelity of the identified PVH→AgRP^{ARC} neuron hunger circuit is to be high, then it should

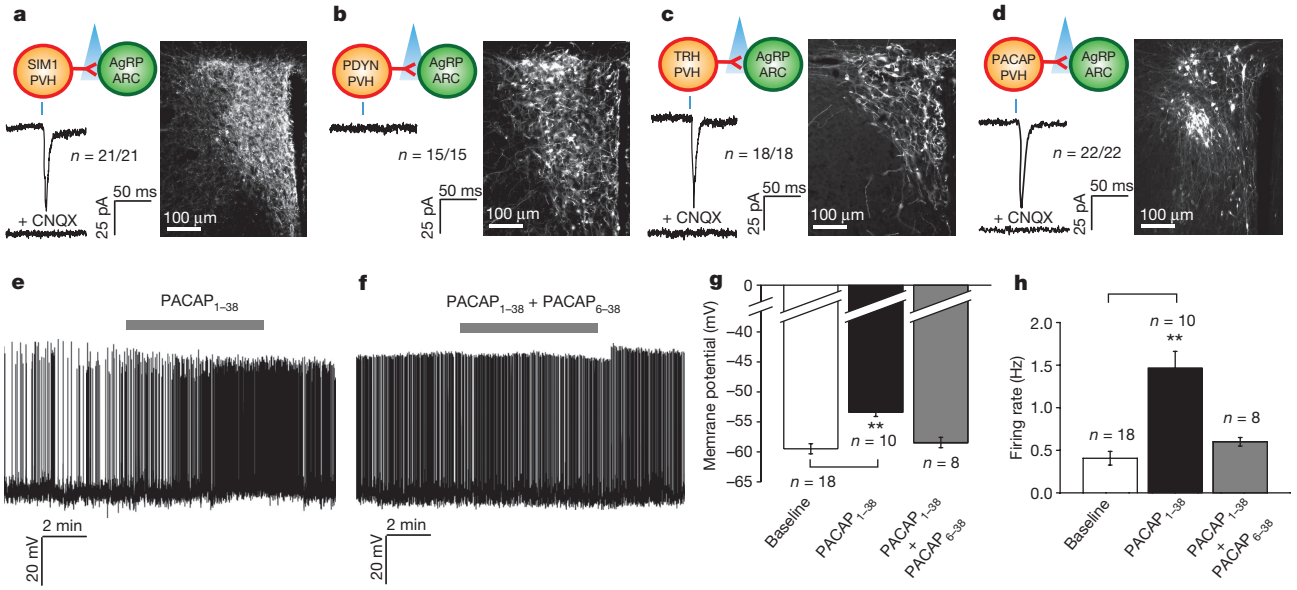


Figure 2 | TRH^{PVH} and PACAP^{PVH} neurons provide excitatory input to AgRP neurons. **a–d**, Top, schematic shows connections being tested. Right, representative brain sections of PVH injected with AAV8-DIO-ChR2–mCherry (white = ChR2–mCherry). Bottom, representative traces showing assessment of light-evoked EPSCs with blue tic indicating the light pulse (473 nm wavelength, 2 msec). Mice used were *Sim1-Cre; Npy-hrGFP* (**a**), *Pdyn-IRES-Cre; Npy-hrGFP* (**b**), *Trh-IRES-Cre; Npy-hrGFP* (**c**) and *Pacap-IRES-Cre; Npy-hrGFP* (**d**). CNQX is an AMPA receptor blocker. **e–h**, Baseline and effects of PACAP (100 nM) with or without PAC1R blocker ($PACAP_{6–38}$; 200 nM) on membrane potential and firing rate of AgRP neurons ($n = 18$, $n = 10$, $n = 8$, respectively). Results are shown as mean ± s.e.m. **g**, **h**, $n = 18$ for baseline group; $n = 10$ for $PACAP_{1–38}$ group; $n = 8$ for $PACAP_{1–38}$ + $PACAP_{6–38}$ group. *P* values for pair-wise comparisons (baseline versus $PACAP_{1–38}$ group) and unpaired comparisons (baseline versus $PACAP_{1–38}$ + $PACAP_{6–38}$ group) were calculated by two-tailed Student's *t*-test. $^{**}P < 0.001$.

not also engage satiety-promoting POMC neurons. We evaluated this and found that the vast majority of TRH^{PVH}→POMC^{ARC} neurons (21 of 22) and all PACAP^{PVH}→POMC^{ARC} neurons recorded are not monosynaptically connected (Fig. 3a, b, respectively). However, consistent with our observation that VGLUT2^{VMH}→POMC^{ARC} neurons are connected (Supplementary Fig. 3e), and as a positive control for studies investigating PACAP→POMC^{ARC} connections, we detected PACAP^{VMH}→POMC^{ARC} monosynaptic connections (latency = 4.5 ± 0.2 ms; Fig. 3c). These studies clearly demonstrate that the identified PVH→AgRP^{ARC} neuron circuit selectively drives hunger-promoting AgRP neurons, but not satiety-promoting POMC neurons.

Fidelity should also exist in the downstream connections of orexigenic AgRP neurons, which have reciprocal connections with the PVH. GABAergic AgRP neurons are monosynaptically connected with a subset of satiety-promoting neurons in the PVH, and this inhibitory connection drives feeding[5]. Here we confirm this inhibitory AgRP^{ARC}→PVH connection following ChR2 expression in AgRP neurons, as we detected light-evoked IPSCs (latency = 5.2 ± 0.3 ms) in a subset (55%) of AGRP^{ARC}→SIM1^{PVH} neurons (Fig. 3d), which were blocked by PTX. If fidelity of this hunger-promoting, GABAergic reciprocal AgRP^{ARC}→PVH connection is to be high, then it should not also engage and consequently inhibit the TRH^{PVH} and PACAP^{PVH} neurons. As expected, we detected no AgRP^{ARC}→TRH^{PVH} connections (Fig. 3e) and only rare AgRP^{ARC}→PACAP^{PVH} connections (1 of 12 neurons tested) (Fig. 3f). The above studies demonstrate marked fidelity in the reciprocal TRH/PACAP^{PVH} glutamatergic to AgRP^{ARC} GABAergic to satiety^{PVH} neuron circuit that drives feeding (Fig. 3g).

To confirm function of the TRH^{PVH} and PACAP^{PVH} neurons, we targeted DREADDs (designer receptors exclusively activated by designer drugs; AAV8-DIO-hM3Dq-mCherry)[3] to the PVH. Upon binding of

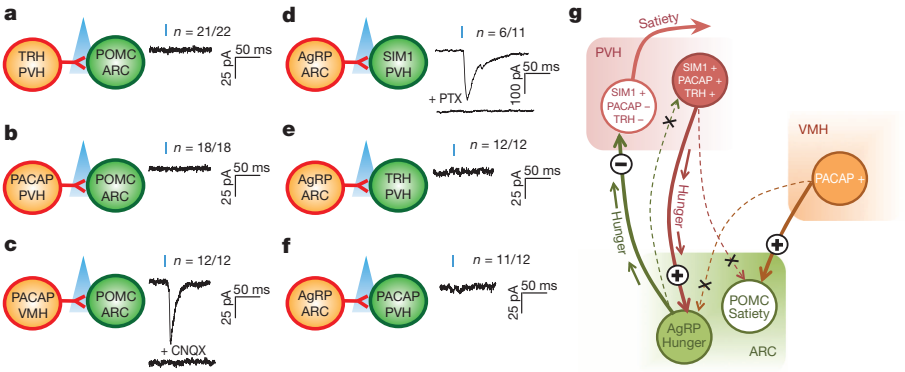


Figure 3 | Fidelity of TRH^{PVH}/PACAP^{PVH}→ARC and AgRP^{ARC}→PVH circuitry. **a–f**, Left, schematic shows connections being tested. Right, representative traces showing assessment of light-evoked EPSCs (**a–c**) or IPSCs (**d–f**) with blue tic indicating the light pulse (473 nm wavelength, 2 msec). Mice used were *Trh-IRES-Cre; Pomc-hrGFP* (**a**), *Pacap-IRES-Cre; Pomc-hrGFP* (**b, c**), *Agrp-IRES-Cre; Sim1-Cre; R26-loxSTOPlox-L10-GFP* (**d**) mice for visualization of SIM1 neurons, *Agrp-IRES-Cre; Trh-IRES-Cre* (**e**) mice injected with AAV8-DIO-mCherry into the PVH for visualization of TRH neurons and *Agrp-IRES-Cre; Pacap-IRES-Cre; R26-loxSTOPlox-L10-GFP* (**f**) mice for visualization of PACAP neurons. CNQX is an AMPA receptor blocker. PTX is a $GABA_A$ receptor blocker. **g**, Model summarizing reciprocal circuitry and its fidelity.

clozapine-N-oxide (CNO), hM3Dq activates neurons through the Gq signalling cascade[24]. After verifying in brain slices that CNO depolarized and increased firing frequency of TRH^{PVH} and PACAP^{PVH} neurons (Fig. 4a, d, respectively), we found that acutely stimulating upstream TRH^{PVH} or PACAP^{PVH} neurons markedly induced Fos activity in AgRP cells, ipsilateral to the DREADD-activated upstream PVH neurons (Fig. 4b, e, respectively).

We next assessed effects of bilateral h3MDq-DREADD-mediated TRH^{PVH} or PACAP^{PVH} neuron stimulation on feeding behaviour during the light cycle, a time when food intake is usually low because of feeding during the preceding dark cycle (as observed following saline injections, black lines in Fig. 4c, f, respectively). Importantly, acute activation by CNO injection of either TRH^{PVH} or PACAP^{PVH} neurons caused robust feeding (red lines in Fig. 4c, f, respectively). This effect was absent following an overnight fast which itself elevates excitatory drive and consequently AgRP neuron activity[1,8,9] (Supplementary Fig. 12). To directly demonstrate that this enhanced feeding was mediated through AgRP neurons, we activated TRH^{PVH} neurons (AAV8-DIO-hM3Dq-mCherry) while simultaneously inhibiting AgRP^{ARC} neurons (AAV8-DIO-hM4Di-mCherry[3]; Fig. 4g). This significantly attenuated feeding (Fig. 4h). Finally, to examine whether endogenous activity of the TRH^{PVH} →AgRP^{ARC} pathway is physiologically relevant for feeding, we bilaterally

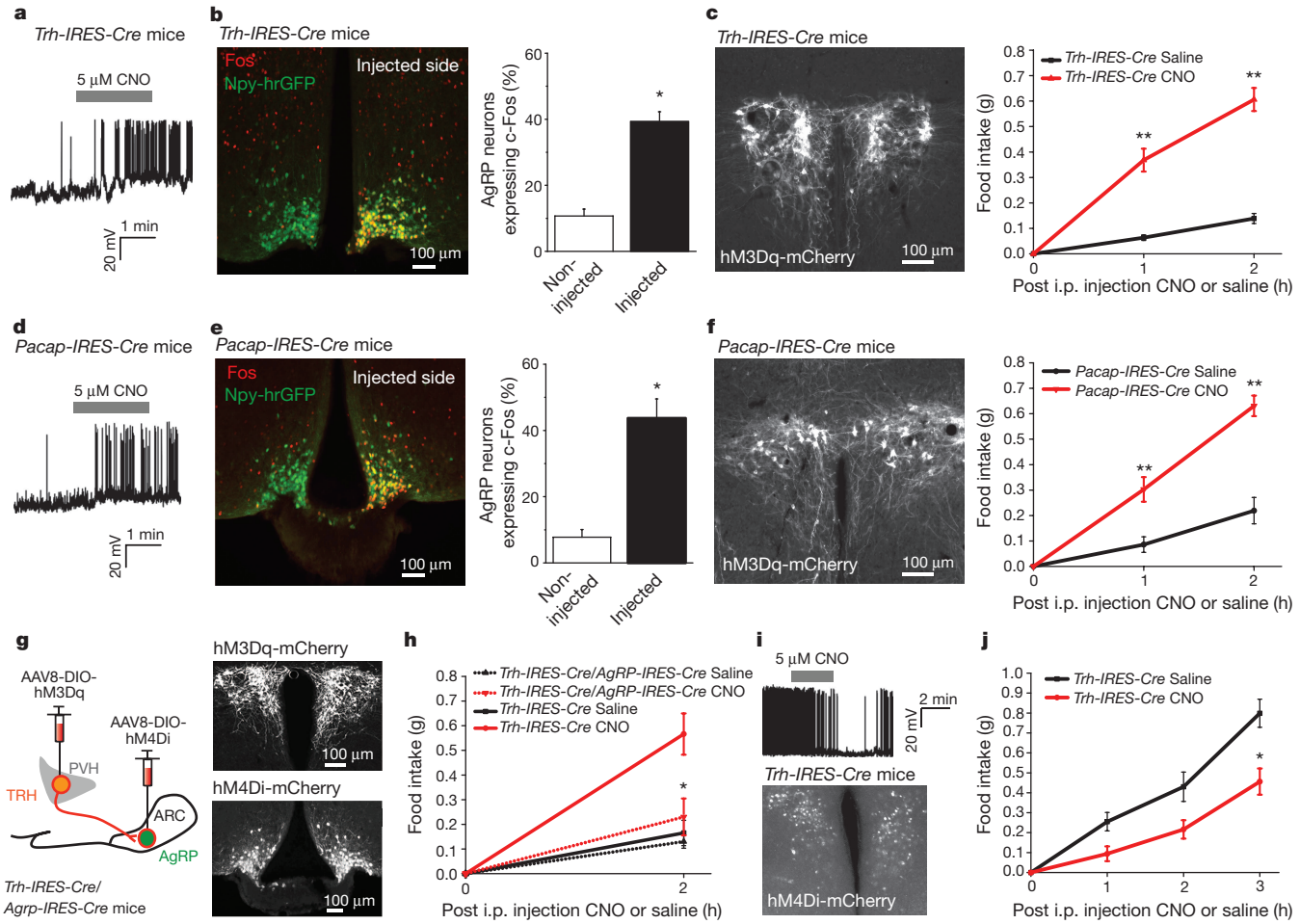


Figure 4 | DREADD-mediated manipulation of TRH^{PVH} or PACAP^{PVH} neurons mediates feeding through AgRP neurons. **a**, **d**, Membrane potential and firing rate of TRH^{PVH} (**a**) or PACAP^{PVH} (**d**) neurons transduced with AAV8-DIO-hM3Dq-mCherry upon CNO application. **b**, **e**, AAV8-DIO-hM3Dq-mCherry was transduced unilaterally into the PVH of *Trh-IRES-Cre* (**b**) or *Pacap-IRES-Cre* (**e**) mice and ipsilateral induction of Fos (red) was assessed in AgRP neurons (marked by hrGFP, green) 3 h following CNO (0.3 mg per kg) injection ($n = 3$, $n = 3$; respectively). Data represent mean ± s.e.m. *P* values for unpaired comparisons were calculated by two-tailed Student's *t*-test. $^*P < 0.05$. **c**, **f**, Left, AAV8-DIO-hM3Dq-mCherry (white) was transduced bilaterally into the PVH of *Trh-IRES-Cre* (**c**) or *Pacap-IRES-Cre* (**f**) mice. Right, light-cycle food intake after injection of saline (black) or CNO (red; 0.3 mg per kg) ($n = 7$–8 animals per condition; experiment replicated three times per animal). $n = 8$ for *Trh-IRES-Cre* group; $n = 7$ for *Pacap-IRES-Cre* group. Data represent mean ± s.e.m. Two-way repeated measures ANOVA detected significant interaction of 'time' and 'treatment'. $F_{3,20} = 56.13$, $^{**}P < 0.001$; $F_{3,18} = 0.056$, $^{**}P < 0.001$; respectively. Sidak's post-hoc test shows significant difference between 'time' and 'treatment' at 1 and 2 h as indicated, respectively. **g, h**, Simultaneous inhibition of AgRP^{ARC} neurons with activation of TRH^{PVH} neurons attenuates food intake. **g**, Sagittal schematic of occlusion study, whereby AAV8-DIO-hM3Dq-mCherry (white) was transduced bilaterally into the PVH (top) and AAV8-DIO-hM4Di-mCherry (white) was transduced bilaterally into the ARC (bottom) of *Trh-IRES-Cre*; *Agrp-IRES-Cre* mice. **h**, Light-cycle food intake after injection of saline (black) or CNO (red; 0.3 mg per kg) in *Trh-IRES-Cre*; *Agrp-IRES-Cre* (dotted; hM3Dq in PVH and hM4Di in ARC) and *Trh-IRES-Cre* (solid; hM3Dq in PVH) mice ($n = 4$ animals per condition; experiment replicated three times per animal). Data represent mean ± s.e.m. Two-way repeated measures ANOVA detected significant interaction of 'genotype' and 'treatment'. $F_{3,18} = 6.083$, $^*P < 0.05$. Sidak's post-hoc test shows significant difference between 'genotype' and 'treatment' as indicated. **i**, Membrane potential and firing rate of TRH^{PVH} neurons transduced with AAV8-DIO-hM4Di-mCherry upon CNO application (top). AAV8-DIO-hM4Di-mCherry (white) was transduced bilaterally into the PVH of *Trh-IRES-Cre* mice (bottom). **j**, Dark-cycle food intake after injection of saline (black) or CNO (red; 0.3 mg per kg) ($n = 4$ animals per condition; experiment replicated three times per animal). Results are shown as mean ± s.e.m., $^*P < 0.05$, $^{**}P < 0.001$; see Supplementary Information for statistical analyses. Data represent mean ± s.e.m. Two-way repeated measures ANOVA detected significant interaction of 'time' and 'treatment'. $F_{3,18} = 3.962$, $^*P < 0.05$. Sidak's post-hoc test shows significant difference between 'time' and 'treatment' at 3 h as indicated.

transduced TRH^{PVH} neurons with inhibitory DREADDs, which led to rapid hyperpolarization and decreased firing rate upon CNO application (as assessed in slice studies, Fig. 4i), and found that this manipulation drastically reduced food intake during the dark cycle (Fig. 4j). These studies demonstrate both the sufficiency and necessity of this $PVH \rightarrow AgRP^{ARC}$ orexigenic circuit in regulating food intake.

Given the vast number of studies implicating an anorexigenic role for the PVH[5,20,25–28], it is remarkable to now discover embedded with the PVH a subset of neurons that drive feeding, through a reciprocal circuit (Fig. 3g). Failure of prior studies to detect orexigenic activity in the PVH is probably due to the reciprocal nature of this hunger circuit (leaving and returning to the PVH) and the inhibitory aspect of its return arm ($AgRP^{ARC} \rightarrow satiety^{PVH}$ neurons).

METHODS SUMMARY

All experiments were conducted according to US National Institutes of Health guidelines for animal research and were approved by the Beth Israel Deaconess Medical Center Institutional Animal Care and Use Committee.

Mice. *Pdyn-IRES Cre, Avp-IRES-Cre, Crh-IRES-Cre, Trh-IRES-Cre* and *Pacap-IRES-Cre* and *R26-loxSTOPlox-L10-GFP* mice were generated using recombineering techniques as previously described[13,20].

Viral injections. Stereotaxic injections were performed as previously described[3,4,29].

Rabies cell counting. For each brain, images were taken throughout one entire brain series, $EGFP^{+}$ cells were quantified and assigned to a specific anatomical structure, and subsequently expressed as a percentage of the total number of cells counted for each mouse, averaged over the entire cohort.

Electrophysiology. Electrophysiology experiments were performed as previously described[29]. Latencies between onset of light and post synaptic currents are expressed as mean ± s.e.m.

***In situ* hybridization.** Digoxigenin-labelled riboprobes against *Pacap, Pdyn* or *Trh* were hybridized to brain sections of *Pacap-IRES-Cre; R26-loxSTOPlox-L10-GFP, Pdyn-IRES-Cre; R26-loxSTOPlox-L10-GFP* or *Trh-IRES-Cre* mice injected bilaterally with AAV-DIO-GFP into the PVH as previously described with minor modifications[30]. Signals from intrinsic Cre-mediated GFP expression and *in situ* hybridization were compared to determine the degree of co-localization.

Fos assay. Assessment of Fos induction used a previously developed method[8] modified for fluorescent co-localization with hrGFP in AgRP neurons.

Food intake. Food intake studies on chow were performed as previously described[3,4].

Statistics. Statistical analyses were performed using Origin Pro 8.6 and Prism 6.0 (GraphPad) software. Feeding studies were run as within-subject design and a final consumption value for each animal obtained from an average of 3 trials. Data was analysed using a two-way repeated-measures ANOVA, interaction of 'time' and 'treatment' or 'genotype' and 'treatment'. $*P < 0.05$, $**P < 0.001$. Error bars indicate mean +/− s.e.m. For slice electrophysiology experiments and cell counting analyses, *P* values for pair-wise or unpaired comparisons were calculated by two-tailed Student's *t*-test. $*P < 0.05$, $**P < 0.001$.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 April; accepted 16 December 2013.
Published online 2 February 2014.

1. Takahashi, K. A. & Cone, R. D. Fasting induces a large, leptin-dependent increase in the intrinsic action potential frequency of orexigenic arcuate nucleus neuropeptide Y/Agouti-related protein neurons. *Endocrinology* **146,** 1043–1047 (2005).
2. Aponte, Y., Atasoy, D. & Sternson, S. M. AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training. *Nature Neurosci.* **14,** 351–355 (2011).
3. Krashes, M. J. *et al.* Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *J. Clin. Invest.* **121,** 1424–1428 (2011).
4. Krashes, M. J., Shah, B. P., Koda, S. & Lowell, B. B. Rapid versus delayed stimulation of feeding by the endogenously released AgRP neuron mediators GABA, NPY, and AgRP. *Cell Metab.* **18,** 588–595 (2013).
5. Atasoy, D., Betley, J. N., Su, H. H. & Sternson, S. M. Deconstruction of a neural circuit for hunger. *Nature* **488,** 172–177 (2012).
6. Luquet, S., Perez, F. A., Hnasko, T. S. & Palmiter, R. D. NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* **310,** 683–685 (2005).
7. Gropp, E. *et al.* Agouti-related peptide-expressing neurons are mandatory for feeding. *Nature Neurosci.* **8,** 1289–1291 (2005).
8. Liu, T. *et al.* Fasting activation of AgRP neurons requires NMDA receptors and involves spinogenesis and increased excitatory tone. *Neuron* **73,** 511–522 (2012).
9. Yang, Y., Atasoy, D., Su, H. H. & Sternson, S. M. Hunger states switch a flip-flop memory circuit via a synaptic AMPK-dependent positive feedback loop. *Cell* **146,** 992–1003 (2011).
10. Pinto, S. *et al.* Rapid rewiring of arcuate nucleus feeding circuits by leptin. *Science* **304,** 110–115 (2004).
11. Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nature Methods* **4,** 47–49 (2007).
12. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74,** 858–873 (2012).
13. Vong, L. *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71,** 142–154 (2011).
14. Petreanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10,** 663–668 (2007).
15. Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A. FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28,** 7025–7030 (2008).
16. Hahn, T. M., Breininger, J. F., Baskin, D. G. & Schwartz, M. W. Coexpression of Agrp and NPY in fasting-activated hypothalamic neurons. *Nature Neurosci.* **1,** 271–272 (1998).
17. van den Pol, A. N. *et al.* Neuromedin B and gastrin-releasing peptide excite arcuate nucleus neuropeptide Y neurons in a novel transgenic mouse expressing strong Renilla green fluorescent protein in NPY neurons. *J. Neurosci.* **29,** 4622–4639 (2009).
18. Sternson, S. M., Shepherd, G. M. & Friedman, J. M. Topographic mapping of VMH→ arcuate nucleus microcircuits and their reorganization by fasting. *Nature Neurosci.* **8,** 1356–1363 (2005).
19. Michaud, J. L., Rosenquist, T., May, N. R. & Fan, C. M. Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1. *Genes Dev.* **12,** 3264–3275 (1998).
20. Balthasar, N. *et al.* Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell* **123,** 493–505 (2005).
21. Wu, Z. *et al.* An obligate role of oxytocin neurons in diet induced energy expenditure. *PLoS ONE* **7,** e45167 (2012).
22. Fukushima, Y. *et al.* Role of endogenous PACAP in catecholamine secretion from the rat adrenal gland. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **281,** R1562–R1567 (2001).
23. Zhan, C. *et al.* Acute and long-term suppression of feeding behavior by POMC neurons in the brainstem and hypothalamus, respectively. *J. Neurosci.* **33,** 3624–3632 (2013).
24. Alexander, G. M. *et al.* Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. *Neuron* **63,** 27–39 (2009).
25. Leibowitz, S. F., Hammer, N. J. & Chang, K. Hypothalamic paraventricular nucleus lesions produce overeating and obesity in the rat. *Physiol. Behav.* **27,** 1031–1040 (1981).
26. Gold, R. M., Jones, A. P. & Sawchenko, P. E. Paraventricular area: critical focus of a longitudinal neurocircuitry mediating food intake. *Physiol. Behav.* **18,** 1111–1119 (1977).
27. Stanley, B. G. & Leibowitz, S. F. Neuropeptide Y injected in the paraventricular hypothalamus: a powerful stimulant of feeding behavior. *Proc. Natl Acad. Sci. USA* **82,** 3940–3943 (1985).
28. Cowley, M. A. *et al.* Integration of NPY, AGRP, and melanocortin signals in the hypothalamic paraventricular nucleus: evidence of a cellular basis for the adipostat. *Neuron* **24,** 155–163 (1999).
29. Kong, D. *et al.* GABAergic RIP-Cre neurons in the arcuate nucleus selectively regulate energy expenditure. *Cell* **151,** 645–657 (2012).
30. Montmayeur, J. P., Liberles, S. D., Matsunami, H. & Buck, L. B. A candidate taste receptor gene near a sweet taste locus. *Nature Neurosci.* **4,** 492–498 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was funded by the following NIH grants to B.B.L.: R01 DK096010, R01 DK089044, R01 DK071051, R01 DK075632, R37 DK053477, BNORC Transgenic Core P30 DK046200 and BADERC Transgenic Core P30 DK57521; to M.J.K.: F32 DK089710; to D.P.O.: K08 DK071561; to L.V.: F32 DK078478; to N.U.: R01 MH095953; and an ADA Mentor-Based Fellowship to B.P.S. and B.B.L. We thank D. Cusher and Y. Li for mouse genotyping, J. Carroll and Y. Guo for perfusions and tissue removal, B. L. Roth and S. C. Rogan for generating the AAV-DIO-hM3Dq-mCherry plasmid, K. Deisseroth for generating the AAV-DIO-ChR2-mCherry plasmid and A. N. Hollenberg for the TRH riboprobe.

Author Contributions M.J.K., B.P.S. and B.B.L. designed the experiments and analysed data. M.J.K., B.P.S., J.C.M., D.E.S., A.S.G., L.V. and H.P. performed experiments. M.J.K. generated *Pdyn-IRES-Cre, Trh-IRES-Cre* and *Pacap-IRES-Cre* mice, D.P.O. generated *Oxt-IRES-Cre, Avp-IRES-Cre, Crh-IRES-Cre* and *R26-loxSTOPlox-L10-GFP* mice and L.V. generated *Vglut2-IRES-Cre* mice. M.W.-U. and N.U. generated, provided and advised on use of AAV-FLEX-TVA-mCherry, AAV-FLEX-RG and SADΔG–EGFP (EnvA) viruses. S.D.L. advised on *in situ* hybridization experiments. M.J.K and B.B.L. wrote the manuscript with comments from all of the authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.B.L. (blowell@bidmc.harvard.edu).

METHODS

Animals. All animal care and experimental procedures were approved by the Beth Israel Deaconess Medical Center Institutional Animal Care and Use Committee. Mice were housed at 22–24 °C with a 12 h light:12 h dark cycle with standard mouse chow (Teklad F6 Rodent Diet 8664; 4.05 kcal g^{-1}, 3.3 kcal g^{-1} metabolizable energy, 12.5% kcal from fat; Harlan Teklad) and water provided *ad libitum*. All diets were provided as pellets. Mice were euthanized by CO_2 narcosis.

Generation of mice. *Pdyn-IRES Cre, Avp-IRES-Cre, Crh-IRES-Cre, Trh-IRES-Cre* and *Pacap-IRES-Cre* mice were generated using recombineering techniques as previously described[13,21]. Briefly, a selection cassette containing an internal ribosomal entry sequence (IRES) linked to Cre- recombinase and an Frt-flanked kanamycin resistance gene was targeted just downstream of the stop codon of the prodynorphin, arginine vasopressin, corticotropin releasing hormone, thyrotropin releasing hormone or adenylate cyclase activating peptide 1 gene, respectively, in a bacterial artificial chromosome, so that Cre recombinase expression was driven by the endogenous genes. A targeting plasmid containing the Cre-containing selection cassette and 4 kb genomic sequence upstream and downstream of the prodynorphin, arginine vasopressin, corticotropin releasing hormone, thyrotropin releasing hormone or adenylate cyclase activating peptide 1 stop codon, respectively was isolated and used for embryonic stem cell targeting. Correctly targeted clones were identified by long range PCR and injected into blastocysts. Chimaeric animals generated from blastocyst implantation were then bred for germline transmission of the altered prodynorphin, arginine vasopressin, corticotropin releasing hormone, thyrotropin releasing hormone or adenylate cyclase activating peptide 1-allele, respectively. Flp-deletion mice were then used to remove the neomycin selection cassette.

Generation of an enhanced Cre-dependent GFP reporter mice (*R26-loxSTOPlox-L10-GFP*) were generated using recombineering techniques as previously described[20]. A transgene containing a lox-flanked transcriptional blocking cassette followed by EGFP fused to the L10-ribosomal subunit[31] was placed under the control of a CMV-enhancer/chicken β-actin promoter and targeted to the Rosa26 locus using standard techniques[32]. Correctly targeted blastocysts were identified by long range PCR and confirmed by Southern blotting and injected into blastocysts.

Characterization of mice. *Pdyn-IRES-Cre, Avp-IRES-Cre, Crh-IRES-Cre,* and *Pacap-IRES-Cre* mice were crossed to *R26-loxSTOPlox-L10-GFP* mice, euthanized and sectioned at 30 μm. Brain sections were washed in 0.1 M phosphate-buffered saline with Tween 20, pH 7.4 (PBST, 2 changes) and then incubated in the primary antiserum (anti-GFP, Abcam (1:1,000), rabbit anti-hrGFP. After several washes in PBS, sections were incubated in Alexa fluorophore secondary antibody (Molecular Probes, 1:200) for 2 h at room temperature. After several washes in PBS, sections were mounted onto gelatin-coated slides and fluorescent images were captured with Olympus VS120 slide scanner microscope.

Trh-IRES-Cre mice were stereotaxically injected with Cre-dependent AAV8-DIO-mCherry[8], due to transient, early embryonic expression and subsequent deletion of floxed alleles by this mouse line, euthanized and sectioned at 30 μm. Native fluorescence was used and fluorescent images were captured with Olympus VS120 slide scanner microscope. All digital images were processed in the same way between experimental conditions to avoid artificial manipulation between different data sets.

In situ mRNA hybridization experiments were performed for *Pdyn-IRES-Cre, Trh-IRES-Cre* and *Pacap-IRES-Cre* mice (details below). Antibody staining experiments were performed for *Avp-IRES-Cre* and *Crh-IRES-Cre* mice. Briefly, *Avp-IRES-Cre; R26-loxSTOPlox-L10-GFP* mice were co-localized with anti-AVP, Sigma (1:500), rabbit anti-AVP and anti-GFP, Abcam (1:1,000), chicken anti-GFP. After 48 h intracerebroventricular colchicine treatment (Sigma, 10 μg), *Crh-IRES-Cre; R26-loxSTOPlox-L10-GFP* mice were co-localized with anti-CRF, Peninsula Laboratories (1:2,500), rabbit anti-CRF and anti-GFP, Abcam (1:1,000), chicken anti-GFP.

All images were subsequently compared to *in situ* mRNA expression profiles generated by the Allen Institute for Brain Science; Allen Mouse Brain Atlas 2013 (http://mouse.brain-map.org/).

Breeding strategies. *Vglut2-IRES-Cre, Sim1-Cre, Pdyn-IRES-Cre, Oxt-IRES-Cre, Avp-IRES-Cre, Crh-IRES-Cre, Trh-IRES-Cre* and *Pacap-IRES-Cre* mice were bred to *Npy-hrGFP* transgenic mice to generate heterozygous *Vglut2-IRES-Cre; Npy-hrGFP, Sim1-Cre; Npy-hrGFP, Pdyn-IRES-Cre; Npy-hrGFP, Oxt-IRES-Cre; Npy-hrGFP, Avp-IRES-Cre; Npy-hrGFP, Crh-IRES-Cre; Npy-hrGFP, Trh-IRES-Cre; Npy-hrGFP* and *Pacap-IRES-Cre; Npy-hrGFP* mice, respectively as previously reported[8]. *Vglut2-IRES-Cre* and *Pacap-IRES-Cre* mice were bred to *Pomc-hrGFP* transgenic mice to generate heterozygous *Vglut2-IRES-Cre; Pomc-hrGFP* and *Pacap-IRES-Cre; Pomc-hrGFP* mice, respectively as previously reported[13].

Agrp-IRES-Cre, Pdyn-IRES-Cre, Avp-IRES-Cre, Crh-IRES-Cre and *Pacap-IRES-Cre* mice were bred to *R26-loxSTOPlox-L10-GFP* mice to generate heterozygous *Agrp-IRES-Cre; R26-loxSTOPlox-L10-GFP, Pdyn-IRES-Cre; R26-loxSTOPlox-L10-GFP, Avp-IRES-Cre; R26-loxSTOPlox-L10-GFP, Crh-IRES-Cre; R26-loxSTOPlox-L10-GFP* and *Pacap-IRES-Cre; R26-loxSTOPlox-L10-GFP* mice, respectively. *Agrp-IRES-Cre; R26-loxSTOPlox-L10-GFP* mice were then bred to *Sim1-Cre* and *Pacap-IRES-Cre* to generate heterozygous *Agrp-IRES-Cre; Sim1-Cre R26-loxSTOPlox-L10-GFP* and *Agrp-IRES-Cre; Pacap-IRES-Cre; R26-loxSTOPlox-L10-GFP* mice, respectively. All mice are on a mixed background.

Viral injections. Stereotaxic injections were performed as previously described[3]. Mice were anaesthetized with xylazine (5 mg per kg) and ketamine (75 mg per kg) diluted in saline (350 mg per kg) and placed into a stereotaxic apparatus (KOPF Model 963). After exposing the skull via small incision, a small hole was drilled for injection. A pulled-glass pipette with 20–40 mm tip diameter was inserted into the brain and virus was injected by an air pressure system. A micromanipulator (Grass Technologies, Model S48 Stimulator) was used to control injection speed at 25 nl min^{-1} and the pipette was withdrawn 5 min after injection. For retrograde rabies tracing, AAV-FLEX-TVA-mCherry, serotype 8 (titre 1.1×10^{12} genomes copies per ml) and AAV-FLEX-RG, serotype 8 (titre 1.4×10^{12} genomes copies per ml)[12] were injected unilaterally into the ARC (200 nl, coordinates, bregma: AP: −1.50 mm, DV: −5.80 mm, L: −0.20 mm) of 5–6-weeks-old mice. Then 21 days later SADΔG–EGFP (EnvA) rabies (titre 10^7 genomes copies per ml) was unilaterally injected into the ARC (400 nl, coordinates, bregma: AP: −1.50 mm, DV: −5.80 mm, L: −0.20 mm). For electrophysiology experiments, AAV-DIO-ChR2(H134R)-mCherry, serotype 8 (titre 1.3×10^{12} genomes copies per ml)[29] was injected unilaterally into the DMH (20 nl, coordinates, bregma: AP: −1.70 mm, DV: −5.00 mm, L: −0.25 mm), PVH (25 nl, coordinates, bregma: AP: −0.70 mm, DV: −4.75 mm, L: −0.20 mm), VMH (25 nl, coordinates, bregma: AP: −1.40 mm, DV: −5.60 mm, L: −0.40 mm) or ARC (20 nl, coordinates, bregma: AP: –1.50 mm, DV: −5.80 mm, L: −0.20 mm) of 5–6-weeks-old mice. For feeding and Fos studies, AAV-DIO-hM3Dq-mCherry, serotype 8 (titre 1.2×10^{12} genomes copies per ml)[3] or AAV-DIO-hM4Di-mCherry, serotype 8 (titre 1.7×10^{12} genomes copies per ml)[3] was injected bilaterally or unilaterally, respectively into the PVH (25 nl, coordinates, bregma: AP: –0.70 mm, DV: −4.75 mm, L: −0.20 mm) of 5–6-weeks-old mice. For validation of *Trh-IRES-Cre* mice, AAV-DIO-mCherry, serotype 8 (titre 1.2×10^{12} genomes copies per ml)[8] or AAV-DIO-GFP, serotype 8 (titre 1.4×10^{12} genomes copies per ml) was injected bilaterally into the PVH (25 nl, coordinates, bregma: AP: −0.70 mm, DV: −4.75 mm, L: −0.20 mm) of 5−6-weeks-old mice. All viruses were packaged at the Gene Therapy Center at the University of North Carolina. For postoperative care, mice were injected intraperitoneally with meloxicam (0.5 mg per kg). All stereotaxic injection sites were verified under electrophysiological microscopy (for electrophysiology-related studies) or by immunohistochemistry (for anatomy and *in vivo* studies). All 'misses' or 'partial hits' were excluded from data analyses.

SADΔG–EGFP (EnvA) rabies cell counting. One week after SADΔG–EGFP (EnvA) rabies injection, mice were perfused and brains were sectioned and mounted as described above. For each brain ($n = 6$), ×10 magnification images were taken throughout one entire brain series using a Zeiss AxioImager Z.1 microscope and $EGFP^+$ cells were quantified using these images. Each $EGFP^+$ cell was assigned to a specific anatomical structure of the hypothalamus using The Mouse Brain in Stereotaxic Coordinates (Franklin & Paxinos). As the number of labelled cell varied over multiple animals depending on the transduction rate of the viruses, the counted cells in each anatomical structure were expressed as a percentage of the total number of cells counted throughout each mouse brain. The percentages for each anatomical structure in a given mouse were then averaged over the entire cohort of mice (Supplementary Fig. 2).

Electrophysiology and circuit mapping. For brain slice preparation, 6–10-weeks-old mice were anaesthetized with isoflurane before decapitation and removal of the entire brain. Brains were immediately submerged in ice-cold, carbogen-saturated (95% O_2, 5% CO_2) high sucrose solution (238 mM sucrose, 26 mM $NaHCO_3$, 2.5 mM KCl, 1.0 mM NaH_2PO_4, 5.0 mM $MgCl_2$, 10.0 mM $CaCl_2$, 11 mM glucose). Then, 300-μM thick coronal sections were cut with a Leica VT1000S Vibratome and incubated in oxygenated aCSF (126 mM NaCl, 21.4 mM $NaHCO_3$, 2.5 mM KCl, 1.2 mM NaH_2PO_4, 1.2 mM $MgCl_2$, 2.4 mM $CaCl_2$, 10 mM glucose) at 34 °C for 30 min. Then, slices were maintained and recorded at room temperature (20–24 °C). The intracellular solution for voltage clamp recording contained the following (in mM): 140 CsCl, 1 BAPTA, 10 HEPES, 5 $MgCl_2$, 5 Mg-ATP, 0.3 Na_2GTP, and 10 lidocaine *N*-ethyl bromide (QX-314), pH 7.35 and 290 mOsm. The intracellular solution for current clamp recordings contained the following (in mM): 128 K gluconate, 10 KCl, 10 HEPES, 1 EGTA, 1 $MgCl_2$, 0.3 $CaCl_2$, 5 Na_2ATP, 0.3 NaGTP, adjusted to pH 7.3 with KOH.

Photostimulation-evoked EPSCs and IPSCs were recorded in the whole cell voltage clamp mode, with membrane potential clamped at −60 mV. Photostimulation-evoked EPSCs was recorded in presence of picrotoxin (100 μM) to block inhibitory postsynaptic currents. All recordings were made using multiclamp 700B amplifier, and data were filtered at 2 kHz and digitized at 10 kHz. To photostimulate channelrhodopsin2-positive fibres, a laser or LED light source (473 nm; Opto

Engine; Thorlabs) was used. The blue light was focused on to the back aperture of the microscope objective, producing a wide-field exposure around the recorded cell of 10–15 mW mm^2. The light power at the specimen was measured using an optical power meter PM100D (Thorlabs). The light output is controlled by a programmable pulse stimulator, Master-8 (A.M.P.I.) and the pClamp 10.2 software (Axon Instruments). Photostimulation-evoked EPSCs/IPSCs detection protocol constitutes four blue light laser pulses administered 1 s apart during the first 4 s of a 8 s sweep, repeating for a total of 30 sweeps. When recording photostimulation-evoked action potentials in AgRP neurons, current was injected into cells to keep the base line membrane potential at approximately −60 mV. Mice with total misses, partial expression or expression outside the intended area were excluded from analysis after examination of mCherry expression. $VGLUT2^{DMH}\rightarrow AgRP^{ARC}$ $n = 4$; $VGLUT2^{PVHH}\rightarrow AgRP^{ARC}$ $n = 5$; $VGLUT2^{VMH}\rightarrow AgRP^{ARC}$ $n = 4$; $VGLUT2^{LH}\rightarrow AgRP^{ARC}$ $n = 3$; $VGLUT2^{ARC}\rightarrow AgRP^{ARC}$ $n = 3$; $VGLUT2^{VMH}\rightarrow POMC^{ARC}$ $n = 3$; $SIM1^{PVH}\rightarrow AgRP^{ARC}$ $n = 3$; $PDYN^{PVH}\rightarrow AgRP^{ARC}$ $n = 3$; $OXT^{PVH}\rightarrow AgRP^{ARC}$ $n = 3$; $AVP^{PVH}\rightarrow AgRP^{ARC}$ $n = 3$; $CRH^{PVH}\rightarrow AgRP^{ARC}$ $n = 3$; $TRH^{PVH}\rightarrow AgRP^{ARC}$ $n = 3$; $PACAP^{PVH}\rightarrow AgRP^{ARC}$ $n = 4$; $TRH^{PVH}\rightarrow POMC^{ARC}$ $n = 4$; $PACAP^{PVH}\rightarrow POMC^{ARC}$ $n = 3$; $PACAP^{VMH}\rightarrow POMC^{ARC}$ $n = 2$; $AgRP^{ARC}\rightarrow SIM1^{PVH}$ $n = 2$; $AgRP^{ARC}\rightarrow TRH^{PVH}$ $n = 3$; $AgRP^{ARC}\rightarrow PACAP^{PVH}$ $n = 3$.

To assess the effect of $PACAP_{1-38}$ (100 nM)[22] and $PACAP_{6-38}$ (200 nM) onto AgRP neurons, we performed whole cell current clamp recordings on to 5–8-weeks-old *Npy-hrGFP mice*. Synaptic blockers (1 mM kynurenate and 100 μM picrotoxin) were added in aCSF to synaptically isolate AgRP neurons.

To assess the effect of CNO on to TRH and PACAP neurons, 5- to 7-week-old *Pacap-IRES-Cre and Trh-IRES-Cre* mice were injected with AAV8-DIO-hM3Dq-mCherry into the PVH 2–3 weeks before recording. CNO was applied to bath solution through perfusion as previously reported[3]. After acquisition of stable whole-cell recordings for 2–5 min, aCSF solution containing 5 μM CNO was perfused into the brain slice preparation.

Fos analysis. Animals (*Trh-IRES-Cre; Npy-hrGFP*; $n = 3$ and *Pacap-IRES-Cre; Npy-hrGFP*; $n = 3$ mice), were handled for 10 consecutive days before the assay to reduce stress response, and then injected with CNO (0.3 mg per kg; i.p.) at 9:00. Then 150 min later, the animals were euthanized with 7% chloral hydrate diluted in saline (350 mg per kg) for histological assay. The mice were perfused and brains were sectioned as described above. Assessment of Fos induction was performed using a previously developed method[8] modified for fluorescent co-localization with hrGFP in AgRP neurons. Brain sections were processed for immunohistochemical detection of Fos and hrGFP and counting.

Brain sections were washed in 0.1 M phosphate-buffered saline with Tween 20, pH 7.4 (PBST, 2 changes) and then incubated in the primary antiserum (anti-GFP, Abcam (1:1,000), rabbit anti-hrGFP and anti-Fos, Calbiochem (1:10,000). After several washes in PBS, sections were incubated with Alexa fluorophore secondary antibodies (Molecular Probes, 1:200) for 2 h at room temperature. After several washes in PBS, sections were mounted onto gelatin-coated slides and fluorescent ×10 magnification images were taken initially using a Zeiss AxioImager Z.1 microscope. Co-localization and quantification was further determined using fluorescent ×20 magnification images taken using a Zeiss AxioImager Z.1 microscope. Data are expressed as the percentage of all AgRP neurons (that is, all hrGFP-positive neurons) that were double-positive for Fos and hrGFP.

Food intake studies. Food intake studies on chow were performed as previously described[3,8,29]. All animals (10- to 12-week-old male mice) were singly housed for at least 2.5 weeks following surgery and handled for 10 consecutive days before the assay to reduce stress response. Feeding studies were performed in home cages with *ad libitum* food access. CNO was administered at 0.3 mg per kg of body weight. Saline was delivered at the same volume to maintain consistency in the studies. Previous publications suggest the duration of the drugs effect at the dosage used throughout the studies is approximately 8 h[24]. Mice with 'missed' injections, incomplete 'hits' or expression outside the area of interest were excluded from analysis after post hoc examination of mCherry expression. In this way, all food intake measurements were randomized and blind to the experimenter.

For light cycle measurements, animals (*Trh-IRES-Cre*; $n = 8$, *Pacap-IRES-Cre*; $n = 7$ mice; Fig. 4c, f, or *Trh-IRES-Cre; Agrp-IRES-Cre*; $n = 4$, *Trh-IRES-Cre*; $n = 4$ mice; Fig. 4h) were injected with either saline or CNO (0.3 mg per kg; i.p.) at 9:00 and food intake was monitored 1 h and/or 2 h after i.p. injection from 9:00 to 11:00. A full trial consisted of assessing food intake from the study subjects after they received injections of saline on day 1 and CNO on day 2. Animals received a day 'off' between trials before another trial was initiated. The food intake data from all days following saline/CNO ($n = 24$ trials for *Trh-IRES-Cre*; $n = 21$ trials for *Pacap-IRES-Cre*; Fig. 4c, f; $n = 12$ for *Trh-IRES-Cre; Agrp-IRES-Cre;* $n = 12$ trials for *Trh-ires-Cre*; Fig. 4h) injections were then averaged and combined for analysis.

For dark cycle measurements, animals (*Trh-IRES-Cre*; $n = 4$ mice; Fig. 4j) were injected with either saline or CNO (0.3 mg per kg; i.p.) at 18:00 and food intake was monitored 1, 2 and 3 h after i.p. injection from 18:00 to 21:00. A full trial consisted of assessing food intake from the study subjects after they received injections of saline on day 1 and CNO on day 2. Animals received a day 'off' between trials before another trial was initiated. The food intake data from all days following saline/CNO ($n = 12$ trials for *Trh-IRES-Cre*; Fig. 4j) injections was then averaged and combined for analysis.

For fast-refeed measurements, animals (*Trh-IRES-Cre*; $n = 8$, *Pacap-IRES-Cre*; $n = 7$ mice; Supplementary Fig. 12) were fasted overnight at 17:00 and the following day were injected with either saline or CNO (0.3 mg per kg; i.p.) at 8:30 and food intake was monitored 1.5, 2.5 and 4.5 h after i.p. injection from 9:00 to 13:00. A full trial consisted of assessing food intake from the study subjects after they received injections of saline on week 1 and CNO on week 2. Animals received a week 'off' between trials before another trial was initiated. The food intake data from all days following saline/CNO ($n = 16$ trials for *Trh-IRES-Cre*; $n = 14$ trials for *Pacap-IRES-Cre*; Supplementary Fig. 12) injections were then averaged and combined for analysis.

***In situ* hybridization.** *Pacap-IRES-Cre; R26-loxSTOPlox-L10-GFP* and P*dyn-IRES-Cre; R26-loxSTOPlox-L10-GFP* mice were euthanized and brains were frozen fresh. *Trh-IRES-Cre* mice with bilateral AAV-DIO-GFP injections into the PVH were perfused with 4% PFA, post-fixed for 2 h and equilibrated in 20% sucrose overnight. Then 14-μM cryosections were analysed by *in situ* hybridization as previously described[30] with minor modifications. Slides were fixed (4% PFA), washed (2× RNase-free PBS), coverslips were added and samples imaged for intrinsic Cre-mediated GFP fluorescence. Digoxigenin labelled anti-sense cRNA probes were generated by T3 (Roche) *in vitro* transcription reactions using the complete coding sequence of *Pdyn* (5′ primer: 5′-ATGGCGTGGTCCAGGCTGATGC-3′; 3′ primer: 5′-TCAAACATCTAAATCTTCAGAATAGG-3′) and a 914-bp fragment of *Pacap* cDNA (5′ primer: 5′-CTGCGTGACGCTTACGCCCT-3′; 3′ primer: 5′-TTGCCCCTGCAACCAGTGGG-3′). A previously described murine cDNA[33] (gift of Masanobu Yamada) was used to generate a *Trh* probe with SP6 RNA polymerase (Roche). Following hybridization, slides were incubated with anti-digoxigenin antibody conjugated to alkaline phosphatase (Roche, 1:200, 1 h room temperature), washed and incubated (3–6 h) in NBT/BCIP chromogenic substrate according to the manufacturer's specifications (Roche). Slides had coverslips added and brightfield images were captured. Images were pseudocoloured and compared with intrinsic Cre-mediated GFP fluorescence for co-localization and quantification.

Statistical analysis. Statistical analyses were performed using Origin Pro 8.6 and Prism 6.0 (GraphPad) software.

31. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135,** 738–748 (2008).
32. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1,** 4 (2001).
33. Satoh, T., Yamada, M., Monden, T., Iizuka, M. & Mori, M. Cloning of the mouse hypothalamic preprothyrotropin-releasing hormone (TRH) cDNA and tissue distribution of its mRNA. *Brain Res. Mol. Brain Res.* **14,** 131–135 (1992).

LETTER

doi:10.1038/nature12967

L-Myc expression by dendritic cells is required for optimal T-cell priming

Wumesh KC[1], Ansuman T. Satpathy[1], Aaron S. Rapaport[1], Carlos G. Briseño[1], Xiaodi Wu[1], Jörn C. Albring[2], Emilie V. Russler-Germain[1], Nicole M. Kretzer[1], Vivek Durai[1], Stephen P. Persaud[1], Brian T. Edelson[1], Jakob Loschko[3], Marina Cella[1], Paul M. Allen[1], Michel C. Nussenzweig[3], Marco Colonna[1], Barry P. Sleckman[1], Theresa L. Murphy[1] & Kenneth M. Murphy[1,4]

The transcription factors c-Myc and N-Myc—encoded by *Myc* and *Mycn*, respectively—regulate cellular growth[1] and are required for embryonic development[2,3]. A third paralogue, *Mycl1*, is dispensable for normal embryonic development but its biological function has remained unclear[4]. To examine the *in vivo* function of *Mycl1* in mice, we generated an inactivating *Mycl1*^{gfp} allele that also reports *Mycl1* expression. We find that *Mycl1* is selectively expressed in dendritic cells (DCs) of the immune system and controlled by IRF8, and that during DC development, *Mycl1* expression is initiated in the common DC progenitor[5] concurrent with reduction in c-Myc expression. Mature DCs lack expression of c-Myc and N-Myc but maintain L-Myc expression even in the presence of inflammatory signals such as granulocyte–macrophage colony-stimulating factor. All DC subsets develop in *Mycl1*-deficient mice, but some subsets such as migratory CD103$^+$ conventional DCs in the lung and liver are greatly reduced at steady state. Importantly, loss of L-Myc by DCs causes a significant decrease in *in vivo* T-cell priming during infection by *Listeria monocytogenes* and vesicular stomatitis virus. The replacement of c-Myc by L-Myc in immature DCs may provide for Myc transcriptional activity in the setting of inflammation that is required for optimal T-cell priming[6].

c-Myc regulates cellular proliferation, metabolism and maintenance of progenitor populations[7,8] and globally amplifies transcription by direct interactions with RNA polymerases I, II and III, accounting for its regulation of disparate and context-dependent target loci across cell types[9–11]. Forced expression of L-Myc exerts weaker effects than c-Myc for cell growth, apoptosis and transformation[12] but is more efficient in reprogramming fibroblasts towards induced pluripotent stem cells[13]. However, the *in vivo* function of L-Myc has not been established[4].

Mature DCs exhibit substantial proliferative activity[14,15]. In a 4-h *in vivo* 5-bromo-2′deoxyuridine (BrdU) pulse labelling, B cells, monocytes and neutrophils showed a low rate of ~1% BrdU uptake (Extended Data Fig. 1a), whereas splenic conventional DC (cDC) subsets showed 4–8% BrdU uptake, consistent with previous studies[14,15]. In agreement, 4–7% of cDCs were found to be in S/G2/M phase by DAPI (4′,6-diamidino-2-phenylindole) staining and 17–34% were in active cell cycle by Ki-67 staining (Extended Data Fig. 1b). Although splenic plasmacytoid DCs (pDCs) had little proliferative capacity, a significant fraction of pDCs in bone marrow were in S/G2/M phase by DAPI staining (Extended Data Fig. 1c).

Myc and *Mycn* were highly expressed in various haematopoietic progenitor populations but were significantly reduced in mature DCs (Fig. 1a). By contrast, *Mycl1* was expressed by common DC precursors (CDPs), committed precursors to cDCs[5] (pre-cDCs) and by mature splenic DCs (Fig. 1a), but not by other haematopoietic lineages[8] (Extended Data Fig. 1d, e), indicating that L-Myc expression replaces c-Myc and

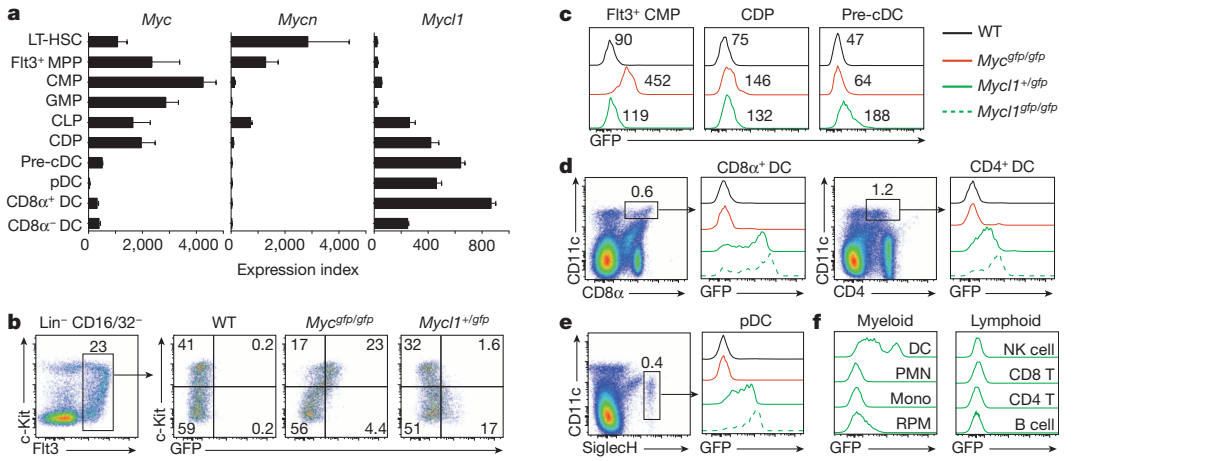


Figure 1 | DCs selectively express *Mycl1* but not *Myc* or *Mycn*. **a**, Shown are mean values for *Myc*, *Mycn* and *Mycl1* expression in progenitors and DC subsets as described in the Methods (error bars, s.d., $n = 2$–4 biological replicates). CLP, common lymphoid progenitor; GMP, granulocyte–macrophage progenitor; LT-HSC, long-term haematopoietic stem cell; MPP, multipotent progenitor. **b**, c-Kit and GFP expression by Flt3$^+$ Lin$^-$ CD16/CD32$^-$ bone marrow cells from wild-type (WT), *Myc*$^{gfp/gfp}$ and *Mycl1*$^{+/gfp}$ mice. **c**, c-Myc–GFP and L-Myc–GFP expression in the indicated progenitors and mice. Numbers indicate mean fluorescence intensities. **d**–**f**, GFP expression for mice in **c** for the indicated cell populations. Data are representative of at least 4 experiments ($n = 10$ mice). Mono, monocyte; PMN, neutrophil; RPM, red pulp macrophage.

[1]Department of Pathology and Immunology, Washington University School of Medicine, 660 S. Euclid Avenue, St Louis, Missouri 63110, USA. [2]Department of Medicine A, Hematology and Oncology, University of Muenster, 48149 Muenster, Germany. [3]Laboratory of Molecular Immunology, Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA. [4]Howard Hughes Medical Institute, Washington University School of Medicine, 660 S. Euclid Avenue, St Louis, Missouri 63110, USA.

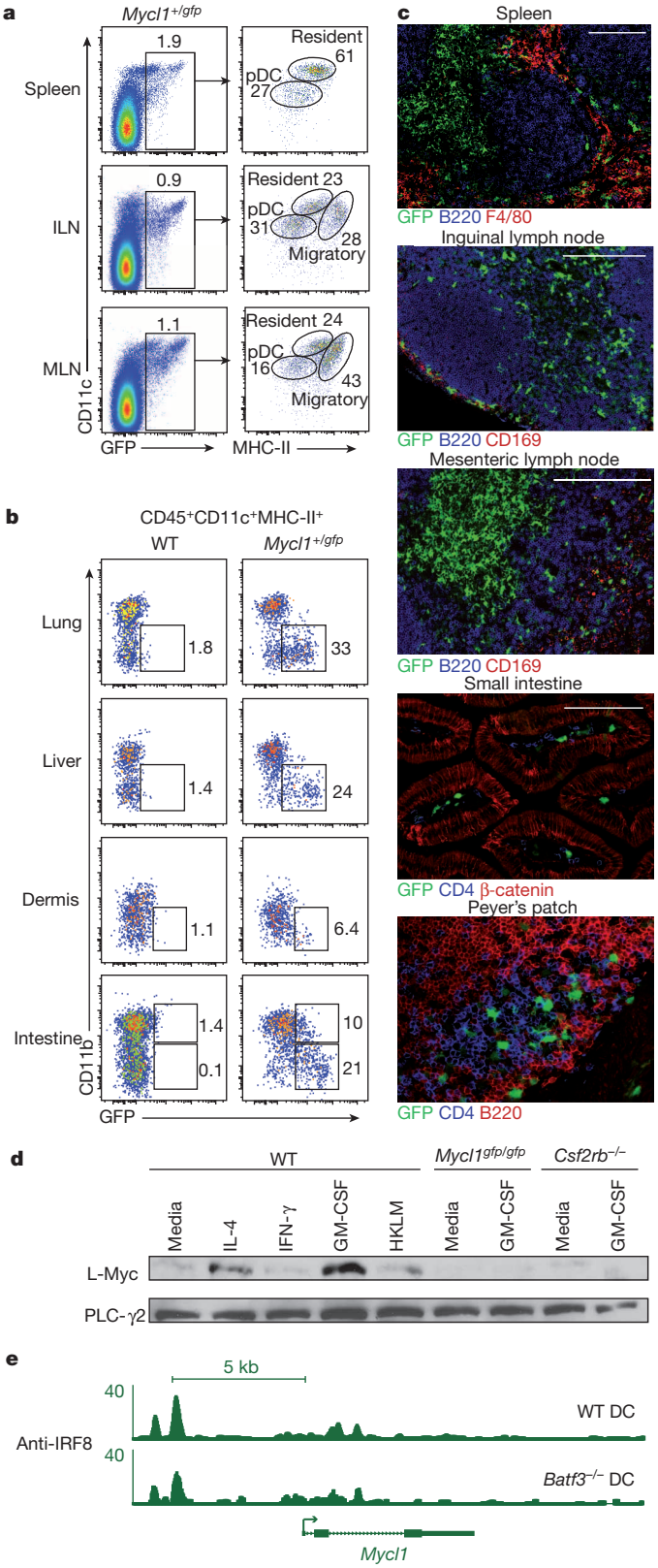


Figure 2 | *Mycl1* is restricted to DCs in lymphoid and peripheral tissues and regulated by IRF8 and GM-CSF. **a**, CD11c and major histocompatibility complex class II (MHC-II) expression is shown for GFP$^+$ cells from spleen, inguinal lymph node (ILN) and mesenteric lymph node (MLN) from *Mycl1*$^{+/gfp}$ mice ($n = 5$ mice). **b**, CD11b and GFP expression in CD45$^+$ CD11c$^+$ MHC-II$^+$ cells from indicated tissues of wild-type (WT) and *Mycl1*$^{+/gfp}$ mice ($n = 5$ mice). **c**, Immunofluorescence for indicated markers from *Mycl1*$^{+/gfp}$ mice. Scale bars, 200 μm. **d**, Western blot analysis for L-Myc and phospholipase C (PLC)-γ2 from Flt3L-derived DCs from WT, *Mycl1*$^{gfp/gfp}$ and *Csf2rb*$^{-/-}$ mice, treated as indicated for 24 h. HKLM, heat-killed *L. monocytogenes* strain EGD. Data in **a**–**d** are representative of at least 3 experiments. **e**, IRF8 binding in the *Mycl1* locus determined by ChIP-seq in wild-type or *Batf3*$^{-/-}$ DCs. Numbers represent normalized reads.

N-Myc during DC development. We compared c-Myc and L-Myc expression at a single-cell resolution using *Myc*^{gfp} reporter mice[16] encoding a green fluorescent protein (GFP)–c-Myc amino-terminal fusion protein and a novel L-Myc allele, *Mycl1*^{gfp}, that substitutes *gfp* for the first coding exon (Extended Data Fig. 1f and Extended Data Fig. 2). c-Myc protein was highly expressed in Flt3$^+$ common myeloid progenitors (CMPs), was greatly reduced in CDPs and pre-cDCs (Fig. 1b, c) and was undetectable in mature splenic DCs (Fig. 1d, e). By contrast, *Mycl1*^{gfp} was absent in CMPs, became detectible in CDPs and pre-cDCs (Fig. 1b, c), and was highly expressed in mature splenic CD8α$^+$ cDCs, CD8α$^-$ cDCs, and pDCs, but not in neutrophils, monocytes, red pulp macrophages, natural killer (NK) cells and T and B cells (Fig. 1d–f), consistent with *Mycl1* messenger RNA expression (Extended Data Fig. 1d).

DC subsets that developed in *Mycl1*$^{gfp/gfp}$ (L-Myc-deficient) mice showed no compensatory induction of *Myc* expression (Extended Data Fig. 1g, h). *Mycl1*^{gfp} expression was observed in DCs that developed from Flt3 ligand (Flt3L)-treated bone marrow cultures *in vitro* (Extended Data Fig. 3a, b). Retroviral overexpression of c-Myc, but not L-Myc, into Flt3$^+$ CMPs reduced the proportion of mature cDCs and pDCs that developed in Flt3L cultures (Extended Data Fig. 3c, d), suggesting that L-Myc may be non-redundant with c-Myc for DC development.

We also compared *Myc*^{gfp} and *Mycl1*^{gfp} expression in other tissues. In inguinal and mesenteric lymph nodes, *Mycl1*^{gfp}, but not *Myc*^{gfp}, was expressed by pDCs and by migratory and resident cDCs (Fig. 2a and Extended Data Fig. 3e, f). In the lung, liver and dermis, *Mycl1*^{gfp} was expressed predominantly by CD11b$^-$ cDCs, but, in the small intestine, *Mycl1*^{gfp} was expressed by CD11b$^+$ and CD11b$^-$ cDCs (Fig. 2b). *Mycl1*^{gfp} was more highly expressed in resident CD8α$^+$ and the migratory CD103$^+$ cDCs than in CD11b$^+$ cDC subsets (Extended Data Fig. 3f) and was absent in macrophages in the peritoneum, kidney and liver (Extended Data Fig. 3g). *Mycl1*^{gfp}-expressing DCs were abundant in the T-cell zones of spleen and lymph nodes, and less frequent within B-cell follicles and the splenic red pulp (Fig. 2c). Sparse *Mycl1*^{gfp}-expressing DCs were present in the sub-capsular sinus of inguinal lymph nodes where *Zbtb46*$^+$ cDCs reside[17] (Fig. 2c). In addition, *Mycl1*^{gfp} was expressed by CD4$^-$ B220$^-$ cells in small intestinal lamina propria, inside villi and within Peyer's patches (Fig. 2c). *Mycl1*^{gfp} expression was not expressed in vascular endothelium, unlike *Zbtb46*^{gfp} (Extended Data Fig. 4a, b). Thus, *Mycl*^{gfp} expression identifies DCs in both lymphoid and non-lymphoid peripheral tissues.

Mycl1 expression by CMPs has been reported to be IRF8-dependent[18]. In mice homozygous for the *Irf8*^{R294C} point mutation[19] that interrupts interactions between IRF8 and the transcription factor PU.1, *Mycl1*^{gfp} expression was absent in DC progenitors and substantially reduced in pDCs (Extended Data Fig. 5a, b). Moreover, *Mycl1*^{gfp} was expressed in DCs differentiated from wild-type, but not *Irf8*^{R294C}, Ly6C^{hi} monocytes using interleukin (IL)-4 and granulocyte–macrophage colony-stimulating factor (GM-CSF)[17,20] (Extended Data Fig. 5a, c). L-Myc protein was maintained in cDCs under various inflammatory conditions including treatment with interferon (IFN)-γ and increased by treatment with GM-CSF (Fig. 2d and Extended Data Fig. 5d). Chromatin immunoprecipitation combined with massively parallel sequencing (ChIP-seq) analysis identified several IRF8 binding regions across *Mycl1* that did not require the transcription factor BATF3 (ref. 21) (Fig. 2e and Extended Data Fig. 5e–g). Together, these results support a role for IRF8–PU.1 interactions in *Mycl1* expression.

In lymphoid and peripheral tissues, absence of L-Myc decreased the total number and relative frequency of DCs, which competitive mixed bone marrow chimaeras suggested was due to a cell-intrinsic defect (Fig. 3a–c and Extended Data Fig. 6a–e). The largest reduction was to

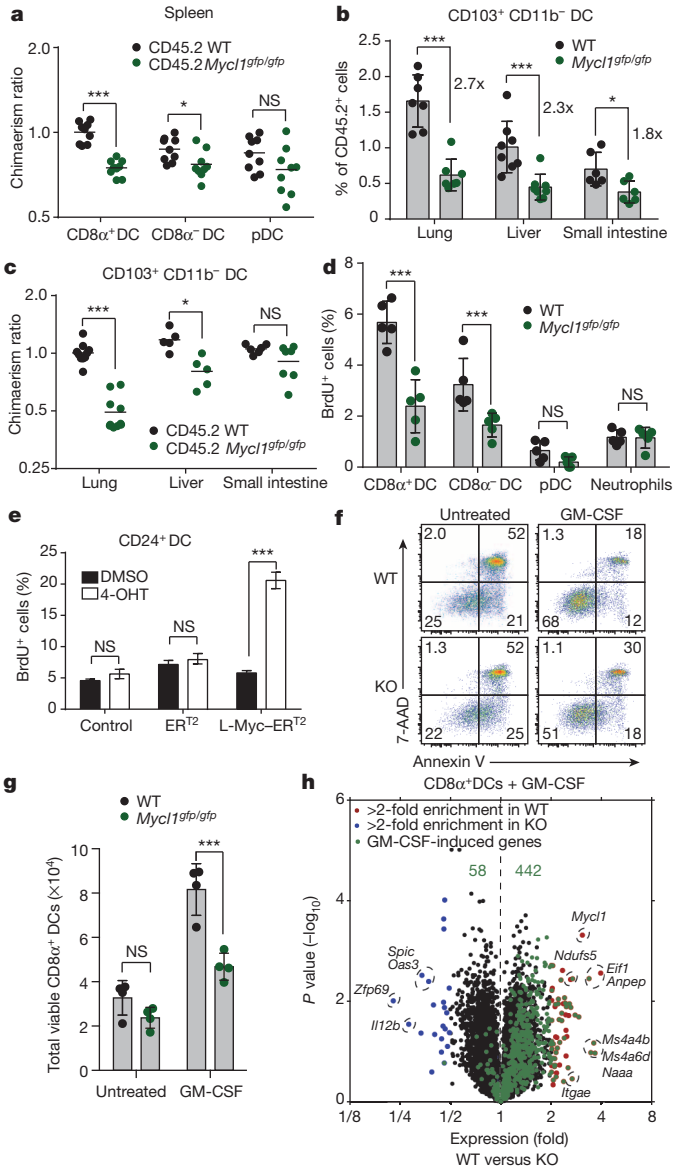


Figure 3 | *Mycl1* regulates DC proliferation and survival. **a**, Normalized donor contribution for indicated DC subsets in mixed bone marrow chimaeras (error bars, ± s.d., $n = 9$ mice, Mann–Whitney *U*-test). **b**, CD103$^+$ CD11b$^-$ DCs from tissues from the indicated mice as percentage of CD45.2$^+$ cells (error bars, ± s.d., $n = 5$–7 mice, two-way ANOVA Holm–Sidak post-hoc test). **c**, Donor contribution of CD103$^+$ CD11b$^-$ DCs as in **a** shown for indicated tissues. **d**, 1 h BrdU incorporation for indicated cells and mice (error bars, ± s.d., $n = 5$, Student's *t*-test). **e**, BrdU incorporation of CD24$^+$ cDCs expressing ER^{T2} or L-Myc–ER^{T2} fusion proteins in response to tamoxifen (4-OHT) treatment (error bars, ± s.d., $n = 4$, two-way ANOVA Holm–Sidak post-hoc test). **f**, 7-Aminoactinomycin D (7-AAD)/Annexin V staining of splenic CD8α$^+$ DCs from wild-type or *Mycl1*$^{gfp/gfp}$ (KO) mice treated with GM-CSF. **g**, Viable splenic CD8α$^+$ DCs from **f** for wild-type or *Mycl1*$^{gfp/gfp}$ mice (error bars, ± s.d, $n = 4$, Student's *t*-test). Data in **e**–**g** are representative of 2–3 experiments. **h**, Volcano plot of wild-type and *Mycl1*$^{gfp/gfp}$ CD8α$^+$ DCs treated with or without GM-CSF. Shown are genes increased >2-fold (red) or decreased >2-fold (blue) in wild-type relative to *Mycl1*$^{gfp/gfp}$ mice. The top 500 genes induced in wild-type cells following GM-CSF treatment are shown (green) ($n = 3$ biological replicates, Welch's *t*-test). *$P < 0.05$, ***$P < 0.001$, NS, $P > 0.05$.

the CD103$^+$ CD11b$^-$ cDCs in the lung, an organ rich in GM-CSF[22]. Gene set enrichment analysis (GSEA) also revealed significant enrichment of cell-cycle-related transcripts in lung CD103$^+$ CD11b$^-$ cDCs as compared to migratory CD103$^+$ CD11b$^-$ cDCs in draining lymph nodes (Extended Data Fig. 6f, g). Thus, variations in the abundance of cell-extrinsic factors such as GM-CSF in the local tissue microenvironment result in different homeostatic requirements for DCs that are revealed by the loss of L-Myc.

Analysis of gene expression microarrays from wild-type and L-Myc-deficient pDCs, CD8α$^+$ cDCs, and CD8α$^-$ cDCs identified a few presumed L-Myc target genes associated with cellular proliferation and apoptosis (Extended Data Fig. 7a, b). In agreement, splenic CD8α$^+$ cDCs from L-Myc-deficient mice showed a 50% reduction in BrdU incorporation *in vivo* relative to wild-type mice (Fig. 3d). Also, a tamoxifen-activated L-Myc fusion protein (L-Myc–ER^{T2}, in which ER^{T2} represents the mutant human oestrogen receptor) markedly and specifically increased proliferation of CD8α$^+$ cDCs (Fig. 3e), suggesting that L-Myc can regulate DC proliferation. GM-CSF, a cytokine known to regulate DC homeostasis, increased CD8α$^+$ cDC cell size and expression of >500 genes, including many involved in regulation of apoptosis (Extended Data Fig. 7c–f). Notably, the ability of GM-CSF to increase CD8α$^+$ cDC survival was impaired in the absence of L-Myc (Fig. 3f, g). Genes that appear to be targets of L-Myc in CD8α$^+$ cDCs include eukaryotic translation initiation factor 1 (*Eif1*) and NADH dehydrogenase (ubiquinone) Fe-S protein 5 (*Ndufs5*), which could affect global protein translation and energy metabolism. Furthermore, of the 500 genes induced by GM-CSF treatment in wild-type DCs (Extended Data Fig. 7e), 442 are reduced in expression in L-Myc-deficient CD8α$^+$ cDCs (Fig. 3h), suggesting that the absence of L-Myc broadly limits inducible genes expressed in activated CD8α$^+$ cDCs.

Finally, we assessed whether L-Myc expression was required for T-cell priming and other functions attributed to DCs[23,24]. We measured antigen-specific CD8$^+$ and CD4$^+$ T-cell responses after infection with *L. monocytogenes* expressing soluble ovalbumin (LM-OVA). Loss of L-Myc significantly decreased the total number of IFN-γ-producing OVA-specific CD8$^+$ and CD4$^+$ T cells (Fig. 4a, b and Extended Data Fig. 8a, b). To demonstrate that these effects were not the result of a requirement for L-Myc in T cells, we adoptively transferred congenically marked L-Myc-sufficient OVA-specific OT-I CD8$^+$ T cells into wild-type and L-Myc-deficient mice. After infection with LM-OVA, OT-I CD8$^+$ T-cell expansion was considerably reduced in L-Myc-deficient mice as compared to wild-type mice (Fig. 4c, d). L-Myc-deficient mice also showed impaired CD8$^+$ T-cell priming after infection with vesicular stomatitis virus expressing ovalbumin (VSV-OVA) (Extended Data Fig. 8c, d). These priming defects were attributable to the action of L-Myc in CD8α$^+$ cDCs, as depletion of pDCs[25] or Notch2-dependent CD11b$^+$ cDCs[26] had no effect on CD8$^+$ T-cell priming after infection with LM-OVA (Extended Data Fig. 9a–c). Further, the defect did not appear to involve processing and presentation of soluble antigen (Extended Data Fig. 10a, b). Mixed chimaera analysis using *Zbtb46*$^{dtr/dtr}$ mice[27] indicated that the priming defect is intrinsic to cDCs (Extended Data Fig. 9d–g).

We recently showed that CD8α$^+$ cDCs are required for *L. monocytogenes* to establish infection in mice via the intravenous route, as *Batf3*$^{-/-}$ mice lacking these cells are entirely resistant to lethal infection[28]. Further, for the first 24 h after infection, bacteria grow entirely within CD8α$^+$ cDCs[29], which are the initial reservoir for bacterial expansion. Because L-Myc is most highly expressed in the CD8α$^+$ cDCs in spleen, we asked whether L-Myc deficiency might influence infection by *L. monocytogenes*. L-Myc-deficient mice were remarkably resistant to lethal infection by *L. monocytogenes* relative to wild-type mice (Fig. 4e). This resistance was caused by significant reduction in the intracellular growth of bacteria within L-Myc-deficient CD8α$^+$ cDCs, and was not due to reduced bacterial capture or DC viability during the first 24 h of infection (Fig. 4f and Extended Data Fig. 8e–h). This reduced growth of *L. monocytogenes* was cell-intrinsic to CD8α$^+$ cDCs (Fig. 4g and Extended Data Fig. 10c, d) and was sufficient to prevent the subsequent spread of bacteria to other lineages (Fig. 4h and Extended Data Fig. 10e).

The functional relationship of L-Myc to other Myc factors has remained uncertain[4]. We show that L-Myc is selectively expressed in DCs, maintained

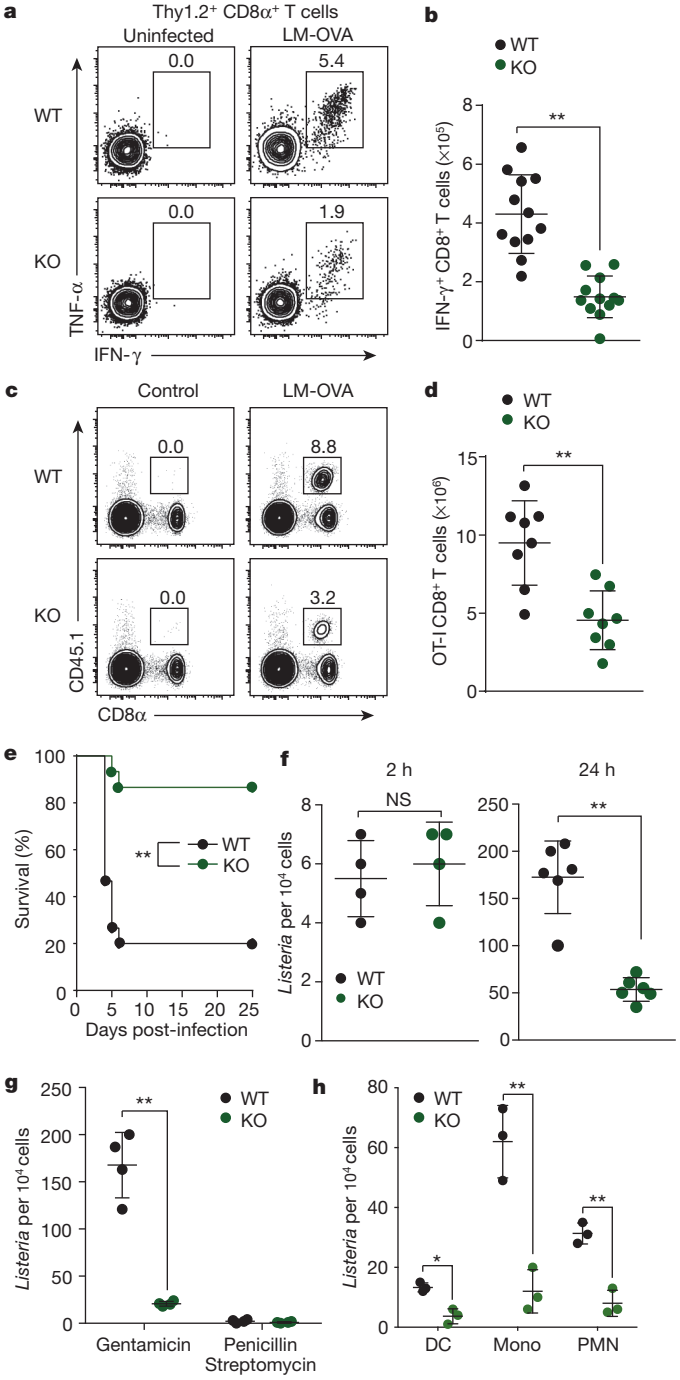


Figure 4 | *Mycl1* supports normal T-cell priming by DCs following infection but mediates resistance to lethal *L. monocytogenes* challenge. **a**, $CD8^+$ T cells from wild-type (WT) and $Mycl1^{gfp/gfp}$ (KO) mice infected with LM-OVA treated with SIINFEKL were analysed for TNF-α and IFN-γ production. **b**, Total IFN-γ^+ $CD8^+$ T cells were measured from the indicated mice after infection as described in **a** (error bars, ± s.d., $n = 12$, Student's *t*-test). Data in **a**, **b** are from 2 independent experiments. **c**, $CD45.1^+$ OT-I T cells transferred into the indicated mice and infected with LM-OVA were measured after 7 days. Numbers are OT-I T cells as a percentage of all splenocytes. **d**, Total OT-I $CD8^+$ T cells were measured from the indicated recipient mice after infection as described in **c** (error bars, ± s.d., $n = 8$, Student's *t*-test). Data in **c**, **d** are from 2 independent experiments. **e**, Survival of WT and KO mice after infection with *L. monocytogenes* ($n = 15$, log-rank Mantel–Cox test). Data are from 3 independent experiments. **f**, *L. monocytogenes* was measured in purified $CD8\alpha^+$ DCs after 2 h or 24 h of infection from **e** (error bars, ± s.d., $n = 4$–6 biological replicates, Student's *t*-test). **g**, Splenic $CD8\alpha^+$ DCs from mice infected for 2 h were cultured *in vitro* in media with the indicated antibiotic for 12 h and viable intracellular bacteria quantified (error bars, ± s.d., $n = 4$, Student's *t*-test). Data in **f**, **g** are from 2 independent experiments. **h**, Viable intracellular bacteria were measured as in **g** from the indicated cells from WT and KO mice infected with *L. monocytogenes* for 60 h (error bars, ± s.d., $n = 3$, Student's *t*-test). $^{**}P < 0.001$, NS, $P > 0.05$.

during inflammation, and required by cDCs for optimal priming of T cells in bacterial and viral infection. As c-Myc is repressed by interferons[30] and inducible genes can depend on c-Myc, L-Myc may provide a means to support transcriptional responses required during T-cell priming by cDCs.

METHODS SUMMARY

Mice. Wild-type 129S6/SvEv mice were from Taconic. Wild-type C57BL/6 mice, $Csf2rb^{-/-}$ mice and the congenic strain B6.SJL-$Ptprc^aPepc^b$/BoyJ (B6.SJL) were from The Jackson Laboratory. Mice were maintained in our specific pathogen-free animal facility according to institutional guidelines. Generation of $Myc^{gfp/gfp}$, $Zbtb46^{dtr/dtr}$ and $Irf8^{R294C}$ mice were described[16,19,26]. $Irf8^{R294C}$ mice were backcrossed to C57BL/6 for 11 generations. Experiments used sex- and age-matched mice at 6–16 weeks of age. All pathogen infections were performed on mice of the 129S6/SvEv genetic background unless indicated.

Myc, *Mycn* and *Mycl1* expression in Fig. 1a was determined by microarrays for the long-term haematopoietic stem cell, $Flt3^+$ multipotent progenitor, CMP, granulocyte–macrophage progenitor, common lymphoid progenitor, CDP, bone marrow pre-cDC, splenic pDC, splenic $CD8\alpha^+$ DC and splenic $CD8\alpha^-$ DC. In Fig. 1b, lineage markers included Ter119, NK1.1, B220, MHC-II, CD3 and CD11b. For Fig. 1c, CMPs were Lin^- $CD16/CD32^-$ $Flt3^+$ $cKit^+$, CDPs were Lin^- $CD16/CD32^-$ $Flt3^+$ $cKit^{int/-}$ $CD115^+$ and pre-cDCs were Lin^- $CD16/CD32^-$ $Flt3^+$ $cKit^-$ $CD11c^+$. For Fig. 1d–f, gating is as follows: DCs, $CD11c^+$ MHC-II^+; neutrophils, $Ly6G^+$ $CD11b^+$; monocytes, $Ly6C^+$ $CD11b^+$ $Ly6G^-$; red pulp macrophages, autofluorescent $F4/80^+$; NK cells, $NK1.1^+$ $CD3\varepsilon^-$; CD8 T cells, $CD3\varepsilon^+$ $CD8\alpha^+$ $CD4^-$; CD4 T cells, $CD3\varepsilon^+$ $CD4^+$ $CD8\alpha^-$; B cells, $CD19^+$ $B220^+$ $SiglecH^-$. In Fig. 2a, pDCs were gated as $CD11c^{int}$ MHC-$II^{int/-}$ cells; resident DCs as $CD11c^+$ MHC-II^{int} cells; and migratory DCs as $CD11c^{int}$ MHC-II^+ cells. In Fig. 2d, bone marrow cells were cultured for 9 days with Flt3L and treated with media, IL-4, IFN-γ, GM-CSF or heat-killed *L. monocytogenes* strain EGD for 24 h.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 November 2012; accepted 19 December 2013.
Published online 9 February 2014.

1. Dang, C. V. *MYC* on the path to cancer. *Cell* **149,** 22–35 (2012).
2. Charron, J. *et al.* Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. *Genes Dev.* **6,** 2248–2257 (1992).
3. Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R. & Bradley, A. A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev.* **7,** 671–682 (1993).
4. Hatton, K. S. *et al.* Expression and activity of L-Myc in normal mouse development. *Mol. Cell. Biol.* **16,** 1794–1804 (1996).
5. Liu, K. *et al. In vivo* analysis of dendritic cell development and homeostasis. *Science* **324,** 392–397 (2009).
6. Lauvau, G. *et al.* Priming of memory but not effector CD8 T cells by a killed bacterial vaccine. *Science* **294,** 1735–1739 (2001).
7. Wang, R. *et al.* The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35,** 871–882 (2011).
8. Laurenti, E. *et al.* Hematopoietic stem cell function and survival depend on c-Myc and N-Myc activity. *Cell Stem Cell* **3,** 611–624 (2008).
9. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141,** 432–445 (2010).
10. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151,** 56–67 (2012).
11. Nie, Z. *et al.* c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* **151,** 68–79 (2012).
12. Wasylishen, A. R. *et al.* New model systems provide insights into Myc-induced transformation. *Oncogene* **30,** 3727–3734 (2011).
13. Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T. & Yamanaka, S. Promotion of direct reprogramming by transformation-deficient Myc. *Proc. Natl Acad. Sci. USA* **107,** 14152–14157 (2010).
14. Kabashima, K. *et al.* Intrinsic lymphotoxin-β receptor requirement for homeostasis of lymphoid tissue dendritic cells. *Immunity* **22,** 439–450 (2005).
15. Liu, K. *et al.* Origin of dendritic cells in peripheral lymphoid organs of mice. *Nature Immunol.* **8,** 578–583 (2007).
16. Huang, C. Y., Bredemeyer, A. L., Walker, L. M., Bassing, C. H. & Sleckman, B. P. Dynamic regulation of *c-Myc* proto-oncogene expression during lymphocyte

development revealed by a *GFP-c-Myc* knock-in mouse. *Eur. J. Immunol.* **38,** 342–349 (2008).
17. Satpathy, A. T. *et al.* Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J. Exp. Med.* **209,** 1135–1152 (2012).
18. Becker, A. M. *et al.* IRF-8 extinguishes neutrophil production and promotes dendritic cell lineage commitment in both myeloid and lymphoid mouse progenitors. *Blood* **119,** 2003–2012 (2012).
19. Tailor, P., Tamura, T., Morse, H. C. & Ozato, K. The BXH2 mutation in IRF8 differentially impairs dendritic cell subset development in the mouse. *Blood* **111,** 1942–1945 (2008).
20. Inaba, K. *et al.* Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **176,** 1693–1702 (1992).
21. Tussiwand, R. *et al.* Compensatory dendritic cell development mediated by BATF–IRF interactions. *Nature* **490,** 502–507 (2012).
22. Scott, C. L. *et al.* Functional analysis of mature hematopoietic cells from mice lacking the βc chain of the granulocyte-macrophage colony-stimulating factor receptor. *Blood* **92,** 4119–4127 (1998).
23. Shortman, K. & Heath, W. R. The $CD8^+$ dendritic cell subset. *Immunol. Rev.* **234,** 18–31 (2010).
24. Hildner, K. *et al. Batf3* deficiency reveals a critical role for $CD8\alpha^+$ dendritic cells in cytotoxic T cell immunity. *Science* **322,** 1097–1100 (2008).
25. Swiecki, M., Gilfillan, S., Vermi, W., Wang, Y. & Colonna, M. Plasmacytoid dendritic cell ablation impacts early interferon responses and antiviral NK and $CD8^+$ T cell accrual. *Immunity* **33,** 955–966 (2010).
26. Satpathy, A. T. *et al.* Notch2-dependent classical dendritic cells orchestrate intestinal immunity to attaching-and-effacing bacterial pathogens. *Nature Immunol.* **14,** 937–948 (2013).
27. Meredith, M. M. *et al.* Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage. *J. Exp. Med.* **209,** 1153–1165 (2012).
28. Edelson, B. T. *et al.* $CD8\alpha^+$ dendritic cells are an obligate cellular entry point for productive infection by *Listeria monocytogenes*. *Immunity* **35,** 236–248 (2011).
29. Neuenhahn, M. *et al.* $CD8\alpha^+$ dendritic cells are required for efficient entry of *Listeria monocytogenes* into the spleen. *Immunity* **25,** 619–630 (2006).
30. Resnitzky, D. & Kimchi, A. Deregulated *c-myc* expression abrogates the interferon- and interleukin 6-mediated G_0/G_1 cell cycle arrest but not other inhibitory responses in M1 myeloblastic cells. *Cell Growth Differ.* **2,** 33–41 (1991).

Acknowledgements This work was supported by the Howard Hughes Medical Institute and the Siteman Cancer Center (K.M.M.), American Heart Association (12PRE8610005 to A.T.S. and 12PRE12050419 to W.K.), German Research Foundation (AL 1038/1-1; J.C.A.), American Society of Hematology Scholar Award and Burroughs Wellcome Fund Career Award for Medical Scientists (B.T.E.). We thank J. M. White for blastocyst injections and generation of mouse chimaeras, the Alvin J. Siteman Cancer Center for use of the Center for Biomedical Informatics and the Multiplexed Gene Analysis Core Facility. The Siteman Cancer Center is supported in part by the NCI Cancer Center Support Grant P30 CA91842.

Author Contributions W.K., T.L.M. and K.M.M. directed the work and wrote the manuscript. S.P.P. and P.M.A. helped with *L. monocytogenes* infections. A.S.R., M. Cella and M. Colonna helped with VSV infections. B.P.S. provided mice expressing GFP–c-Myc. J.L. and M.C.N. provided *Zbtb46*$^{dtr/dtr}$ mice. A.T.S. and J.C.A. aided gene targeting and analysis of DC populations. C.G.B., X.W., B.T.E., V.D., N.M.K. and E.V.R.-G. aided analysis of infections and histological analysis. X.W. aided microarray analysis. All authors discussed the results and contributed to the manuscript.

Author Information Aligned and raw data can be found online under the Gene Expression Omnibus (GEO) series accession GSE53311. All original microarray data have been deposited in GEO series accession GSE53312. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M.M. (kmurphy@wustl.edu).

LETTER

doi:10.1038/nature12920

A transcriptional switch underlies commitment to sexual development in malaria parasites

Björn F. C. Kafsack[1]†, Núria Rovira-Graells[2,3], Taane G. Clark[4,5], Cristina Bancells[2], Valerie M. Crowley[1,3]†, Susana G. Campino[6], April E. Williams[7], Laura G. Drought[4], Dominic P. Kwiatkowski[6,8], David A. Baker[4], Alfred Cortés[2,3,9] & Manuel Llinás[1,7]†

The life cycles of many parasites involve transitions between disparate host species, requiring these parasites to go through multiple developmental stages adapted to each of these specialized niches. Transmission of malaria parasites (*Plasmodium* spp.) from humans to the mosquito vector requires differentiation from asexual stages replicating within red blood cells into non-dividing male and female gametocytes. Although gametocytes were first described in 1880, our understanding of the molecular mechanisms involved in commitment to gametocyte formation is extremely limited, and disrupting this critical developmental transition remains a long-standing goal[1]. Here we show that expression levels of the DNA-binding protein PfAP2-G correlate strongly with levels of gametocyte formation. Using independent forward and reverse genetics approaches, we demonstrate that PfAP2-G function is essential for parasite sexual differentiation. By combining genome-wide PfAP2-G cognate motif occurrence with global transcriptional changes resulting from PfAP2-G ablation, we identify early gametocyte genes as probable targets of PfAP2-G and show that their regulation by PfAP2-G is critical for their wild-type level expression. In the asexual blood-stage parasites *pfap2-g* appears to be among a set of epigenetically silenced loci[2,3] prone to spontaneous activation[4]. Stochastic activation presents a simple mechanism for a low baseline of gametocyte production. Overall, these findings identify PfAP2-G as a master regulator of sexual-stage development in malaria parasites and mark the first discovery of a transcriptional switch controlling a differentiation decision in protozoan parasites.

From its uptake in a mosquito blood meal to initial infection of red blood cells in the subsequent host, the malaria parasite *Plasmodium falciparum* goes through at least seven key developmental changes (asexual red cell stage → gametocyte → gamete → ookinete → oocyst → sporozoite → liver stage → asexual red cell stage). In all but one case, as the parasite reaches its subsequent niche within the host, differentiation into the appropriate developmental stage is a necessity for continuation of the life cycle. The lone exception occurs once the parasite has started replicating in red blood cells. During the 48-h intraerythrocytic developmental cycle following each new red blood cell invasion, a developmental decision is made that determines whether daughter parasites will continue replicating asexually and maintain the infection of the current host or differentiate into non-dividing male or female gametocytes. Although the latter decision is a dead-end for replication within the current host it is essential for infection of mosquitoes and thus transmission to the next host[5,6].

A recent study on transcriptional variation identified differentially expressed genes linked to early gametocyte development in two stocks (3D7-A and 3D7-B) of the common 3D7 *P. falciparum* parasite line[7]. Within this expression cluster of early gametocyte markers, we noted the presence of a potential transcriptional regulator, PfAP2-G (PFL1085w/PF3D7_1222600; http://www.plasmodb.org), which belongs to the apicomplexan AP2 (ApiAP2) family of DNA-binding proteins (Supplementary Fig. 1) and is conserved among most members of the phylum (Supplementary Fig. 2). ApiAP2 proteins represent the main family of transcriptional regulators in malaria parasites[8] and have thus far been found to regulate several of the parasite's developmental transitions, including ookinete formation[9] and oocyst sporozoite maturation[10] within the mosquito, and development in the mammalian liver[11]. Follow-up quantitative PCR with reverse transcription (qRT–PCR) analysis in blood-stage parasites confirmed higher *pfap2-g* transcript abundance in 3D7-B compared to 3D7-A and also revealed significant variation in expression levels between individual 3D7-B subclones (Fig. 1a). Notably, when gametocyte formation was measured in these lines, *pfap2-g* transcript levels were highly predictive ($R^2 > 0.99$) of relative gametocyte production (Fig. 1b).

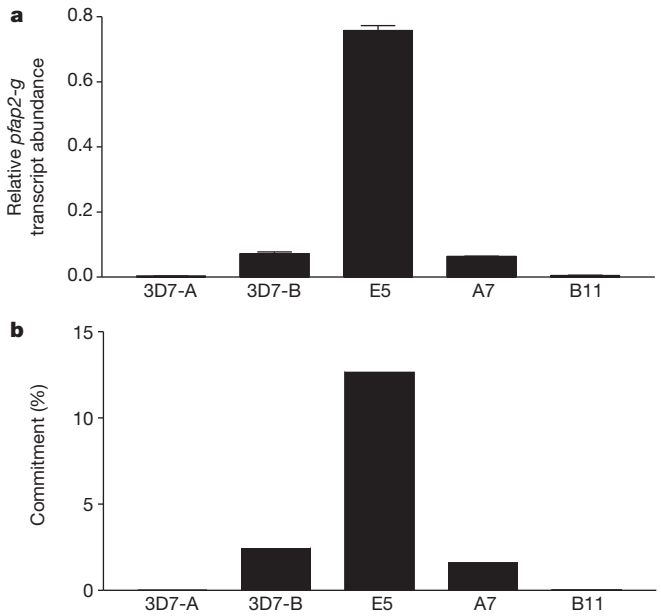


Figure 1 | *pfap2-g* transcript levels mirror gametocyte production. a, *pfap2-g* relative transcript abundance in synchronized (early schizont stage) cultures as measured by qPCR varies significantly between 3D7-A and 3D7-B populations as well as the 3D7-B subclones E5, A7 and B11. Values are normalized against seryl transfer RNA synthetase (PF07_0073) ($n = 3$, standard deviation shown). **b**, Per cent commitment to gametocyte differentiation in these lines mirrors relative *pfap2-g* transcript levels (mean of $n = 2$).

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. [2]Barcelona Centre for International Health Research (CRESIB, Hospital Clínic-Universitat de Barcelona), Barcelona, 08036 Catalonia, Spain. [3]Institute for Research in Biomedicine (IRB), Barcelona, 08028 Catalonia, Spain. [4]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. [5]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. [6]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. [7]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [8]Wellcome Trust Sanger Centre for Human Genetics, Oxford OX3 7BN, UK. [9]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, 08010 Catalonia, Spain. †Present addresses: Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA (B.F.C.K.); Department of Molecular Biology and Center for Infectious Disease Dynamics, The Pennsylvania State University, State College, Pennsylvania 16802, USA (V.M.C., M.L.).

In a parallel line of inquiry, we screened the well-studied gametocyte non-producer line F12 (refs 5, 6, 12), as well as a second parasite line (GNP-A4) that had also spontaneously lost its ability to produce gametocytes, for mutations in protein-coding regions. Whole genome sequencing of these lines revealed that the only gene containing mutations in both F12 and GNP-A4 was *pfap2-g* (Supplementary Table 1), resulting in the introduction of stop codons upstream of or within the AP2 DNA-binding domain (Fig. 2a and Supplementary Fig. 3). Previous studies identified subtelomeric deletions in the right arm of *P. falciparum* chromosome 9 that are associated with defective gametocyte production[13,14]. The F12 and GNP-A4 clones do not have coding-sequence mutations or deletions within the chromosome 9 region, nor within any of the 16 genes recently implicated in gametocyte development by random transposon mutagenesis[15]. The presence of *pfap2-g* mutations in two independently derived gametocyte non-producer lines provides a second, independent connection between PfAP2-G and gametocyte formation, pointing to this locus as a key determinant of sexual differentiation. Although spontaneous inactivation of *pfap2-g* has occurred repeatedly *in vitro*, no loss-of-function mutations could be found in the genomes of nearly 300 distinct field isolates[16], further underlining its potential importance to transmission.

To directly test the contribution of PfAP2-G function to gametocyte formation, we generated a PfAP2-G null mutant (Δ*pfap2-g*) via double homologous recombination in the high-gametocyte-producing 3D7-B subclone E5 (Fig. 2a, b and Supplementary Fig. 4). As predicted based on our earlier sequencing results, the Δ*pfap2-g* mutant completely lost the ability to produce gametocytes (Fig. 2c). To identify any additional mutations that may have been acquired in the extended process of generating the *pfap2-g* knockout, we sequenced the genomes of both Δ*pfap2-g* and its E5 parent. Apart from the targeted deletion, we found only a limited number of additional mutations within coding regions, none of which are shared with the other non-producer lines that we sequenced or found in genes previously linked to gametocyte development[13–15] (Supplementary Table 1). This combination of forward and reverse genetic evidence strongly implies the essentiality of PfAP2-G for the production of gametocytes in *P. falciparum*. In direct competition cultures Δ*pfap2-g* consistently outgrew its parent E5, consistent with the fact that PfAP2-G action occurs at or before the asexual/sexual decision but not thereafter, as only a failure to initiate gametocytogenesis would provide an *in vitro* growth advantage (Supplementary Fig. 5).

Attempts at generating full-length complementation expression constructs were unsuccessful, probably owing to the considerable length (7.3 kilobases (kb)) of the coding sequence and its very low complexity (21.8% GC and long repeat sequences). As an alternative confirmation for the role of PfAP2-G in gametocyte formation, we made PfAP2-G function ligand-regulatable by appending the FKBP-derived destabilization domain (ddFKBP) to the 3′ end of the endogenous coding sequence (*pfap2-g-ddfkbp*, Supplementary Fig. 6a, b). In the absence of the synthetic ligand Shield-1 (Shld1) the ddFKBP domain is unstable and targets fusion proteins for proteolytic degradation[17,18], thus making PfAP2-G protein levels regulatable by the addition of Shld1 (Supplementary Fig. 6c). Indeed, in the *pfap2-g-ddfkbp* line gametocyte formation was completely dependent on the addition of Shld1, whereas its presence had no effect on gametocyte production by the E5 parent (Fig. 2d, e), demonstrating that PfAP2-G function is essential for gametocyte formation.

On the basis of the localization of haemagglutinin (HA)-tagged PfAP2-G to the parasite nucleus (Fig. 3a and Supplementary Fig. 7) and the fact that several ApiAP2 proteins act as transcriptional regulators, we aimed to identify possible regulatory targets of PfAP2-G. To do this, we compared the global transcriptional pattern over the 48-h intraerythrocytic cycle for the gametocyte-producing parent E5 to those of the mutant non-producers Δ*pfap2-g* and F12. As expected, only a small number of transcripts changed by greater than twofold in both mutants; with four transcripts increasing and 23 transcripts decreasing in abundance (Fig. 3b and Supplementary Table 2). All four upregulated genes are located in subtelomeric regions and have previously been shown to undergo spontaneous transcriptional variation and were therefore not considered further[7]. However, the cluster of downregulated genes is highly enriched for genes expressed during the first stages of gametocyte formation ($P < 0.003$), including some of the earliest known markers of sexual commitment: *pfs16, pfg27/25* and *pfg14.744* (refs 19, 20).

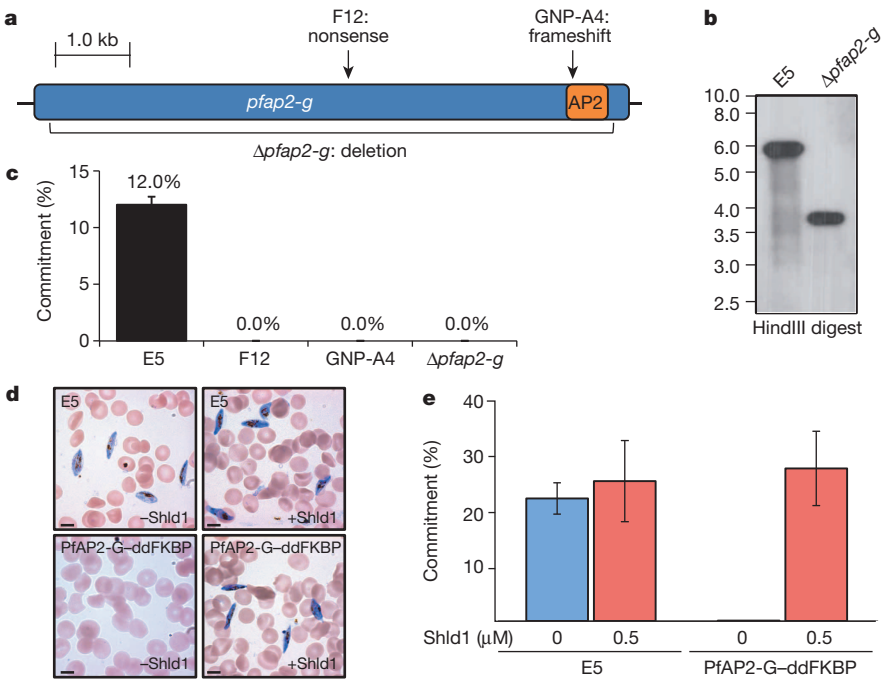


Figure 2 | Disrupting PfAP2-G function results in loss of gametocyte production. **a**, Positions of *pfap2-g* mutations in the gametocyte non-producer lines F12 and GNP-A4 and the targeted deletion of Δ*pfap2-g*. **b**, Southern blot showing successful disruption of the *pfap2-g* locus by homologous recombination (also see Supplementary Fig. 4). Single replicate. **c**, *pfap2-g* mutants fail to produce gametocytes ($n = 3$, standard error shown). **d**, Ligand-regulatable gametocyte formation in PfAP2-G–ddFKBP (bottom row images) but not in the E5 parent (top row images). Representative of $n = 4$. Scale bars, 5 μm. **e**, Quantification of ligand-regulatable gametocyte formation ($n = 4$, standard error shown).

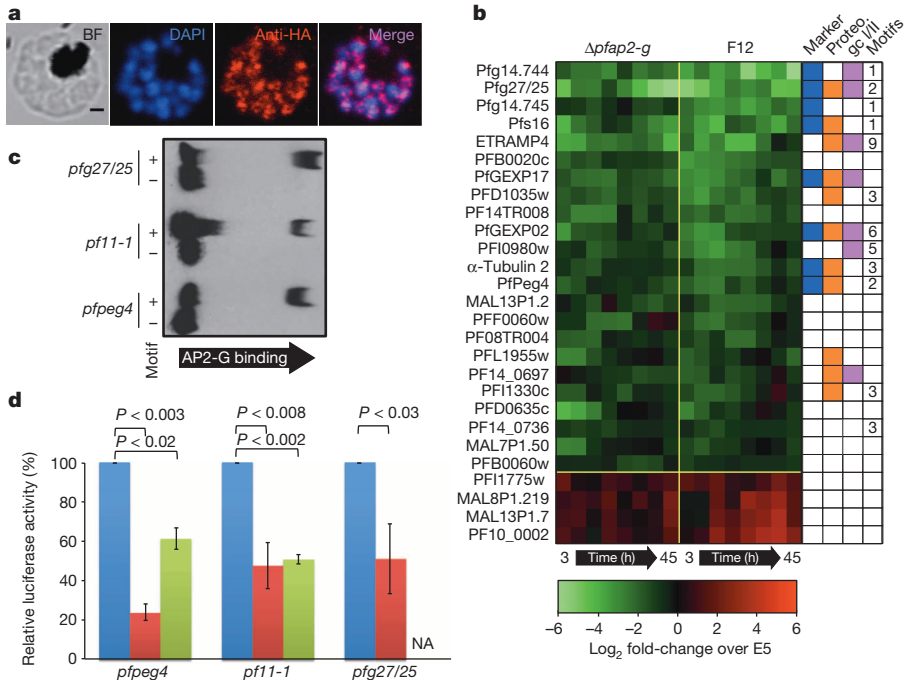


Figure 3 | Identification of PfAP2-G targets. **a**, PfAP2-G–HA×3 localizes to the nuclei of schizonts in asexually growing parasites (see Supplementary Fig. 7 for additional stages). Scale bar, 1 μm. Representative of $n = 8$. BF, bright field; DAPI, 4′,6-diamidino-2-phenylindole. **b**, Relative abundance of transcripts with greater than twofold average difference in both Δ*pfap2-g* and F12 with respect to 3D7-B clone E5 across the intra-erythrocytic developmental cycle at 6-h intervals. Columns on the right indicate whether genes are known gametocyte markers (blue), detected in two or more gametocyte proteomes (orange), enriched in early gametocyte proteome (purple), and the number of PfAP2-G cognate motifs within 2 kb upstream of the start codon. **c**, Binding of the recombinant PfAP2-G AP2 domain to three gametocyte promoters occurs only in the presence of the wild-type cognate motif (+). Representative of $n = 3$. **d**, Relative luciferase activity under the control of wild-type gametocyte promoters in 3D7-B E5 (blue) and Δ*pfap2-g* (red), or in 3D7-B E5 under control of promoters lacking the PfAP2-G motif (green). (18–30 h post-invasion, $n = 3$, standard error is shown, two-sided *t*-test used). NA, not tested.

qRT–PCR measurements of early gametocyte markers confirmed the lower relative abundance levels in Δ*pfap2-g* (Supplementary Fig. 8). Analysis of the upstream regions of most downregulated genes showed that they were also enriched in the DNA motif recognized by PfAP2-G[21] ($P < 0.017$). These results implicate PfAP2-G as a transcriptional switch that controls sexual differentiation by activating the transcription of early gametocyte genes.

Using electrophoretic mobility shift assays we confirmed that the recombinant PfAP2-G DNA-binding domain could interact with three gametocyte promoters in a motif-dependent manner *in vitro* (Fig. 3c). To test whether this interaction occurs within the parasite, we transfected E5 and Δ*pfap2-g* with luciferase reporter constructs under the control of these gametocyte promoters (Fig. 3d). There was a significant reduction in luciferase activity in the Δ*pfap2-g* background compared to its E5 parent for all three constructs. In addition, luciferase levels were also significantly diminished in the parental E5 line when we altered the PfAP2-G recognition sequence in the two promoters tested, indicating that PfAP2-G probably acts as a direct transcriptional activator of the earliest gametocyte genes.

The *pfap2-g* locus shares many features that have been associated with the epigenetic silencing of multigene families in *P. falciparum*[4,22] such as high levels of the H3K9me3 histone modification[3], associated binding of heterochromatin protein 1 (PfHP1)[2,23] and perinuclear localization[3]. On the basis of these data PfAP2-G expression is probably regulated epigenetically by reversible formation of repressive chromatin structures. Interestingly, we find that the pattern of histone modifications at this locus is typical of heterochromatin-silenced genes in both the high gametocyte producer E5 and its low-producing A7 sibling clone (Supplementary Fig. 9a–c). This finding suggests that, in predominantly asexual blood-stage cultures, the *pfap2-g* locus is found in a heterochromatic (silenced) state in the majority of parasites and that the transcriptionally permissive state may only occur in a small number of sexually committed parasites. Indeed, the vast majority of asexually growing PfAP2-G–HA×3 (in which HA×3 denotes a triple HA tag) parasites contained no detectable levels of PfAP2-G by immunofluorescence, whereas a small subpopulation exhibited clear nuclear PfAP2-G staining (Fig. 4a). Every newly formed merozoite within PfAP2-G-expressing schizonts stained positive for PfAP2-G, lending further support to the previous findings that all daughter parasites from a given schizont are commited to the same developmental fate[24]. Furthermore, although the PfAP2-G-positive fraction varied between experiments, it was highly predictive of subsequent gametocyte formation in commitment assays ($R^2 = 0.94$, Fig. 4b).

Stochastic, low-frequency activation would provide a simple mechanism for baseline gametocyte production, which may be modulated in response to environmental stimuli. Furthermore, the presence of insulator-like pairing element sequences—which have been suggested to have an important role in the silencing of *var* genes[25]—flanking the *pfap2-g* locus (Supplementary Fig. 9d) raises the intriguing possibility that the expression of *pfap2-g* may be mutually exclusive with that of the *var* gene family[22]. In addition to chromatin-mediated control, PfAP2-G expression may be autoregulated via binding to the eight instances of the PfAP2-G cognate motifs located 2.1–3.6 kb upstream of the PfAP2-G locus (Supplementary Fig. 10). We have integrated these various regulatory mechanisms into a model of how PfAP2-G expression controls the decision of individual cells to commit to gametocyte formation or to continue along the default pathway of asexual replication (Fig. 4c).

Together with the work of Sinha *et al.*[26] (accompanying manuscript), our results demonstrate that AP2-G is an essential regulator of gametocyte formation in malaria parasites and acts as a developmental switch by activating the transcription of early gametocyte genes. This provides the first insight into the molecular mechanisms controlling the asexual/sexual developmental decision in malaria parasites and unveils new targets in the long-standing aim of interrupting malaria transmission by preventing the formation and/or maturation of the parasite's sexual stages[1]. Last, ligand-regulatable PfAP2-G is not only a powerful new

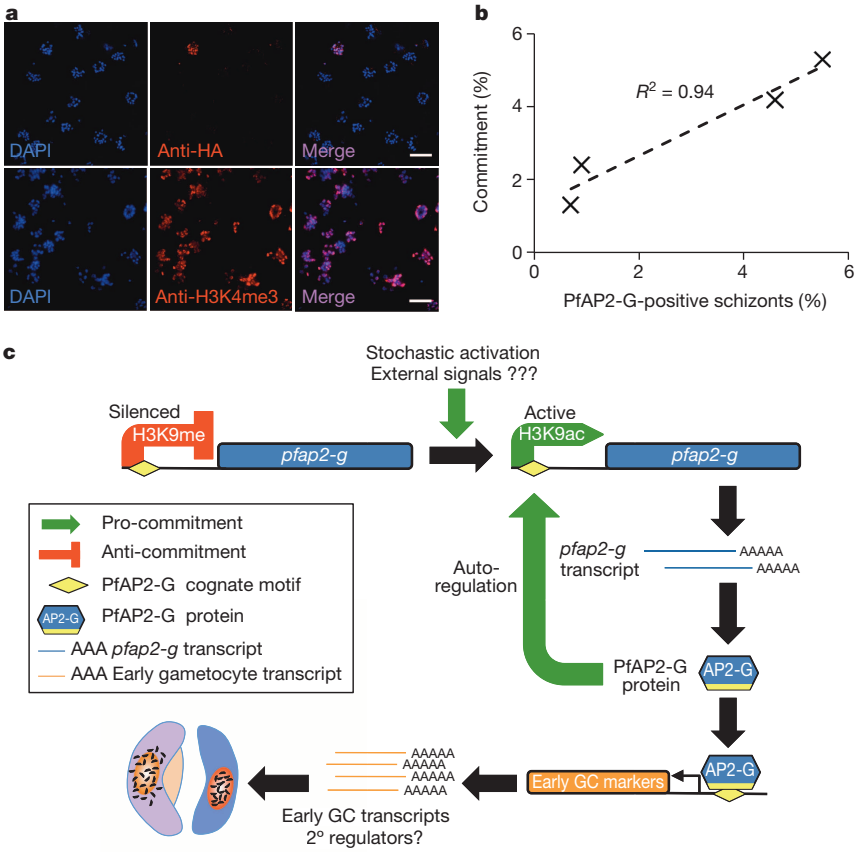


Figure 4 | Activation of PfAP2-G. a, Only a small fraction (1–6%) of asexually growing subclone 9A schizonts (see Supplementary Fig. 7 for details) express detectable levels of PfAP2-G–HA×3 (top row). H3K4me3 staining was performed in parallel to confirm full permeabilization (bottom row). Scale bars, 10 μm. Representative of $n = 4$. **b**, The percentage of PfAP2-G–HA×3-positive cells is highly predictive ($R^2 = 0.94$) of subsequent gametocyte formation levels. **c**, Model of PfAP2-G activation and function.

tool for future studies of sexual-stage development in malaria parasites but also holds great potential for inducible gene expression in general.

METHODS SUMMARY

Parasites and strains. Δ*pfap2-g* knockout parasites were generated by transfection of 3D7-B E5 with pHHT-FCU-pfap2-g (Supplementary Fig. 4) followed by positive (*hdhfr*)/negative (*fcu*) selection. *pfap2-g-ddfkbp* parasites were generated by transfection of 3D7-B E5 with pJDD145-pfap2-g (Supplementary Fig. 6). Parasites expressing PfAP2-G-HAx3 were generated by transfection of 3D7-B E5 with pHH1inv-pfap2-g-HAx3 (Supplementary Fig. 7). All parasites were grown in media containing AlbuMAX II and synchronized by standard methods.

Gametocytogenesis. Gametocyte induction was performed according to published methods[27]. For ligand-regulatable gametocytogenesis (Fig. 2d, e), synchronized parasites were set up at 0.5–1.0% late trophozoites in 3% hematocrit on day 0. Cultures were split in two and treated with 0.5 μM Shld1 or solvent control for the remainder of the experiment.

Gel shifts. Electrophoretic mobility shift assays were performed using Light Shift EMSA kits (Thermo Scientific) using 2 μg of protein and 20 fmol of probe.

Microarrays. Starting at 3 h post-invasion, tightly synchronized parasites were collected at eight time points with 6-h intervals. RNA isolation, complementary DNA generation/labelling, array hybridization, and feature extraction was performed as described previously[28]. Cy5-labelled cDNA was hybridized with a common Cy3-labelled reference pool on the *P. falciparum* 8×15K Agilent nuclear expression array (Gene Expression Omnibus (GEO) platform accession GPL17880). Genes were rank ordered by their average relative transcript abundance differences across the eight time points between the wild type (E5) and mutant (F12 or Δ*pfap2-g*).

Luciferase assays. Equal numbers of synchronized, stably transfected parasites were isolated and saponin-lysed (0.05% in PBS) at ~18–30 h post-invasion and assayed using Bright-Glo Luciferase Assay System (Promega).

Next-generation sequencing and analysis. Next-generation sequencing of the 3D7-B subclone E5 and Δ*pfap2-g* was performed using Illumina TruSeq single-end sequencing runs, analysed and visualized as described previously[29]. Genomic DNA for 3D7A, F12 and GNP-A4 was also used for whole genome sequencing at the Sanger Institute using Illumina GA II technology with 76-base paired-end reads. The raw sequence data were processed as described previously[30]. Experimental confirmation of informative genomic variants was performed using capillary sequencing methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April; accepted 27 November 2013.
Published online 23 February 2014.

1. Wells, T. N. C., Alonso, P. L. & Gutteridge, W. E. New medicines to improve control and contribute to the eradication of malaria. *Nature Rev. Drug Discov.* **8,** 879–891 (2009).
2. Flueck, C. *et al. Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog.* **5,** e1000569 (2009).
3. Lopez-Rubio, J. J., Mancio-Silva, L. & Scherf, A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5,** 179–190 (2009).
4. Cortés, A., Crowley, V. M., Vaquero, A. & Voss, T. S. A view on the role of epigenetics in the biology of malaria parasites. *PLoS Pathog.* **8,** e1002943 (2012).
5. Alano, P. *Plasmodium falciparum* gametocytes: still many secrets of a hidden life. *Mol. Microbiol.* **66,** 291–302 (2007).
6. Dixon, M. W. A., Thompson, J., Gardiner, D. L. & Trenholme, K. R. Sex in *Plasmodium*: a sign of commitment. *Trends Parasitol.* **24,** 168–175 (2008).
7. Rovira-Graells, N. *et al.* Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res.* **22,** 925–938 (2012).
8. Painter, H. J., Campbell, T. L. & Llinás, M. The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. *Mol. Biochem. Parasitol.* **176,** 1–7 (2011).
9. Yuda, M. *et al.* Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol. Microbiol.* **71,** 1402–1414 (2009).
10. Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T. & Kaneko, I. Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol. Microbiol.* **75,** 854–863 (2010).

11. Iwanaga, S., Kaneko, I., Kato, T. & Yuda, M. Identification of an AP2-family protein that is critical for malaria liver stage development. *PLoS ONE* **7,** e47557 (2012).
12. Alano, P. *et al. Plasmodium falciparum*: parasites defective in early stages of gametocytogenesis. *Exp. Parasitol.* **81,** 227–235 (1995).
13. Day, K. P. *et al.* Genes necessary for expression of a virulence determinant and for transmission of *Plasmodium falciparum* are located on a 0.3-megabase region of chromosome 9. *Proc. Natl Acad. Sci. USA* **90,** 8292–8296 (1993).
14. Eksi, S. *et al. Plasmodium falciparum* gametocyte development 1 (*Pfgdv1*) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS Pathog.* **8,** e1002964 (2012).
15. Ikadai, H. *et al.* Transposon mutagenesis identifies genes essential for *Plasmodium falciparum* gametocytogenesis. *Proc. Natl Acad. Sci. USA* **110,** E1676–E1684 (2013).
16. Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487,** 375–379 (2012).
17. Armstrong, C. M. & Goldberg, D. E. An FKBP destabilization domain modulates protein levels in *Plasmodium falciparum*. *Nature Methods* **4,** 1007–1009 (2007).
18. Banaszynski, L. A., Chen, L.-C., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126,** 995–1004 (2006).
19. Pradel, G. Proteins of the malaria parasite sexual stages: expression, function and potential for transmission blocking strategies. *Parasitology* **134,** 1911–1929 (2007).
20. Silvestrini, F. *et al.* Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol. Cell. Proteomics* **9,** 1437–1448 (2010).
21. Campbell, T. L., De Silva, E. K., Olszewski, K. L., Elemento, O. & Llinás, M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* **6,** e1001165 (2010).
22. Guizetti, J. & Scherf, A. Silence, activate, poise, and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. *Cell. Microbiol.* **15,** 718–726 (2013).
23. Pérez-Toledo, K. *et al. Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of *var* genes. *Nucleic Acids Res.* **37,** 2596–2606 (2009).
24. Bruce, M. C., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology* **100,** 191–200 (1990).
25. Avraham, I., Schreier, J. & Dzikowski, R. Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **109,** 52 (2012).
26. Sinha, A. *et al.* A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* http://dx.doi.org/10.1038/nature12970 (this issue).
27. Fivelman, Q. L. *et al.* Improved synchronous production of *Plasmodium falciparum* gametocytes *in vitro*. *Mol. Biochem. Parasitol.* **154,** 119–123 (2007).
28. Kafsack, B. F. C., Painter, H. J. & Llinás, M. New Agilent platform DNA microarrays for transcriptome analysis of *Plasmodium falciparum* and *Plasmodium berghei* for the malaria research community. *Malar. J.* **11,** 187 (2012).
29. Straimer, J. *et al.* Site-specific genome editing in *Plasmodium falciparum* using engineered zinc-finger nucleases. *Nature Methods* **9,** 993–998 (2012).
30. Robinson, T. *et al.* Drug-resistant genotypes and multi-clonality in *Plasmodium falciparum* analysed by direct genome sequencing from peripheral blood of malaria patients. *PLoS ONE* **6,** e23204 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank C. Klein, T. Campbell and A. Schieler for technical assistance and are grateful to O. Billker, C. Flueck, J. Kelly, C. Sutherland, A. Vaidya and A. Waters for discussion and reading of the manuscript. We would also like to thank P. Alano for providing *P. falciparum* clone F12, C. Taylor for providing the *P. falciparum* GNP-A4 clone, E. Thompson for isolating *P. falciparum* DNA for whole genome analysis, Z. Gorvett for assistance with confirming single nucleotide polymorphisms in gametocyte non-producing clones, M. Duraisingh for the ddFKBP tagging construct pJDD145, C. Ben Mamoun for the anti-PP2c antibody and D. Goldberg for Shld1. M.L. is funded by National Institutes of Health (NIH) grant R01 AI076276 with support from the Centre for Quantitative Biology (P50GM071508). B.F.C.K. was supported by a Howard Hughes Medical Institute fellowship of the Damon Runyon Cancer Research Foundation. D.A.B. is funded by Wellcome Trust grant ref. 094752 and European Commission FP7 'MALSIG' (ref. 223044). L.G.D. is supported by a Biotechnology and Biological Sciences Research Council CASE PhD studentship with Pfizer as the Industrial partner. A.C. is funded by the Spanish Ministry of Science and Innovation grant SAF2010-20111. V.M.C. was supported by a fellowship from IRB Barcelona. T.G.C. is supported by the Medical Research Council UK (J005398). D.P.K. and S.G.C. are supported through the Wellcome Trust (098051; 090532/Z/09/Z) and the Medical Research Council UK (G0600230). C.B. is supported by the Catalan Government fellowship 2011-BP-B 00060 (AGAUR, Catalonia, Spain).

Author Contributions M.L. managed the overall project with input from B.F.C.K., D.A.B. and A.C. B.F.C.K. generated the Δ*pfap2-g* knockout, PfAP2-G–ddFKBP and luciferase lines and designed, performed and analysed the microarray, gel shift, luciferase and ligand-regulatable gametocytogenesis experiments. V.M.C. performed qRT–PCR validation. A.E.W. prepared Δ*pfap2-g* sequencing libraries and together with B.F.C.K. analysed the sequencing data. D.A.B., T.G.C. and S.G.C. conceived the sequencing of gametocyte non-producer lines F12 and GNP-A4. T.G.C. analysed the gametocyte non-producer sequencing data and L.G.D. confirmed the SNPs by PCR. S.G.C. and D.P.K. carried out and supervised sequencing of gametocyte non-producer lines, respectively. A.C. and N.R.-G. generated E5 and other 3D7-B subclones and respectively supervised and performed the experiments presented in Figs. 1 and 2b, and provided the analysis presented in Supplementary Fig. 1. V.M.C. and N.R.-G. performed and A.C. supervised chromatin immunoprecipitation experiments. C.B. and A.C. generated the PfAP2-G–HA×3 line and carried out immunofluorescence assays and correlations with gametocyte formation. B.F.C.K. wrote the manuscript with major input from M.L., D.A.B. and A.C.

Author Information Microarray data was submitted to the NCBI GEO repository (series accession GSE52030). Next generation sequencing data was submitted to the NCBI Sequence Read Archive (SRA) (study number ERP000190 for samples F12 (ERS011445), 3D7A (ERS011446) and GNP-A4 (ERS011447) and study number SRP035432 for samples E5 (SRS529791) and Δ*pfap2-g* (SRS529811)). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.L. (manuel@psu.edu).

METHODS

Parasites and strains. Parasite lines 3D7-A[31], 3D7-B[31] and F12[12] have been described previously. Note that 3D7-A is not the same line as the competent gametocyte producer 3D7A[32], which was used only as a reference genome for next-generation sequence analysis. 3D7-B subclones E5, A7 and B11 were generated by limiting dilution. The gametocyte non-producer line GNP-A4 was generated during an attempt to knockout a phosphodiesterase gene (PfPDEδ, PF14_0672). Integration of the knockout construct by single crossover homologous recombination occurred at the targeted locus but this event was not responsible for the clone's inability to produce gametocytes[33]. A subsequent successful knockout of PfPDEδ produced gametocytes at normal rates and the true phenotype was a significantly lower exflagellation rate than parental parasites owing to a reduced ability of gametes to egress from red blood cells[34]. Δ*pfap2-g* knockout parasites were generated by transfection of 3D7-B E5 with pHHT-FCU-pfap2-g (Supplementary Fig. 4) followed by positive (*hdhfr*)/negative (*fcu*) selection using WR99210 and 5-fluoro-cytosine as described previously[35]. Resistant parasites were subcloned and verified by PCR and Southern blot. *pfap2-g-ddfkbp* parasites were generated by transfection of 3D7-B E5 with pJDD145-pfap2-g and selected on WR99210. After subcloning, integration was verified by PCR using a forward primer at position +4,269 and a ddFKBP reverse primer. Displacement of the endogenous downstream sequence was verified using primers at +4,269 and +7,490 with respect to the translation initiation site (Supplementary Fig. 6). Parasites expressing a HA×3-tagged version of PfAP2-G were obtained by transfecting 3D7-B E5 with the plasmid pHH1inv-pfap2-g-HA×3 and cycling twice on/off WR99210 to select for parasites where the plasmid has integrated into the genome. After subcloning by limiting dilution and Southern blot analysis (Supplementary Fig. 7b), a subclone with a single copy of the plasmid integrated at the *pfap2-g* locus (E5-pfap2-g–HA×3 clone 9A) was selected for immunofluorescence assay (IFA) analysis. All parasites were grown in media containing AlbuMAX II and synchronized by standard methods[36].

Knockout and ddFKBP-tagging constructs. Knockout construct: the region from −126 base pairs (bp) to +366 bp and +6,945 to +7,379 bp with respect to the *pfap2-g* initiation codon were cloned into the NcoI/EcoRI and SpeI/SacII sites of pHHT-FCU[35], respectively, to generate pHHT-FCU-pfap2-g. ddFKBP carboxy-terminal tagging construct: *pfap2-g* coding sequence positions +4,740 to +7,296 were cloned into with NotI/XhoI sites of pJDD145[37] (gift from M. Duraisingh).

Luciferase expression constructs. The *hdhfr* selectable marker of pVLbIDh[38] was replaced with blasticidin-S deaminase using the SacI/NotI sites to generate pVL-BSD. The *var7b* promoter was excised with HpaI/KpnI, blunted and re-ligated, destroying these sites. The 1,445 bp, 1,226 bp and 1,159 bp upstream of the *pf11-1*, *pfg27/25* and *pfpeg4* start codons were cloned into the AatII/NcoI sites. The PfAP2-G cognate motifs in the upstream sequences of *pf11-1* (−328 to −323) and *pfpeg4* (−1,138 to 1,131) were converted to adenines using site-directed mutagenesis.

HA×3-tagging construct. The plasmid pHH1inv-pfap2-g-HA×3 was derived from the plasmid E140-0[39]. A triple HA tag (HA×3) was cloned into KpnI-XhoI sites of E140-0, replacing the *eba-140* open reading frame (ORF) and introducing several new restriction sites (plasmid pHH1inv_HA×3). A fragment of the *pfap2-g* ORF from position +6,685 to the stop codon was PCR-amplified and cloned in-frame into KpnI-PstI sites of pHH1inv_HA×3, such that upon integration of the plasmid by single homologous recombination PfAP2-G is expressed as a fusion protein with the HA×3 tag, separated by the sequence YLQ.

Gametocytogenesis. Gametocyte conversion rates for the 3D7-A, 3D7-B and 3D7-B subclones (Fig. 1) were measured by treating synchronized ring-stage parasite cultures at 5% parasitemia with *N*-acetyl-D-glucosamine and counting gametocytaemia 3–4 days later. Gametocyte induction of E5, F12, GNP-A4 and Δ*pfap2-g* (Fig. 2c) was performed according to published methods[27] in media containing 5% AB+ heat-inactivated human serum and 0.25% AlbuMAX II. For ligand-regulatable gametocytogenesis (Fig. 2d, e), synchronized parasites were set up at 0.5–1.0% late trophozoites in 3% hematocrit on day 0. Cultures were split in two and treated with 0.5 μM Shld1 (gift from D. Goldberg) or an equal volume of ethanol solvent control for the remainder of the experiment. Parasitemia was determined on day 3 and 50 mM *N*-acetyl-D-glucosamine was added to all cultures for the remainder of the experiment. Gametocytemia was determined on day 9 and converted into per cent commitment by dividing by day 3 parasitemia. Statistical significance in gametocyte production was determined using unpaired two-sided *t*-tests. Replicates were biological not technical.

Growth competition. E5 and Δ*pfap2-g* parasites were mixed at 1:1 ratio and grown for 7 or more weeks in triplicate. Parasites were diluted 1:20 with uninfected erythrocytes and genomic DNA was isolated whenever parasitemia exceeded 10%. For each time point, gDNA was isolated from the technical replicates, pooled at equal concentration, used for PCR amplification of a 414-nucleotide (nt) (1,718–2,131) region of PFF0275c covering the single nucleotide polymorphism (SNP) (described in Supplementary Table 1) and sequenced using the reverse primer. Difference in the growth rate was determined using the relative sequencing read peak height of A and C across at least 16 replication cycles (32 or more days). Growth rate was fit to the data using:

$$\mathrm{WT}_{t=x} = \mathrm{WT}_{t=0} / \left(\mathrm{WT}_{t=0} + \mathrm{KO}_{t=0} \times (1+\Delta g)^{t/2} \right)$$

$$\mathrm{KO}_{t=x} = 1 - \mathrm{WT}_{t=x}$$

where $\mathrm{WT}_{t=x}$ is the relative peak height of cytosine at day x, $\mathrm{KO}_{t=x}$ is the relative peak height of adenine at day x, and Δg is the per cent difference in growth rate between Δ*pfap2-g* and E5. 95% confidence intervals were determined using 1.96 × standard error of the mean for the difference in growth rate. Replicates were biological not technical.

Chromatin immunoprecipitation (ChIP). ChIP experiments were performed as described previously[40]. In brief, cultures were synchronized to late trophozoites/schizont stage, saponin-lysed and crosslinked using formaldehyde. Nuclei were released using a Dounce homogenizer (Kimble Chase) and DNA was subsequently fragmented using a Bioruptor (Diagenode). Immunoprecipitations were carried out using commercial antibodies against H3K9ac (Millipore 07-352) and H3K9me3 (Millipore 07-442) and analysed by qPCR using the relative standard curve method. The primers used for ChIP analysis of the *pfap2-g* locus amplify positions (relative to the start codon) −4,954 to −4,875 (5′-1), −1,412 to −1,302 (5′-2), −449 to −351 (5′-3), +3,874 to +3,979 (ORF-1), +5,318 to +5,433 (ORF-2) and +8,492 to +8,632 (3′-1). Primers for the control genes *clag3.1* (primer pair 5, beginning of the ORF), *clag3.2* (primer pair 5, beginning of the ORF), *ama-1* (primer pair 2, beginning of the ORF) and the *var* gene PFL1950w (upstream region, presumably 5′ untranslated region) have been described before[40,41]. Replicates were biological not technical.

Western blots. At the late trophozoite stage synchronized parasites were treated with 0.5 μM Shld1 or ethanol solvent control for 36 h. Proteins were sequentially extracted as described previously[2], separated on 4–12% polyacrylamide gels (Life Technologies) and assayed using anti-HA tag (Roche Diagnostics 11 867 423 001), anti-histone 3 (Abcam ab1791) and anti-PfPP2C (gift from C. Ben Mamoun). Replicates were biological not technical.

qRT–PCR. *pfap2-g* transcript abundance measurements were carried out as previously described[7] using a primer set designed to amplify positions +3,874 to +3,979 and normalized to seryl-tRNA synthetase abundance. Statistical significance was determined using a two-sided *t*-test. Replicates presented are technical and representative of several biological replicates.

IFAs. IFAs were performed on smears of E5-pfap2-g–HA×3-9A cultures synchronized to different stages. Air-dried smears were fixed for 10 min with 1% formaldehyde and permeabilized for 10 min in 0.1% Triton X-100 in PBS. Experiments performed on smears fixed with 90% acetone/10% methanol yielded identical results (not shown). Smears were incubated with rabbit anti-HA (1:100; Life technologies 71-5500) or rabbit anti-H3K4me3 (1:10,000; Millipore 05-745) antibodies. Secondary anti-rabbit antibodies were conjugated with Alexa Fluor 488 (Life technologies A-11034). Nuclei were stained with DAPI. Importantly, no wild-type E5 parasites were positive for staining with anti-HA antibody, and secondary antibody controls also yielded no signal. Preparations were observed under a confocal Leica TCS-SP5 microscope with LAS-AF image acquisition software and were processed using ImageJ software. The proportion of HA-positive schizonts was determined by counting >3,000 schizonts (identified by DAPI staining) for each experiment. The gametocyte conversion rate was measured for each of the same parasite cultures used to quantify the proportion of schizonts positive for anti-HA by IFA. Replicates were biological not technical.

Gel shifts. Electrophoretic mobility shift assays were performed using Light Shift EMSA kits (Thermo Scientific) as previously described using 2 μg of protein and 20 fmol of probe[21]. Biotinylated double-stranded probes were designed using the 24-nt flanking the PfAP2-G motif of the indicated upstream sequence. Probe sequences were as follows with capital letters indicating the AP2-G motifs and lowercase letters indicating the flanking sequences: *pfg27/25*: 5′-ttattagtatctGTACAC attggtatttgt-3′, *pf11-1*: 5′-tatatatatattGTACACatacatgtagtt-3′, *pfpeg4*: 5′-gacaataa agaaGTGTACACatatatcaataa-3′. The motifs were replaced by an equal number of adenines for 'no motif' probes. Replicates were conducted using the same materials on separate days.

Transcription profiling and associated analysis. Starting at 3 h post-invasion, tightly synchronized parasites were collected at eight time points with 6-h intervals. RNA isolation, cDNA generation/labelling, array hybridization and feature extraction was performed as described previously[28]. Cy5-labelled cDNA was hybridized

with a common Cy3-labelled reference pool on the *P. falciparum* 8×15K Agilent nuclear expression array (Gene Expression Omnibus (GEO) platform ID GPL17880). Relative transcript abundance was determined using a shared Cy3-labelled reference pool. All microarray data was submitted to the NCBI GEO repository (series accession number GSE52030). Genes were ordered by their average relative transcript abundance differences across the eight time points between the wild type (E5) and mutant (F12/Δ*pfap2-g*). Occurrences of the trimmed (6-nt) PfAP2-G motif were mapped using ScanACE to intergenic regions up to 2,000 bp upstream of the start codon as previously described[21] (see Supplementary Fig. 10 for motif). Significant enrichments of proteomic evidence and PfAP2-G motif occurrence were calculated using an unpaired two-sided *t*-test comparing the occurrences within the cluster of downregulated genes and their frequency genome-wide. Results were validated by qRT–PCR for a subset of downregulated genes using the primers in Supplementary Table 3 and methods described above. Statistically significant differences in relative expression levels were determined by two-sided *t*-test.

Luciferase assays. Equal numbers of synchronized, stably transfected parasites were isolated and saponin-lysed (0.05% in PBS) at ~18–30 h post-invasion and assayed using Bright-Glo Luciferase Assay System (Promega) as per the manufacturer's protocol on a Synergy H1 (Bio-Tek) plate reader. Statistical significance was determined using unpaired two-sided *t*-tests. Replicates were biological not technical.

Next-generation sequencing and analysis. Genomic DNA was extracted (10 μg each) from E5, *Δpfap2-g*, GNP-A4 and F12 parasite lines. This genomic DNA was used to generate barcoded sequencing libraries for an Illumina TruSeq single-end sequencing run, analysed and visualized as described previously[29]. Genomic DNA for 3D7A[32], F12[12] and GNP-A4[33] was also used for whole genome sequencing at the Sanger Institute using Illumina GA II technology with 76-base paired-end reads. The raw sequence data were processed as described previously[30]. In brief, the raw data for each isolate was mapped onto the 3D7 reference genome (version 3) using the SMALT short read alignment algorithm[42]. High-quality SNPs and insertions and deletions (supported by bidirectional reads, and error rates less than one per 1,000 bp) in unique genomic regions were called using SAMtools (http://samtools.sourceforge.net). Regions of interest were inspected using the Artemis alignment viewer (http://www.sanger.ac.uk/resources/software/artemis/), and polymorphisms compared to publically available sequence data[16,30,43] processed as described above. Experimental confirmation of informative genomic variants was performed using capillary sequencing methods.

31. Cortés, A., Benet, A., Cooke, B. M., Barnwell, J. W. & Reeder, J. C. Ability of *Plasmodium falciparum* to invade Southeast Asian ovalocytes varies between parasite lines. *Blood* **104,** 2961–2966 (2004).
32. Walliker, D. *et al.* Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236,** 1661–1666 (1987).
33. Taylor, C. J. *The role of two cyclic nucleotide phosphodiesterases in the sexual development of* Plasmodium falciparum *and* Plasmodium berghei. PhD thesis, Univ. London (2007).
34. Taylor, C. J., McRobert, L. & Baker, D. A. Disruption of a *Plasmodium falciparum* cyclic nucleotide phosphodiesterase gene causes aberrant gametogenesis. *Mol. Microbiol.* **69,** 110–118 (2008).
35. Maier, A. G., Braks, J. A. M., Waters, A. P. & Cowman, A. F. Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in *Plasmodium falciparum* for targeted gene deletion by double crossover recombination. *Mol. Biochem. Parasitol.* **150,** 118–121 (2006).
36. Ménard, R. *Malaria: Methods and Protocols. Methods in Molecular Biology* Vol. 923, 2nd edn, 3–15 (Humana Press, Springer, 2013).
37. Farrell, A. *et al.* A DOC2 protein identified by mutational profiling is essential for apicomplexan parasite exocytosis. *Science.* **335,** 218–221 (2012).
38. Calderwood, M. S., Gannoun-Zaki, L., Wellems, T. E. & Deitsch, K. W. *Plasmodium falciparum var* genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J. Biol. Chem.* **278,** 34125–34132 (2003).
39. Cortés, A. *et al.* Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog.* **3,** e107 (2007).
40. Crowley, V. M., Rovira-Graells, N., Ribas de Pouplana, L. & Cortés, A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. *Mol. Microbiol.* **80,** 391–406 (2011).
41. Jiang, L. *et al.* Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. *Proc. Natl Acad. Sci. USA* **107,** 2224–2229 (2010).
42. SMALT - Wellcome Trust Sanger Institute. (http://www.sanger.ac.uk/resources/software/smalt/).
43. The Wellcome Trust Sanger Institute SRA Study ERP000190 (http://www.ebi.ac.uk/ena/data/view/ERP000190).

LETTER

doi:10.1038/nature12970

A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*

Abhinav Sinha[1]*, Katie R. Hughes[1]*, Katarzyna K. Modrzynska[2]*, Thomas D. Otto[2], Claudia Pfander[2], Nicholas J. Dickens[1], Agnieszka A. Religa[1], Ellen Bushell[2], Anne L. Graham[1], Rachael Cameron[1], Bjorn F. C. Kafsack[3], April E. Williams[3,4], Manuel Llinás[3,4]†, Matthew Berriman[2], Oliver Billker[2] & Andrew P. Waters[1]

Commitment to and completion of sexual development are essential for malaria parasites (protists of the genus *Plasmodium*) to be transmitted through mosquitoes[1]. The molecular mechanism(s) responsible for commitment have been hitherto unknown. Here we show that PbAP2-G, a conserved member of the apicomplexan AP2 (ApiAP2) family of DNA-binding proteins, is essential for the commitment of asexually replicating forms to sexual development in *Plasmodium berghei*, a malaria parasite of rodents. PbAP2-G was identified from mutations in its encoding gene, PBANKA_143750, which account for the loss of sexual development frequently observed in parasites transmitted artificially by blood passage. Systematic gene deletion of conserved ApiAP2 genes in *Plasmodium* confirmed the role of PbAP2-G and revealed a second ApiAP2 member (PBANKA_103430, here termed PbAP2-G2) that significantly modulates but does not abolish gametocytogenesis, indicating that a cascade of ApiAP2 proteins are involved in commitment to the production and maturation of gametocytes. The data suggest a mechanism of commitment to gametocytogenesis in *Plasmodium* consistent with a positive feedback loop involving PbAP2-G that could be exploited to prevent the transmission of this pernicious parasite.

Malaria parasites spontaneously and stochastically produce sexual forms (gametocytes) required for mosquito transmission. Asexual parasites commit to sexual development in the erythrocyte and the cell-cycle-arrested male and female gametocytes are available to initiate transmission when ingested within the blood meal of a female anopheline mosquito. Gametocyte production may be lost when *Plasmodium* parasites are maintained either in continuous culture or by blood transfer between vertebrate hosts[1]. In a parasite line that produces fluorescently tagged gametocytes[2,3] we generated three gametocyte non-producer (GNP) lines (GNPm7, GNPm8 and GNPm9) that had verifiably lost the ability to undertake gametocytogenesis after 52 weeks of mechanical passage (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1).

Subsequent developmental stages (gametes, ookinetes) were absent and none of the GNP lines could be transmitted through mosquitoes (Supplementary Fig. 2 and Supplementary Table 2). Whole-genome sequencing of these and an existing GNP line (ANKA 2.33) revealed numerous single nucleotide polymorphisms (SNPs) and insertions or deletions (indels) per line (Supplementary Fig. 3 and Supplementary Table 3); however, only a single gene, PBANKA_143750, carried a different and therefore independent nonsense or missense mutation in each line (Fig. 1b). PBANKA_143750 (here termed *pbap2-g*) encodes a putative transcription factor predicted to be composed of 2,330 amino acids with a single 55-amino-acid AP2 class DNA-binding domain (DBD) at its carboxy terminus (Fig. 1b). PbAP2-G belongs to the 27-strong[4,5] *Plasmodium* ApiAP2 family of transcription factors, themselves part of the larger Apetala 2/ethylene response factor (AP2/ERF) family of transcription factors restricted to the Plantae and apicomplexan protists. The role of PbAP2-G in gametocyte production was confirmed either by correcting the mutations in *pbap2-g* in the GNP lines through genomic recombination with a wild-type copy (generating GNPm7REP, GNPm8REP, GNPm9REP and 2.33REP) or genetic complementation of a targeted deletion mutant of *pbap2-g* (Fig. 1c and Supplementary Fig. 4a–g). Functionality of the restored gametocytes was demonstrated in GNPm7REP and 2.33REP by transmission through mosquitoes (Fig. 1d and Supplementary Table 4). Disruption of a second ApiAP2 gene, PBANKA_103430 (*pbap2-g2*) (Fig. 1b), resulted in the nearly complete (>95%) loss of mature gametocytes, but in contrast to *pbap2-g*$^{-}$ parasites, small numbers of female gametocytes were occasionally observed (Fig. 1c). These were not, however, transmitted successfully to mosquitoes. In direct growth competition assays *pbap2-g*$^{-}$ parasites outgrew wild-type *P. berghei* and *pbap2-g2*$^{-}$ parasites, which had wild-type growth rates (Fig. 1e and Supplementary Fig. 5). *pbap2-g*$^{-}$ mutants are therefore uniquely capable of converting a loss of gametocytes into increased asexual growth, which confers an advantage during asexual growth and explains why continued blood passage invariably selects for mutations in *pbap2-g*. This demonstrates that PbAP2-G functions specifically at the point of commitment, whereas PbAP2-G2 is required downstream, once sexual differentiation has become irreversible (Fig. 1e).

In a protein-binding microarray the recombinant DBD of PbAP2-G[6,7] recognized closely related DNA motifs (Fig. 2a and Supplementary Table 5) identical to the previously derived motif for the DBD from the orthologous ApiAP2 protein of *Plasmodium falciparum* (PF3D7_1222600)[6], confirming that both DBDs bind primarily to the same (GxGTACxC) motif (in which x denotes any residue). Electrophoretic mobility shift assay (EMSA) analyses (Fig. 2a) refined the motif to two 6-mers (GxGTAC and GTACxC, which are essentially palindromes of each other) that are sufficient and necessary for binding. A single point mutation in the core GTAC was sufficient to abrogate binding (Fig. 2a). These two motifs occurred within 2 kilobases (kb) upstream of 49% of all genes (2,359 of 4,803 considered), yet more frequently in genes designated as upregulated in gametocytes (246 (54%) of 452 genes; $P < 0.002$, hypergeometric test). The occurrence of both motifs upstream of *pbap2-g* itself suggested the potential for an autoregulatory feedback mechanism, and the regions of the genome containing these motifs upstream of *pbap2-g* were both recognized by PbAP2-G in EMSA analysis (Fig. 2a). Expression analysis demonstrated transcription of *pbap2-g* in blood-stage parasites; however, epitope tagging of full-length *pbap2-g* produced no detectable protein (Supplementary Fig. 6) yet gametocytogenesis was unaltered, implying that tagged PbAP2-G activity is unaffected. However, a truncated cyan fluorescent protein (CFP)-tagged transgene product could be detected in nuclei of female gametocytes (Fig. 2c and Supplementary Fig. 7).

Comparative microarray analyses showed that gametocyte-specific genes were highly enriched among the 500 most downregulated genes

[1]Wellcome Trust Centre for Molecular Parasitology, University of Glasgow, Glasgow G12 8QQ, UK. [2]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [3]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544-1014, USA. [4]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. †Present address: Department of Biochemistry and Molecular Biology and Center for Infectious Disease Dynamics, The Pennsylvania State University, State College, Pennsylvania 16802, USA.
*These authors contributed equally to this work.

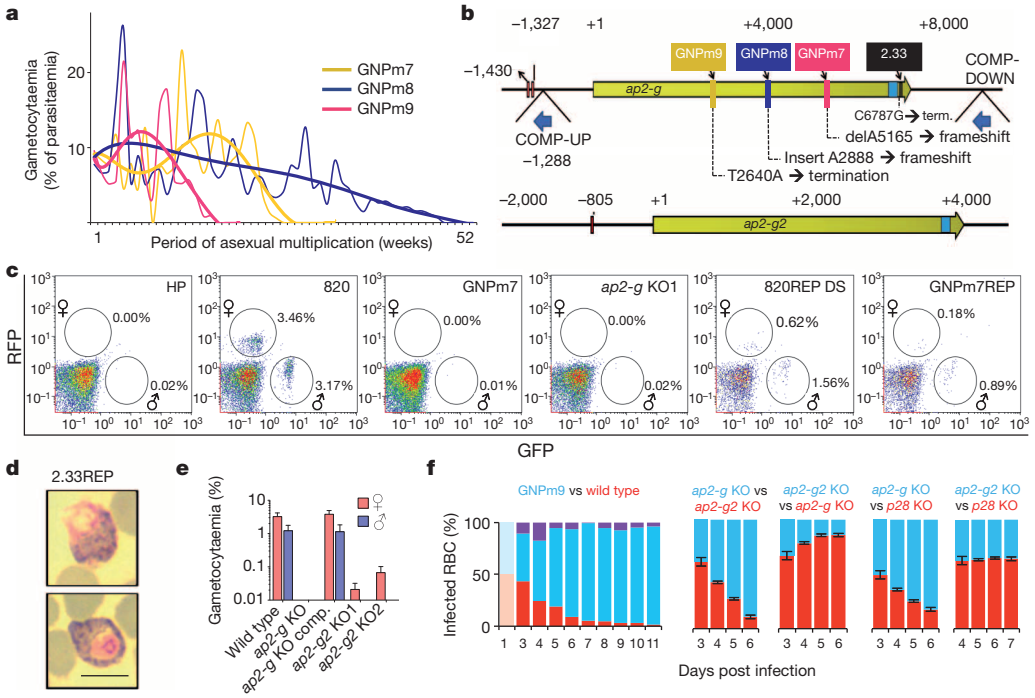


Figure 1 | Identification of mutations in *pbap2-g* that account for the repeated spontaneous loss of commitment to gametocytogenesis. **a**, Gametocyte production during a year of continuous mechanical passage of *P. berghei*. Best-fit polynomial trend (thick) lines of gametocytaemia on individual weekly observations (thin lines). **b**, Open reading frames (ORF) (yellow) of *pbap2-g* (PBANKA_143750) and *pbap2-g2* (PBANKA_103430) with point mutations in new GNP lines shown in **a** and the long-established line 2.33. Predicted DBDs (light blue) and DBD recognition motifs for PbAP2-G upstream of each ORF (brown bars) are indicated. Dark blue arrows show integration sites for selectable marker cassettes as used for genetic complementation of GNPs (COMP-DOWN) or to disrupt the promoter (COMP-UP). Numbering is relative to position 1 of the ORF. **c**, FACS analyses of male and female gametocyte numbers (circled areas) expressed as a percentage of the total parasitized cell counts. From left, *P. berghei* ANKA HP line (which lacks green (GFP) or red (RFP) fluorescent protein reporters, thus having no fluorescent signal and from which all subsequent lines reported in this study were derived) served as a negative control. Line 820 is the reporter line from which GNP mutants and a targeted knockout (KO) (using vector PbGEM-072446) were derived. 820REP and GNPm7REP were generated with the COMP-DOWN complementation vector. **d**, Giemsa-stained gametocytes in GNP line 2.33 (G756) repaired by the COMP-DOWN construct and after a single transmission through mosquitoes. Scale bar, 6 μm. **e**, Gametocyte quantification from manual counting in Giemsa-stained blood smears of an independently produced *pbap2-g* deletion mutant before and after complementation (comp.) with the DS (downstream) vector and of two independent *pbap2-g2* knockout mutants. Error bars show standard deviations from three replicates. The loss of gametocytes from the knockout mutants was significant ($P < 0.05$). **f**, Relative growth kinetics of GNPm9, *pbap2-g*$^{-}$ and *pbap2-g2*$^{-}$ lines determined by flow cytometry. Left, cloned GNPm9 constitutively expressing CFP (line GNPm9-CFP) was mixed in a 1:1 ratio with wild-type (PBANKA HP) producer line constitutively expressing RFP (line WT-RFP). The daily percentage of the population expressing either RFP (red), CFP (blue) or both (purple; reflecting cells infected with multiple parasites) was calculated. Right four panels, deletion vectors for *pbap2-g*, *pbap2-g2* or *p28* (control gene for neutral growth rate) were transfected in GFP- or mCherry-expressing lines (blue and red bars, respectively) and the relative abundance of each mutant determined in mixed infections of uncloned parasites. Error bars show ± standard deviations from three biological replicates. The competitive advantage was significant for the *pbap2-g*$^{-}$ ($P < 0.01$) but not the *pbap2-g2*$^{-}$ parasites (two tailed Student's *t*-test for change in relative abundance). RBC, red blood cell.

in GNP lines ($P < 10^{-51}$, Fisher's exact test), *pbap2*-deletion parasites ($P < 10^{-74}$) and in the *pbap2-g2* deletion mutant ($P < 10^{-49}$), although less marked in the latter (Table 1 and Supplementary Fig. 6). Comparison of the transcriptomes of wild-type asexual blood-stage parasites with those of various *pbap2-g*$^{-}$ lines was performed in an attempt to identify early-transcribed genes downstream of and under control of PbAP2-G (Fig. 3a). The steady-state transcription levels of 307 genes were identified as being downregulated (>2 s.d. reduced from the mean, Supplementary Table 6) in schizonts.

The activity of 18 promoters consistently downregulated in GNP lines, and which contain one or more candidate PbAP2-G-binding motifs, was analysed in wild-type and GNPm9 parasite backgrounds. Male, female or sex-specific genes downstream of AP2-G in the gametocyte developmental pathway were identified (Fig. 3b, Supplementary Fig. 8 and Supplementary Table 8). Single point mutations in PbAP2-G-binding motifs did not significantly reduce stage- or sex-specific expression of all of a number of reporter genes *in vivo*, even if identical changes ablated DNA binding *in vitro*. Only larger promoter truncations produced an impact on expression (Supplementary Fig. 9). Therefore, the relatively simple and highly abundant PbAP2-G motif is only active in context and its presence not always indicative of a critical role for the activity of a particular promoter. The PbAP2-G motifs upstream of *pbap2-g* do appear to be important as gametocytogenesis is blocked when the allelic motifs are both deleted, supporting the concept that commitment to gametocytogenesis requires a positive feedback loop powered by PbAP2-G itself (Fig. 3c).

The discovery of the ApiAP2 family[4] was the first identification of predicted transcription factors in apicomplexan genomes, otherwise thought to be remarkably lacking in genes encoding transcription factors[8]. The majority of ApiAP2 transcription factors are probably essential, involved in the progression of the intraerythrocytic asexual development of *Plasmodium*. Roles for additional ApiAP2 factors in the continuation of development of parasite forms associated with transmission have been demonstrated, namely for the ookinete (PbAP2-O[9]), sporozoite (PbAP2-S[10]) and liver stages (PbAP2-L[11]) of development. ApiAP2 proteins may also silence genes, possibly through maintenance of heterochromatin[12]. The AP2/ERF family members in *Plasmodium* are predicted to act singly or in combinations that control the continuation of the transcriptional

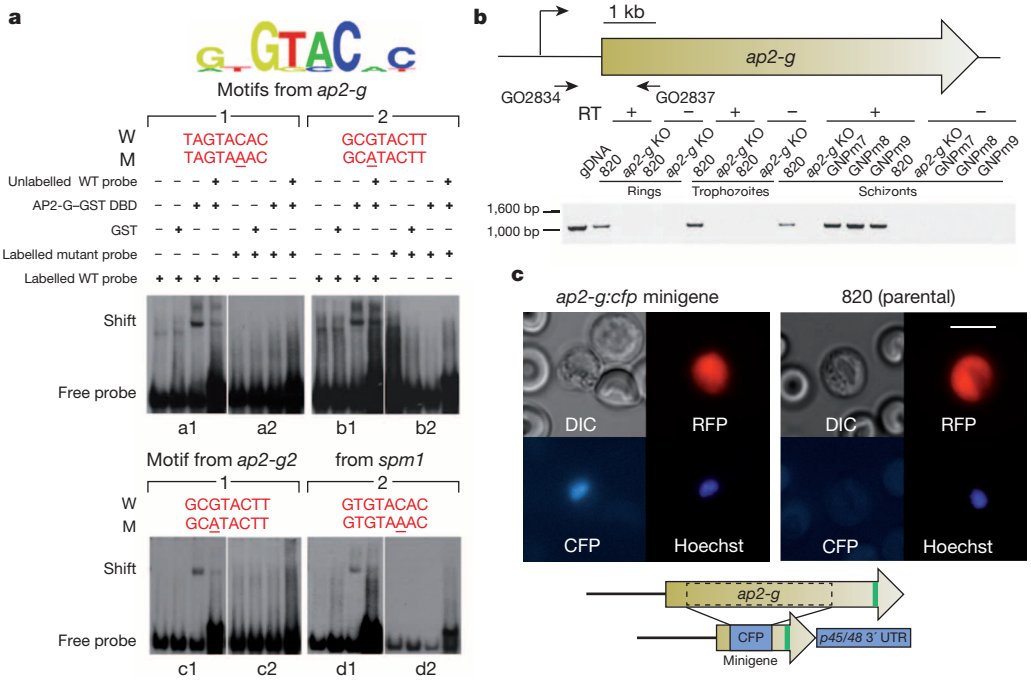


Figure 2 | Characterization of the DNA-binding specificity, expression and subcellular localization of PbAP2-G. **a**, Top, protein binding microarray determination of the DNA binding recognition preference of the recombinant DBD of PbAP2-G. GST, glutathione *S*-transferase. Bottom, EMSA in which a shift indicates whether the PbAP2-G DBD binds to double-stranded DNA containing wild-type (W) or mutated (M) motifs (panels a1–d1 and a2–d2, respectively) from the upstream regions of *pbap2-g* itself, *pbap2-g2*, and position −610 of the hypothetical gene *spm1* (subpellicular microtubule protein 1, PBANKA_081070). **b**, Expression analysis by reverse-transcriptase (RT)–PCR of *pbap2-g* in targeted and spontaneous *pbap2-g* mutants and the wild-type control line, 820. The 1.15-kb product indicates lack of transcript only in the targeted knockout line. Primer positions were as shown in the schematic. See Supplementary Fig. 7 for *pbap2-g* transgene expression data. $n = 3$. gDNA, genomic DNA. GO, Glasgow oligo. **c**, Localization of the *pbap2-g* minigene product to the nucleus of *P. berghei* female gametocytes. CFP was sandwiched between the N-terminal 300 base pairs (bp) and the C-terminal 800 bp of *pbap2-g*, including the DBD, and expressed from 2 kb of the *pbap2-g* promoter in line 820. Expression was only detected in the nuclei of female gametocytes (>50 observations in three experiments). It is the C-terminal segment that determines the nuclear localization of PbAP2-G (Supplementary Fig. 8). Scale bar, 6 μm. Cartoon is not to scale. DIC, differential interference contrast.

programme of the *Plasmodium* life cycle[4,6]. Heritable gene-regulatory strategies include epigenetic marks, stable cytoplasmic factors and transcriptional autoregulatory circuits that can determine distinct cell fates[13]. In the latter, commitment to a specific developmental pathway (for example, gametocytogenesis) is probabilistic, its frequency being defined by the likelihood of the interaction of a fate-determining transcription factor with a critical promoter often triggering a positive autoregulatory feedback loop that commits the cell[14], a paradigm that has been invoked within the *Plasmodium* AP2 transcription factor network[6]. *P. falciparum* uses precise epigenetic control to influence the sub-nuclear location of *pfap2-g*[15] and therefore possibly PfAP2-G binding which, when coupled to an autoregulatory positive feedback loop (Fig. 3c) involving PfAP2-G production, could provide flexible control of gametocytogenesis in a manner that would also be amenable to environmental sensing[16,17]. AP2-G is, at present, unique within the apiAP2 transcription factor family in that it directs a change in developmental fate rather than merely progressing a lineage (Supplementary Fig. 10), distinguishing it from AP2-G2 and from a number of other genes required for gametocyte maturation[18]. This critical role of AP2-G is conserved in *P. falciparum*[19], even though models for the timing of commitment in the two parasites differ[20,21]. Orthologues of the *ap2-g* DBD are present in all sequenced Apicomplexa, raising the possibility that mechanisms of commitment to sexual development may also be conserved (Supplementary Fig. 11). Thus these data identify the earliest known event in parasite transmission. Because it occurs in the blood of the host it is amenable to and suggests novel control strategies largely through drug development and

Table 1 | Changes in gene expression in mutants

Gene ID	Description	Rank	GNP	*P*	*pbap2-g* KO2	*pbap2-g2* KO1
051500	25-kDa ookinete surface antigen	1	−4.56	2.5×10^{-2}	−4.88	−1.72
051490	28-kDa ookinete surface antigen	2	−3.48	2.9×10^{-2}	−6.28	−2.37
133370	Phosphodiesterase delta	125	−3.61	1.3×10^{-2}	−3.89	−1.32
121910	Heat-shock protein 90	175	−3.34	7.6×10^{-2}	−3.67	−1.93
142170	Secreted ookinete protein, putative	62	−3.95	1.0×10^{-1}	−3.98	−1.42
131950	LCCL domain-containing protein CCP2	64	−3.09	6.2×10^{-2}	−3.79	−1.31
146300	Osmiophilic body protein	232	−1.63	1.2×10^{-1}	−2.60	−0.27
112040	Pfs77 homologue, putative	52	−2.68	3.4×10^{-2}	−3.50	−0.78
134040	Oxidoreductase, putative	327	−4.59	5.9×10^{-2}	−2.80	−1.77
123130	Metabolite/drug transporter, putative	26	−3.31	5.0×10^{-2}	−2.82	−1.46

Gene expression was determined on Agilent microarrays for *in vitro*-cultured schizonts, comparing pooled GNP clones and targeted mutants to their parental control lines. Log_2 fold changes are shown for the top 10 genes with good functional annotation that were most strongly deregulated in the targeted mutant *pbap2-g* KO1. Gene IDs are given without their PBANKA_ prefix. Rank refers to the absolute expression rank among 4,553 genes in purified gametocytes determined from three biological replicates. Expression data are means from three biological replicates for each mutant. *P* denotes the *P* value adjusted for multiple testing. For the complete data and all *P* values see Supplementary Table 6.

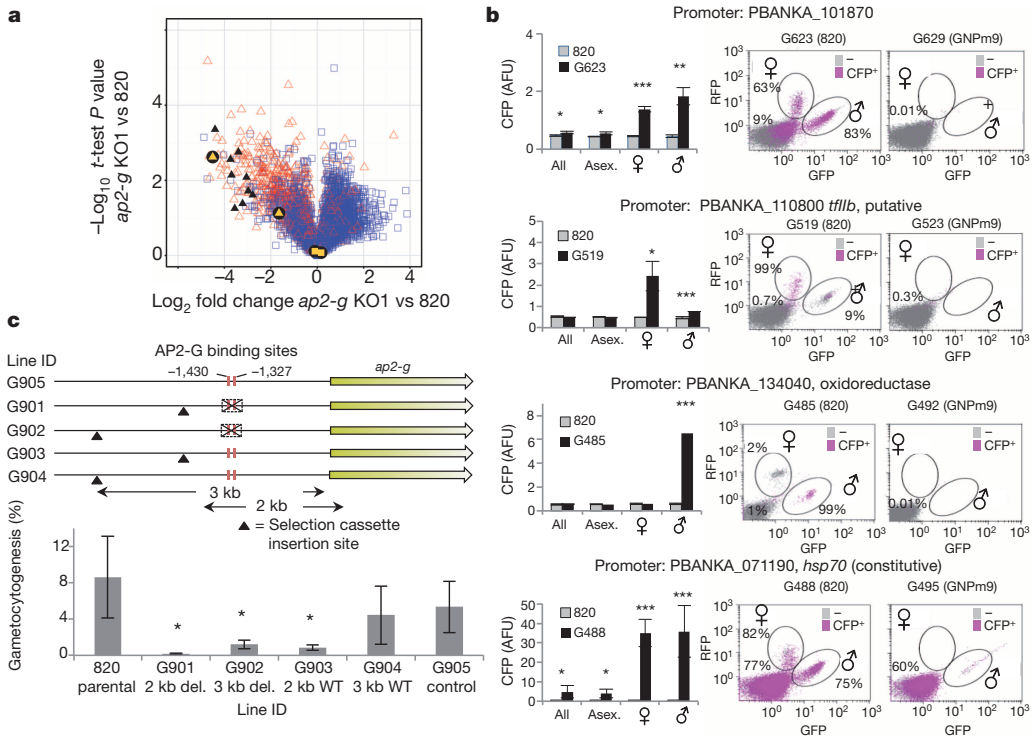


Figure 3 | *pbap2-g* acts upstream of gametocyte gene transcription. **a**, Volcano plot of $\log_2$ fold change in gene expression in schizonts of *pbap2-g* KO1 (whole ORF deletion) versus wild-type line 820 against significance of change ($-\log10$ *t*-test). Red triangles indicate genes upregulated in gametocytes compared to schizonts. Black and yellow shapes are genes detailed in Table 1 and Fig. 3c, respectively. **b**, Reporter-gene expression constructs were transfected into the GNPm9 and 820 control clones to confirm gametocyte-gene-specific promoters. Reporters contained 2 kb of upstream sequence from the indicated genes driving CFP expression with a constitutive 3′ untranslated region. Bar plots show CFP measured by flow cytometry over 3 days in the 820 line. Life cycle stages (asexual, male and female) are separated on the basis of GFP or RFP expression. Mean of three measurements (geometric mean CFP fluorescence) ± s.d.; $*P < 0.05$, $** P < 0.01$, $***P < 0.001$, two-tailed *t*-test. Flow cytometry plots are shown for CFP expression of reporters in 820 (parental) (left) or GNPm9 (right) lines. Plots show GFP (*x* axis) versus RFP (*y* axis) expression for all infected red blood cells and CFP expression in magenta. Numbers on each plot represent the percentage of events within each gate that are positive for CFP (see also Supplementary Fig. 8). **c**, Deletion studies in the *pbap2-g* promoter provide support for a role of PbAP2-G binding motifs in the positive feedback regulation of *pbap2-g* expression. Top, DNA constructs containing a selectable marker were integrated into the promoter region of *pbap2-g* in PBANKA 820. The constructs either deleted 207 bp surrounding the two instances of the PbAP2-G binding motif at the positions indicated (G901 and G902) or did not (G903 and G904). Two sites of selectable marker integration were tested, 2 and 3 kb upstream of the ORF of *pbap2-g*. In addition, interruption at −1,288 upstream of the ORF of *pbap2-g* was shown to disrupt gametocytogenesis (Supplementary Fig. 4f). Control line G905 was transfected with a reporter construct targeted to the *p230p* locus and known not to affect gametocytogenesis. Bottom, gametocytaemia was measured on consecutive days by flow cytometry once the parasitaemia reached >1%. Mean ± s.d. shown, $*P < 0.05$ compared to 820 parental (two-tailed *t*-test). Data shown are pooled from 3 days' observations and representative of three independent experiments.

offers some strategic value in the prevention of sexual development and reduction of transmission.

METHODS SUMMARY

P. berghei ANKA parasites were maintained in female Theiler's original (TO) mice (6–8 weeks old) under appropriate Home Office licences. A fluorescent reporter line 820 (ref. 3) for male (green) and female (red) gametocytes was transmitted weekly by blood passage into a new host for up to 52 weeks in 10 parallel lines and gametocytaemia assessed weekly by flow cytometry. Whole-genome sequencing was followed by *de novo* assembly and variant calling. Targeted gene knockouts were generated using traditional plasmids or *Plasmo*GEM vectors[22]. GNP phenotypes were confirmed by a variety of methods. Genetic complementation was by ends-out recombination over the region mutated in GNP clones and confirmed functionally by FACS and mosquito passage. A *pbap2-g* DBD–GST fusion protein was used in protein binding microarray analysis as described[6]. The purified GST-recombinant protein was used in EMSA assays with 60-mer biotinylated annealed oligonucleotides. Microarray analysis was performed on total RNA on an Agilent array[23] and data submitted to the Gene Expression Omnibus (GEO) database. Reporter constructs were transfected into 820 and GNPm9. Reporter expression was monitored by FACS over several days. The promoter of *pbap2-g* was modified by ends-out integration into 820 and gametocytaemia monitored over several days using flow cytometry.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April; accepted 19 December 2013.
Published online 23 February 2014.

1. Janse, C. J. *et al. Plasmodium berghei*: *in vivo* generation and selection of karyotype mutants and non-gametocyte producer mutants. *Exp. Parasitol.* **74,** 1–10 (1992).
2. Mair, G. R. *et al.* Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. *PLoS Pathog.* **6,** e1000767 (2010).
3. Ponzi, M. *et al.* Egress of *Plasmodium berghei* gametes from their host erythrocyte is mediated by the MDV-1/PEG3 protein. *Cell. Microbiol.* **11,** 1272–1288 (2009).
4. Balaji, S. *et al.* Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.* **33,** 3994–4006 (2005).
5. Painter, H. J., Campbell, T. L. & Llinas, M. The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. *Mol. Biochem. Parasitol.* **176,** 1–7 (2011).
6. Campbell, T. L. *et al.* Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* **6,** e1001165 (2010).
7. Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols* **4,** 393–411 (2009).
8. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419,** 498–511 (2002).
9. Yuda, M. *et al.* Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol. Microbiol.* **71,** 1402–1414 (2009).
10. Yuda, M. *et al.* Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol. Microbiol.* **75,** 854–863 (2010).
11. Iwanaga, S. *et al.* Identification of an AP2-family protein that is critical for malaria liver stage development. *PLoS ONE* **7,** e47557 (2012).

12. Flueck, C. *et al. Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog.* **5,** e1000569 (2009).
13. Burrill, D. R. & Silver, P. A. Synthetic circuit identifies subpopulations with sustained memory of DNA damage. *Genes Dev.* **25,** 434–439 (2011).
14. Shiels, B. R. Should I stay or should I go now? A stochastic model of stage differentiation in *Theileria annulata*. *Parasitol. Today* **15,** 241–245 (1999).
15. Lopez-Rubio, J. J., Mancio-Silva, L. & Scherf, A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5,** 179–190 (2009).
16. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331,** 76–79 (2011).
17. Cameron, A., Reece, S. E., Drew, D. R., Haydon, D. T. & Yates, A. J. Plasticity in transmission strategies of the malaria parasite, *Plasmodium chabaudi*: environmental and genetic effects. *Evol. Appl.* **6,** 365–376 (2013).
18. Ikadai, H. *et al.* Transposon mutagenesis identifies genes essential for *Plasmodium falciparum* gametocytogenesis. *Proc. Natl Acad. Sci. USA* **110,** E1676–E1684 (2013).
19. Kafsack, B. F. C. *et al.* A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* http://dx.doi.org/10.1038/nature12920 (this issue).
20. Janse, C. J. *et al.* Variation in karyotype and gametocyte production during asexual multiplication of *Plasmodium berghei*. *Acta Leiden.* **60,** 43–48 (1991).
21. Bruce, M. C., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology* **100,** 191–200 (1990).
22. Pfander, C. *et al.* A scalable pipeline for highly effective genetic modification of a malaria parasite. *Nature Methods* **8,** 1078–1082 (2011).
23. Kafsack, B. F., Painter, H. J. & Llinás, M. New Agilent platform DNA microarrays for transcriptome analysis of *Plasmodium falciparum* and *Plasmodium berghei* for the malaria research community. *Malar. J.* **11,** 187 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. McBride for technical assistance with mosquito transmissions; A. R. Gomes, J. Tripathi, D. Vaughan and the III flow cytometry facility for assistance; C. Cairney and N. Keith at the Beatson Institute for use of their Agilent microarray scanner; and A. Cortes and D. Baker for reading of the manuscript. A.P.W. is funded by the Wellcome Trust (ref. 083811/Z/07/Z). The Wellcome Trust Centre for Molecular Parasitology is supported by core funding from the Wellcome Trust (085349). A.S. was a student of the University of Glasgow Wellcome Trust 4-year PhD Programme Molecular Functions in Disease. A.P.W., O.B. and M.B. are members of Evimalar (ref. 242095), which funds the work of T.D.O. Work at the Sanger Institute was funded by grants from the Wellcome Trust (098051) and the Medical Research Council (G0501670). M.L. is funded by the National Institutes of Health (R01 AI076276) and the Centre for Quantitative Biology (P50GM071508). B.F.C.K. was supported by a Howard Hughes Medical Institute fellowship of the Damon Runyon Cancer Research Foundation.

Author Contributions A.P.W. and O.B. directed the research. A.S. generated the GNP clones, performed some of the EMSA analyses, made *pbap2-g* gene knockout lines and complementation lines and analysed the latter. K.R.H. performed microarray analyses, generated reporter and minigene constructs, made transgenic parasites and analysed them, performed competition experiments. K.K.M. made the complementation construct, generated and analysed knockout and complemented lines for *pbap2-g* and *pbap2-g2* and performed and analysed competition and microarray experiments. C.P. generated knockout lines for *pbap2-g* and *pbap2-g2* and performed the initial parasitological analysis. E.B. generated recombinase engineered constructs for use at Wellcome Trust Sanger Institute and University of Glasgow. A.L.G. and A.A.R. performed expression analyses. N.J.D. performed statistical analyses of motif distribution and assisted with the microarray analyses. R.C. performed the complementation experiments and transmission experiments. A.E.W. performed EMSA analyses and generated constructs used in the analysis. T.D.O. and M.B. generated the GNP sequence data and SNP analyses. M.L. and B.F.C.K. performed microarray analyses, M.L. and A.E.W. performed EMSA analyses and generated recombinant PbAP2-G DBD. A.P.W., O.B., A.S., K.R.H. and K.K.M. wrote the paper.

Author Information Microarray data has been submitted to the GEO database under accession numbers GSE52859 and GSE53246. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.B. (OB4@sanger.ac.uk) or A.P.W. (Andy.Waters@glasgow.ac.uk).

METHODS

Parasite lines and methods. *P. berghei* ANKA HP was obtained from C. Janse at Leiden University Medical Centre and was originally referred to as clone 15Cy1A (Leiden Malaria Group website). Line 820 was generated from HP. *P. berghei* ANKA clone 2.33 is a non-gametocyte-producing clone line reported in 1990 and is now widely distributed and grown by mechanical passage[24]. All infections were performed on female Theiler's original (TO) mice (age 6–8 weeks; weight 25–30 g) according to Home Office licence regulations and the local ethical committees. All animals were assigned to experiments without pre-selection and no blind assignations were performed. Serial passage of freshly cloned *P. berghei* reference line 820 m1cl1 (ref. 3) was performed as follows: 10 mice (m1–m10) were initially infected with 200 µl of a 1:200 dilution of a mouse infected with line 820 at a parasitaemia of ~2%. In the absence of any a priori information concerning mutation rates in *P. berghei* a sample size of 10 was selected based on concerns of animal welfare, cost and logistics. Each week, the infections were passaged to a further 10 mice in a similar manner when the parasitaemia was >1%. Parasitaemia and gametocytaemia were monitored by examination of Giemsa-stained blood films and by flow cytometry as described[25–29]. The infected blood from each mouse was also cryopreserved each week. Passage to a fresh mouse was halted when a line was negative for gametocyte production for 4 consecutive weeks and designated GNPmx, where x would be 1–10. The experiment was halted after 52 weeks. Lines GNPm7, GNPm8 and GNPm9 were cloned by limiting dilution, clones subjected to negative selection[30] to remove the selectable marker residual in the GFP:RFP selection cassette and cloned once more. Each parasite cloning procedure used 10 mice, and mice were infected by intravenous tail injection with an average of 1.5 parasites, which in our experience will give rise to 4 infected mice. Negative selection involved 3 mice, the infections of which were assayed by PCR for completeness of selection. Lines generated in this way were designated m(7,8,9)mxClx, indicating the mouse and clone number identifiers from the negative selection process. In the main text these cloned negatively selected lines are simply referred to GNPm7, GNPm8 and GNPm9.

Transfection of GFP- and RFP-expressing 'wild-type' parasites from the *P. berghei* line 820 with linearized targeting constructs, selection and cloning of the mutant parasites were performed according to procedures described previously[31]. Genotypic analysis of transfected parasites was performed by Southern analysis of chromosomes separated by field-inversion gel electrophoresis and using diagnostic PCR on genomic DNA. Details of the primers used for PCR are shown in Supplementary Table 9. Phenotype analysis of mutant parasites during blood-stage development, quantification of gametocyte production and ookinete development *in vitro* was performed using standard methods as described previously[26–29]. Mosquito-stage development was analysed in *Anopheles stephensi* mosquitoes using standard methods of mosquito infection, analysis of oocyst and sporozoite production and sporozoite infectivity to TO mice[32]. The capacity of wild-type and engineered parasites to infect mice by mosquito-interrupted feeding was determined by exposure of female TO mice ($n = 2–4$) to 40–50 mosquitoes at day 21 after the infectious blood meal. Infection was monitored by analysis of blood-stage infection in Giemsa-stained films of tail blood at day 4 until day 8 after infection. Infectivity was recorded as 'wild type' if mice developed a parasitaemia of 0.1–0.5% at day 4 after infection. For the 2.33 rescue experiment, images representative of >80 gametocytes at a parasitaemia of 8.2% and gametocytaemia of 5.4% are shown in Fig. 1d; similar results were seen on 3 consecutive days.

DNA-sequencing. To sequence clones 2.33, 820, GNPm7, GNPm8 and GNPm9, libraries of 300–500-bp fragment length were generated following a PCR-free protocol[33]. The libraries were sequenced using an Illumina Genome Analyser II with the V4 chemistry. Summary of reads for each project including accession codes are given in Supplementary Table 3. Data are available at http://www.ebi.ac.uk/ena/data/view/ERP000253.

Sequencing: *de novo* assembly. We generated a *de novo* assembly of reads from the 820 parental clone using with velvet[34] version 1.0.12 and the following parameters: -exp_cov auto -min_contig_lgth 500 -cov_cutoff 10 -ins_length 350 -min_pair_count 20. We obtained 417 supercontigs with an average length (N50) of 240 kb. We processed the assembly as described in the post-assembly genome-improvement toolkit protocol[35]. In short, scaffolds were ordered with ABACAS[36] against the *P. berghei* ANKA reference genomes (GeneDB, version July 2010). This resulted in 16 pseudomolecules (14 chromosomes and 2 plastids) and a 'bin' of 100 contigs that could not be associated with a chromosome. Next, using scaffolds of at least 1 kb as a substrate, IMAGE[37] was used to close 469 (61%) of the 774 sequencing gaps. Single-base and indel errors were corrected using ICORN[38]. This corrected 1,067 single-base errors and 92 indels. 1,589 positions had heterozygous calls, which represented collapsed repeats, mostly in *P. berghei* interspersed repeat (*bir*) genes. Last, the annotation of the *P. berghei* ANKA reference genome was transferred onto the improved *P. berghei* 820 assembly using RATT[39] (Assembly option). In total, 4,821 of the 4,938 gene models were transferred correctly. The assembly is available on ftp://ftp.sanger.ac.uk/pub/pathogens/Plasmodium/berghei/820/.

Sequencing: variant calls. To call variants, SMALT (version 0.6.2, http://www.sanger.ac.uk/resources/software/smalt/, parameters: -r 0, -x, -y 0.8, -i 1000, and for index a k-mer size of 17 (-k) and a step size of 3 (-s)) was used to map reads against the generated 820 assembly. After generating bam files with the SAMtools package[40], variation was called with GATK[41] (parameters -ploidy 1 -glm POOLBOTH -pnrm POOL). For the reads mapped onto the 820 assembly, the variation of each clone, and concordance with other clones was analysed using a PERL script. For the reads mapped onto the ANKA reference genome, the script ignored variants that were called in all m7–m9 clones as well as 820. The quality filter for a variant was 60. The pipeline for whole-genome sequencing and identification of single nucleotide polymorphisms is summarized in Supplementary Fig. 12.

Variant calling in *Plasmodium* from re-sequencing data are inherently noisy, owing to false calls within repeats and low-complexity regions. Thus, 3 independent clones were used to identify coincident site(s). Isolate-specific variation is catalogued in Supplementary Table 3 and the large proportion of heterozygous calls are highlighted (a manifestation of calling variants within repetitive and low-complexity regions).

All data were generated using *ad hoc* scripts (available upon request). The variant (.vcf) files of the each isolate are available from ftp://ftp.sanger.ac.uk/pub/pathogens/Plasmodium/berghei/820/vcf.

Phylogenetic analysis. Data were generated from the results of a BLASTP search of EuPathDB Apicomplexa using the AP2 domain from PBANKA_143750 as the query. Significant hits were defined as those that covered at least 75% of the length of the query domain and had >50% conserved residues. Neighbour joining tree was generated in CLC Genomics Workbench (version 6.5.1) using the Jukes-Cantor protein distance measure. Values shown are for 1,000 bootstrap iterations. The tree is rooted using the most distant *Arabidopsis thaliana* DBD Q9M0l0.2.

Recombinant protein production. N-terminal GST-fused extended ApiAP2 DBDs (cloned into pGEX-4T1) from *P. falciparum ap2-g* (PFL_1085w) and *P. berghei ap2-g* (PBANKA_143750) were expressed in Rosetta (DE3) pLys S-competent cells with 0.2 mM IPTG at 25 °C and batch-purified using affinity chromatography (Glutathione HiCap Matrix slurry; Qiagen). The purity of protein was estimated by 10% SDS–PAGE and the eluted proteins were quantified with spectrophotometry by optical absorbance at 260 nm. The eluted protein yield was concentrated and buffer exchanged using Amicon Ultra-0.5 Centrifugal Filter Devices (30K device; Millipore). The properties of the DBD fusion proteins produced and used in this study are indicated in Supplementary Table 10.

Protein binding microarray analysis. Protein binding microarray analyses were processed and analysed as described previously[42–45].

EMSAs. DNA binding of purified N-terminal GST fusions of AP2 domains of AP2-G of *P. falciparum* (PF3D7_1222600) and *P. berghei* (PBANKA_143750) to their cognate DNA sequences was analysed by EMSA. Single-stranded oligonucleotides containing the recognition motif flanked either by random nucleotides (same for all flanking sequences) or by the actual genome sequence (as they occur naturally in the 5′ upstream regions of potential AP2 target genes) and their corresponding complementary oligonucleotides were synthesized and purchased from MWG Eurofins (Germany) as labelled (5′-biotinylated and HPLC purified) and unlabelled sequences. Complementary single-stranded oligonucleotides were annealed to create double-stranded probes and used for EMSA as labelled and unlabelled target probes for the DBD of AP2G. EMSAs were performed using the LightShift Chemiluminescent EMSA kit (Pierce). In brief, 2 µg of the purified GST fusion of PfAP2-G and PbAP2-G (in separate reactions) was pre-incubated with 0.02 pmol of the labelled probe in 20 µl of the binding reaction containing binding buffer, 1 µg poly(dI-dC), 50% glycerol, 100 mM $MgCl_2$, 1% NP40 and 60 µg BSA at room temperature (22 °C) for 10 min. The unlabelled probe (4 pmol; 200-fold excess to the labelled probe) was then added as a competitor and the reaction was incubated for further 20 min at room temperature. The reaction was fractionated using 12% PAGE and transferred to a nylon membrane (Hybond) as per manufacturer's instructions. Specific binding of the AP2 domain with the target motif was detected as an upward shift using the Chemiluminescence Nucleic Acid Detection Module (Pierce), as per the manufacturer's instructions, and anti-GST antibodies.

Southern blot analysis. Southern blot analysis from wild-type line 820 and three different *pbap2-g* length-variable knockouts was performed to show successful integration of the selectable marker cassette at the desired genetic locus. In brief, approximately 10 µg of Plasmodipur (EuroProxima)-filtered and purified genomic DNA from lines 820 (wild type), G401cl1 (complete ORF knockout), G418cl6cl3 (DBD knockout) and G529cl2 (partial ORF knockout bearing the GNPm7, 8 and 9 mutations) was double-digested each with 7 µl of appropriate restriction enzyme (New England Biolabs) pairs at 37 °C for 4 h with NEB Buffer 4. For comparison with the wild-type line (820), gDNA from wild type and G401cl1, wild type and G418cl6cl3, and wild type and G529cl2 was double-digested with the High-Fidelity

versions of NcoI and SpeI, NcoI and BamHI, and EcoRI and SpeI, respectively. After transfer the membrane was hybridized (60 °C overnight) with P^{32}-labelled single-stranded DNA probe for a specific region from one of the homology arms used for generating the gene targeting vector. The probes were PCR-amplified and purified using the following oligonucleotides: GU1058 and GU1059 for G401cl1, GU1416 and GU1417 for G418cl6cl3 and GU1414 and GU1415 for G529cl2. The membrane was washed three times with decreasing concentration of SSC (3× SSC, 1× SSC, 0.5× SSC) and exposed to a maximum-resolution X-ray film (BioMax MR film; Kodak) for 35 h.

Northern blot analysis. Approximately 5 μg of RNA sample for each line (except G529cl2; which was ~2 μg) was denatured and fractionated in 1.2% agarose gel in 2.2 M (w/v) formaldehyde at 20 V overnight in 1× MOPS as running buffer. After transfer the RNA in the membrane was hybridized (60 °C overnight) with P^{32}-labelled single-stranded DNA probe for *p28* messenger RNA (PBANKA_051490; 0.62 kb ORF) and normalized using *hsp70* mRNA probe (PBANKA_071190; 2.08 kb ORF), washed and exposed to a maximum resolution X-ray film (BioMax MR film; Kodak).

Recombineering methods. Gene knockout vectors for *pbap2-g* and *pbap2-g2* were submitted to the *Plasmo*GEM database as PbGEM-072446 and PbGEM-039238, respectively[22] where details of their construction can be found. Complementation vectors were made using the Red recombination system of phage *lambda* using published protocols[46]. First, *E. coli* harbouring *P. berghei* gDNA clone PbG01-2472c01, which carries a >11-kb genomic insert including *pbap2-g* in the pJAZZ-OK linear plasmid (Lucigen), were rendered competent for recombination by transfection with plasmid pSC101gbaA[47]. A marker cassette for positive and negative selection in *E. coli*, attR1-zeo-pheS-attR2, was then amplified using primer pairs Comp143750UpR1/2 or Comp143750D1R1/2 (see Supplementary Table 11 for primer sequences). The resulting PCR products carried 50-bp extensions homologous to the upstream or downstream intergenic regions of *pbap2-g*, respectively. The PCR products were introduced into the recombination-competent *E. coli* carrying the PbG01-2472c01 library plasmid and the recombination product selected with Zeocin. The bacterial marker was then exchanged for the *P. berghei* selection marker *hdhfr-yfcu* in an *in vitro* Gateway reaction, the product of which was retransformed into *E. coli* and negatively selected on YEG-Cl and kanamycin as described[46]. Clones carrying the correct complementation plasmid were identified by PCR across the boundary of the *hdhfr-yfcu* cassette. Before transfection the constructs were linearized using NotI removing the plasmid backbone.

Reporters: construct generation. The CFP reporter construct pG0148 was generated by inserting CFP into pG073 as follows: CFP was amplified from pL1382 using primers to incorporate XhoI and SmaI restriction sites. This was cloned into the XhoI/SmaI sites of pG073 (KH unpublished) between an *hsp70* (PBANKA_071190) promoter (1.4 kb) and *p45/48* constitutive 3′ UTR. The plasmid also contains a negative-selection cassette[30] and target regions for DXO integration into a *p230p* locus downstream of the GFP/RFP cassette in the 820 line[29]. Candidates for reporter analysis in the first batch (rep 1–14) were chosen on the basis of fold downregulation in GNP versus 820 schizont, the presence of at least one predicted AP2 binding motif (GTACxC or GxGTAC or GGTACxC) and at least moderate expression levels in at least one life cycle stage. For some of the second batch of reporters based on analysis of trophozoite stage transcripts (rep 15–24) the additional criteria of not predicted to be translationally repressed was included. 2 kb of sequence immediately upstream to the predicted translational start site (PlasmoDB) was amplified by PCR using Taq polymerase and primers incorporating KpnI/XhoI restriction sites. pG0148 was digested with KpnI/XhoI to excise the *hsp70* promoter and new reporter promoters ligated in. To introduce mutations into the predicted AP2-G binding sites an overlapping PCR strategy was used to mutate the GTAC to GTAA. A primer designed around the site incorporating the mutation in both forward and reverse complement was used with the original forward and reverse primers for the 2 kb fragment in a two-stage overlapping PCR reaction. The fragment was cloned into pG0148 and sequenced to confirm the mutation. After verification of correct insert 15–30 μg of plasmid DNA was digested with SacII to linearize the integration fragment and subsequently cut with either ScaI or SapI to cut the plasmid backbone and minimise risk of introducing episomes. Fully digested DNA was ethanol precipitated and re-suspended in water before being mixed with 100 μl Nucleofector (Lonza Amaxa) solution for transfection into 820 and GNPm9 lines.

Reporters: transfection. DNA prepared as above (4–12 μg per transfection) was mixed with Nycodenz-purified synchronous *P. berghei* schizont lines 820 or GNPm9 and electroporated using programme U33 of Amaxa machine. Parasites were then immediately injected into the tail vein of a TO mouse. 24–28 h after transfection the parasites were placed on positive selection by including pyrimethamine (Sigma) in drinking water[31].

Reporters: flow cytometric analysis. Analysis was performed on parasites from tail blood on days 6–10 after transfection. 2 μl of tail blood was placed into 500 μl rich PBS (Roche) with 20 mM HEPES, 20 mM glucose, 4 mM $NaHCO_3$, 0.1% BSA) containing 1 μl Vybrant DyeCycle Ruby (Invitrogen) and incubated at 37 °C for 30 min. Parasites were pelleted and re-suspended in 1.5 ml of FACS buffer (PBS (Roche) with 2 mM HEPES, 2 mM glucose, 0.4 mM $NaHCO_3$, 0.01% BSA, 2.5 mM EDTA). Analysis was performed on a CyAn ADP 9 colour flow cytometer (Beckman Coulter) equipped with 405-nm, 488-nm and 642-nm solid-state lasers and 500,000 events were acquired (counting all events except debris). On each day an uninfected control and CFP-negative parental controls were processed in parallel with reporter lines. Data analysis was performed using Kaluza analysis software (Beckman Coulter) following the gating strategy indicated in the following schematic. For histogram analysis the CFP geometric mean expression level (AFU) in each gated population male, female and asexual was calculated as a mean from three day's data and plotted as a bar chart in excel.

All events were plotted as forward scatter (FS) versus side scatter (SS) and gate E drawn to exclude debris. Events in gate E were plotted on FS versus FS (area) and gate J(1) drawn to exclude potentially autofluorescent doublets and clumps. Events in gate J(1) were plotted FS versus Ruby (DNA stain) and gate G drawn to select infected cells. Gate G was drawn on the basis of a negative (uninfected) control population stained in the same way and analysed on the same occasion (Supplementary Fig. 13a).

Events in gate G were plotted SS versus CFP and a CFP positive gate drawn based on a non-CFP-expressing parental line (820, HP or GNP9) stained and processed on the same occasion and at similar parasitaemia. GFP versus RFP was plotted for all infected cells (events in G) and for only those falling into the CFP-positive gate. Gates drawn on female F (RFP-positive) and male M (GFP-positive) populations was used to calculate the percentage of each population that expresses CFP based on the number of cells in each gate in each plot.

For illustrative figures the infected population (G) was plotted on GFP versus RFP and those additionally falling into gate CFP-positive coloured magenta whereas those not CFP-positive were coloured grey. The percentage of the population within each gate expressing CFP (calculated as above) is indicated (Supplementary Fig. 13b).

Microscopy analysis. For some lines the CFP expression was analysed on a Zeiss Axioplan II fluorescent microscope. A drop of tail blood was stained with 5 μM Hoechst in enriched PBS for 10 min then placed on a microscope slide under a coverslip and sealed with nail varnish and visualized under a ×100 oil immersion objective, images were captured and processed using Volocity software.

Methods for promoter interruption experiments. During attempts to rescue gametocytogenesis in GNP lines by complementation rescue techniques we had observed that an interruption to the *pbap2-g* promoter slightly downstream of two GxGTAC motifs led to a loss of gametocyte production. To investigate this further a series of constructs was made to target the *pbap2-g* endogenous promoter and mutate specifically in the region of these GxGTAC motifs. Effect on gametocytogenesis after integration of these constructs into the endogenous AP2-G promoter in the fluorescent 820 parental line could then be monitored using flow cytometry.

Promoter interruption construct generation and transfection. A double-crossover homologous recombination method was used to create targeted interruptions of the *pbap2-g* endogenous promoter. The plasmid pL0035 was used, which contains a selection cassette including human *DHFR* driven by the *pbeef1aa* promoter surrounded by multiple cloning sites. Genomic fragments from the *pbap2-g* promoter region were amplified by PCR from wild type genomic DNA using Kapa Hi-Fi polymerase (KapaBiosystems) and cloned in piecewise as described below to allow for flexibility with the vector for creating multiple mutations. The 207-bp region containing the GxGTAC motifs was synthesized by MWG-Biotech with or without point mutations in the core motif. All regions are described by their distance from the *pbap2-g* gene start. A downstream integration fragment from bp-416 to bp-1,277 was cloned in using SmaI and EcoRI and an upstream integration region from bp-2,695 to bp-1,912 cloned in using HindIII and SacII. The region from bp-1,913 to bp-1,484 was cloned downstream of the selection cassette and in front of the downstream integration region using KpnI and EcoRV to create vector pG266 (2-kb deletion). Using SmaI and EcoRV the synthesized region from −1,913 to −1,484, either wild type or containing single point mutations in the G.GTAC motif, was cloned into vector pG266 to create pG298 (2-kb WT) or pG312 (2 kb MutA). Additionally a clone containing the wild-type 200-bp region in reverse orientation was selected pG299 (2-kb WT Rev). Subsequently the SmaI cloning site in pG298 was removed to created pG313 (2-kb WT-Sma). To extend the region of endogenous promoter remaining between the selection cassette and the *pbap2-g* gene an additional fragment from −2,870 to −1,913 was cloned into the KpnI site downstream of the selection cassette in pG266 and pG313 to create pG266+3 (3-kb del) and pG313+3 (3-kb WT). Constructs were linearized using HindIII and EcoRI, and approximately 10 μg of purified linear DNA was transfected in to *P. berghei* parasites (820 line) as described elsewhere.

Promoter interruption gametocytogenesis essays. Gametocyte levels in transfected parasites were monitored by flow cytometry (on a FACS CyAN, Beckman Coulter) on a drop of tail blood from animals containing the transfected parasites and maintained on pyrimethamine selection throughout from 6 days post-transfection for up to 5 consecutive days. Parasites were passaged into a clean animal maintained on pyrimethamine selection and gametocytaemia followed. As the background gametocyte levels measurable using our methods in the parental 820 line varied from ~3 to ~20% depending on parasitaemia and unknown factors, a control transfection was carried out to enable gametocyte levels to be monitored in a line that had been maintained under exactly the same conditions. This was usually the plasmid pG306, which integrated to the *p230p* locus and contains a CFP gene driven by the PBANKA_101870 promoter. This also enabled us to confirm general transfection efficiency in each batch of transfections. After gating on the infected population using DyeCycle Ruby staining, the percentage of parasites expressing RFP (female) or GFP (male) parasites was calculated. Results shown are the total gametocytaemia (male and female) as a percentage of the parasite population and a mean ± s.d. from three readings from passaged animals. The 820 parental line is a mean from four readings.

Minigene construction and analysis. pG0148 was generated as previously described in reporters section. To generate pG0157 a 2-kb fragment immediately upstream of the *pbap2-g* gene was amplified using primers to incorporate KpnI and XhoI restriction sites and cloned in place of the *hsp70* promoter in pG0148. To generate pG0189 a 300-bp fragment of *pbap2-g* was amplified to incorporate XhoI restriction sites and was cloned in frame with CFP into the XhoI restriction site between the *hsp70* promoter and the CFP gene in pG0148. To generate pG0190, CFP was amplified from pL1382 using primers to exclude the stop codon of CFP and incorporate XhoI and SmaI restriction sites. This was cloned into pG073 to generate pG0188 (not shown). A 900-bp C-terminal fragment of *pbap2-g* incorporating the DBD was amplified from gDNA using primers to incorporate SmaI restriction sites and cloned into the SmaI restriction site downstream of and in-frame with CFP in pG0188. To generate pG0191 the *pbap2-g* promoter and first 300 bp of coding sequence were amplified using primers incorporating KpnI and XhoI restriction sites and was cloned in place of the *hsp70* promoter in pG0190. Plasmids were sequenced and 5–10 μg of linearized purified DNA transfected into either 820 or GNPm9 lines as previously described for reporter genes. Resulting transfected parasites were analysed by flow cytometry and fluorescence microscopy for expression and localization of CFP signal. Each experiment was performed independently three times.

Competitive growth assays. GNPm9M1Cl1 was transfected with construct pG0148 to constitutively express CFP from an *hsp70* promoter to generate line GNP-CFP. An analogous construct with RFP driven by the *hsp70* promoter was generated (pG0161) and transfected into wild-type (HP) producer line to generate WT-RFP. Also generated was a wild-type (HP) producer line expressing CFP from construct pG0148 (WT-CFP). Each line was individually grown in a TO mouse under pyrimethamine selection. 2 μl tail blood from each mouse was stained with Vybrant Dyecycle Ruby (Invitrogen) to label infected red blood cells and then run on a CyAn ADP 9 Colour flow cytometer (Beckman coulter). After gating on infected cells the CFP or RFP expression was analysed showing that nearly 100% of each population after gating for infected cells expressed the fluorescent marker. Parasites were mixed to create a 50:50 mix of parasites containing either WT-CFP and WT-RFP or GNP-CFP and WT-RFP. These were injected intravenously into mice. Parasites were monitored daily by flow cytometry and after gating for infected cells the percentage of the population expressing either RFP (gate AF −+), CFP (gate AF +−) or both (gate AF++) reflecting mixed-multiply infected cells was calculated and plotted. On day 6, blood from each mouse was passaged into a new host and the time course continued. After day 11 parasites were cryopreserved. For the competition assays between the *pbap2-g* KO1, *pbap2-g2* KO and *p28* KO, the *Plasmo*GEM knockout vectors were transfected into the GFP- and mCherry-expressing parasites. Once the parasitaemia in transfected animals reached ~5%, they were used to generate an inoculum containing an equal proportion of red and green parasites. Accuracy of each inoculum was tested using flow cytometry. New mice were injected (1×10^5 parasites per animal) and kept under continued pyrimethamine treatment to prevent the emergence of untransfected parasites. The proportion of red and green parasites in the mixture was followed daily using flow cytometry. Three infected mice were used for each comparison.

Microarray methods. A 8×15k custom microarray (Agilent) providing coverage of the *P. berghei* genome at >1 probe per kb of coding sequence was used[23]. Samples were prepared from parasites maintained using standard parasitological procedures. For schizont cultures parasites were obtained from cardiac puncture and grown overnight in culture. For ring-stage cultures parasites were matured *in vitro* to schizont stage in order to synchronise the population, then injected into a new host and allowed to reinvade. Blood was collected at 24 + 6 h post infection and filtered through a magnetic column (variomacsD) to deplete of mature stages and gametocytes. For trophozoite-stage parasites, parasites were prepared as for ring stages were then cultured for a further 6 h. All samples were filtered through a Plasmodipur filter to remove mouse leucocyte contamination before RNA preparation using a standard TRIzol method. Samples were processed for microarray using methods as described[23]. For GNP and *pbap2-g* KO1 a two-colour microarray hybridization was performed with a background pool of complementary DNA made from material from all life cycle stages (except late mosquito and liver stages). Parental control lines and experimental samples were then hybridized with the same background pool sample for all experiments. For *pbap2-g* KO2 and *pbap2-g* KO, the mutant samples were hybridized against the equivalent samples from the parental line and against each other. Arrays were scanned on an Agilent Microarray Scanner. Normalized intensities were then extracted using Agilent Feature Extractor. All expression data are available from the Gene Expression Omnibus database (http//:www.ncbi.nih.gov/geo) under the accession numbers GSE52859 and GSE53246.

Statistical methods for microarrays. Three biological replicates were performed for each life cycle stage of *pbap2-g* KO1 line and the 820 parental line. Naturally derived GNP line (schizonts only) microarray results are representative of two technical replicates each from three independently derived GNP lines. These technical replicates were performed in different laboratories using the same methods. The *pbap2-g* KO1 and GNP microarray data was uploaded to PUMADB (http://puma.princeton.edu/) for further processing. The data was extracted as a $\log_2$ of the fold change of red (sample) versus green (common pool) with minimal filtering to exclude background signal and median centred. The fold change between the GNP sample and the 820 parental line was calculated for each transcript and the mean and standard deviation of the replicates calculated (using Microsoft Excel). The distribution of these samples was confirmed to be normal ($P < 2.2 \times 10^{-16}$, Kolmogorov–Smirnov test in R version 2.10), and the transcripts classed as down regulated in GNP lines were those 2 s.d. below the mean fold change. For plotting volcano plots (Fig. 3) a two-tailed *t*-test was performed on the independent replicates and a $-\log_{10}$ transform of this result plotted. This was plotted against the $\log_2$ fold change using R ggplot2 library. To determine which transcripts were gametocyte-specific the fold change between three replicates of gametocyte-stage wild-type parasites was compared to three replicates of schizont-stage wild-type parasites. A one-tailed *t*-test was then used to determine those upregulated in gametocytes as highlighted in volcano plots in Fig. 3. For *pbap2-g* KO2 and *pbap2-g2* the biological triplicates of each of the hybridizations (both mutants against the wild type and against each other) were processed using the R version 2.15.0 software[48] with limma package[49]. The data was background-corrected and normalized between the arrays (LOESS normalization). Fold changes between the strains and *P* values for differential expression were calculated with a linear statistical model. The *P* values from all experiments were adjusted using the false discovery rate correction.

For the gametocyte expression rank (Fig. 3a and Supplementary Table 6) the absolute intensity values from microarrays from three independent replicates of wild-type gametocytes was used and ranked from highest (1) to lowest (~4,553) expression rank. To test for the deregulation of the gametocyte-specific genes in all the strains, the enrichment in gametocyte-specific genes (expression rank 1 to 500) in the top 500 genes showing the highest fold change in each of the mutants was tested using the Fisher's exact test. Comparisons of the variances of the microarray data were carried out in R and all the variances were similar; none of the samples were significantly different ($P < 10 \times 10^{-16}$, *F*-test). Microarray data has been submitted to the GEO database (accession numbers: GSE52859 and GSE53246).

Search for DNA-binding motifs. The genomic sequences for all *P. berghei* genes were identified using PlasmoDB (version 9.1) and defined as a 2-kb region upstream of the transcription start site to the first base of the transcription start site (4,803 entries). A file was also created for the gametocyte-specific genes (452 entries). Differences in usable entries were due to genes close to the ends of chromosomes or poorly assembled regions, and regions that overlapped other genes. A custom Perl script was used to count occurrences of the PbAP2-G and PbAP2-G2 motifs in the sequences using a regular expression (PbAP2-G was defined as /GxGTAC|GTACxC/ and PbAP2-G2 was defined by orthology as /TGCxACC|GGTxGCA/; ref. 6) The script counts the occurrence of each pattern per-region and also provides a total number of sequences that contain at least one occurrence, and is available on request. Hypergeometric *P* values were calculated interactively using R version 2.10.

24. Dearsly, A. L., Sinden, R. E. & Self, I. A. Sexual development in malarial parasites: gametocyte production, fertility and infectivity to the mosquito vector. *Parasitology* **100,** 359–368 (1990).
25. Franke-Fayard, B. *et al.* A *Plasmodium berghei* reference line that constitutively expresses GFP at a high level throughout the complete life cycle. *Mol. Biochem. Parasitol.* **137,** 23–33 (2004).
26. Khan, S. M. *et al.* Proteome analysis of separated male and female gametocytes reveals novel sex specific *Plasmodium* biology. *Cell* **121,** 675–687 (2005).
27. van Dijk, M. R. *et al.* A central role for P48/45 in malaria parasite male gamete fertility. *Cell* **104,** 153–164 (2001).

28. Mair, G. R. *et al.* Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. *PLoS Pathog.* **6,** e1000767 (2010).
29. Ponzi, M. *et al.* Egress of *Plasmodium berghei* gametes from their host erythrocyte is mediated by the MDV-1/PEG3 protein. *Cell. Microbiol.* **11,** 1272 (2009).
30. Orr, R. Y., Philip, N. & Waters, A. P. Improved negative selection protocol for *Plasmodium berghei* in the rodent malarial model. *Malar. J.* **11,** 103 (2012).
31. Janse, C. J., Ramesar, J. & Waters, A. P. High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite *Plasmodium berghei. Nature Protocols* **1,** 346–356 (2006).
32. Sinden, R. *Molecular Biology of Insect Diseases Vectors: a Methods Manual* (eds Crampton J. M., Beard, C. B. & Louis, C.) (Chapman and Hall, 1997).
33. Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina M-Seq sequensors. *BMC Genomics* **13,** 341 (2012).
34. Zerbino, D. R. & Birney, E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).
35. Swain, M. T. *et al.* A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes. *Nature Protocols* **7,** 1260–1284 (2012).
36. Assefa, S. *et al*. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* **25,** 1968–1969 (2009).
37. Tsai, I. J., Otto, T. D. & Berriman, M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* **11,** R41 (2010).
38. Otto, T. D. *et al.* Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* **26,** 1704–1707 (2010).
39. Otto, T. D., Dillon, G. P., Degrave, W. S. & Berriman, M. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res.* **39,** e57 (2011).
40. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
41. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
42. Campbell, T. L., De Silva, E. K., Olszewski, K. L., Elemento, O. & Llinás, M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* **6,** e1001165 (2010).
43. Berger, M. F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnol.* **24,** 1429–1435 (2006).
44. Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols* **4,** 393–411 (2009).
45. Workman, C. T. *et al.* enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res.* **33,** W389–392 (2005).
46. Pfander, C., Anar, B., Brochet, M., Rayner, J. C. & Billker, O. Recombination-mediated genetic engineering of *Plasmodium berghei* DNA. *Methods Mol. Biol.* **923,** 127–138 (2013).
47. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genet.* **20,** 123–128 (1998).
48. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (2012).
49. Smyth, G. K. in: *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Gentleman, R., Carey, V., Dudoit, S., Irizarry, R. & Huber, W.) 397–420 (Springer, 2005).

LETTER

doi:10.1038/nature12971

DNA-guided DNA interference by a prokaryotic Argonaute

Daan C. Swarts[1]*, Matthijs M. Jore[1]*, Edze R. Westra[1], Yifan Zhu[1], Jorijn H. Janssen[1], Ambrosius P. Snijders[2], Yanli Wang[3], Dinshaw J. Patel[4], José Berenguer[5], Stan J. J. Brouns[1] & John van der Oost[1]

RNA interference is widely distributed in eukaryotes and has a variety of functions, including antiviral defence and gene regulation[1,2]. All RNA interference pathways use small single-stranded RNA (ssRNA) molecules that guide proteins of the Argonaute (Ago) family to complementary ssRNA targets: RNA-guided RNA interference[1,2]. The role of prokaryotic Ago variants has remained elusive, although bioinformatics analysis has suggested their involvement in host defence[3]. Here we demonstrate that Ago of the bacterium *Thermus thermophilus* (TtAgo) acts as a barrier for the uptake and propagation of foreign DNA. *In vivo,* TtAgo is loaded with 5′-phosphorylated DNA guides, 13–25 nucleotides in length, that are mostly plasmid derived and have a strong bias for a 5′-end deoxycytidine. These small interfering DNAs guide TtAgo to cleave complementary DNA strands. Hence, despite structural homology to its eukaryotic counterparts, TtAgo functions in host defence by DNA-guided DNA interference.

To elucidate the physiological role of Ago in prokaryotes, we studied Ago from *T. thermophilus*. Comparison of the *ago* genes of the type strain HB27 (refs 4, 5) and a derivative with enhanced competence (HB27^{EC}; Fig. 1a and Extended Data Fig. 1a), revealed that an insertion sequence (IS*Tth7*)[6] disrupts *ago* in HB27^{EC}. In line with a role of TtAgo in reducing competence, a generated Δ*ago* mutant (HB27Δ*ago*; Fig. 1a) has a natural transformation efficiency that is a factor of ten higher than the wild-type HB27 ($P < 0.02$, Fig. 1b). Complementation of the knockout strain with *ago* (HB27Δ*ago*::^{s}*ago* (HB27Δ*ago* complemented with a *strep(II)-tag-ago* gene fusion insert); Fig. 1a, b) almost completely restores the wild-type phenotype. Moreover, isolation of plasmid and total DNA from the wild-type and the *ago* knockout strains revealed lower plasmid yields from the wild-type strain, indicating that TtAgo reduces the intracellular plasmid concentration ($P < 0.02$, Fig. 1c; $P < 0.02$, Fig. 1d).

We performed transcriptome analysis of HB27 and HB27Δ*ago* to determine whether TtAgo-mediated interference proceeds directly by targeting plasmid DNA, or indirectly by regulating gene expression. Although the comparison revealed pleiotropic changes in gene expression (Extended Data Fig. 2), we did not observe substantial differential expression of genes involved in plasmid uptake or host defence (Extended Data Table 1). Hence, RNA sequencing (RNA-seq) analysis suggests that TtAgo does not influence plasmid uptake and plasmid copy number at the level of transcriptional control.

We therefore studied whether TtAgo interacts with plasmid DNA. In agreement with the RNA-seq analysis (Extended Data Fig. 2), affinity-purified TtAgo expressed from the chromosome of HB27Δ*ago*::^{s}*ago* could be detected by protein mass spectrometry (Extended Data Table 2). Unfortunately, molecular analysis of TtAgo expressed in *T. thermophilus* was hampered by the low TtAgo yield, and attempts to overexpress TtAgo in *T. thermophilus* from a plasmid were unsuccessful. By contrast, expression of Strep(II)-tagged TtAgo (Fig. 2a) in *Escherichia coli* was successful when performed at 20 °C. Under these conditions, TtAgo has no effect on plasmid content (Extended Data Fig. 1b). Analysis of co-purified nucleic acids revealed that TtAgo-associated RNA (10–150 nucleotides) is preferentially ^{32}P-labelled in a polynucleotide kinase

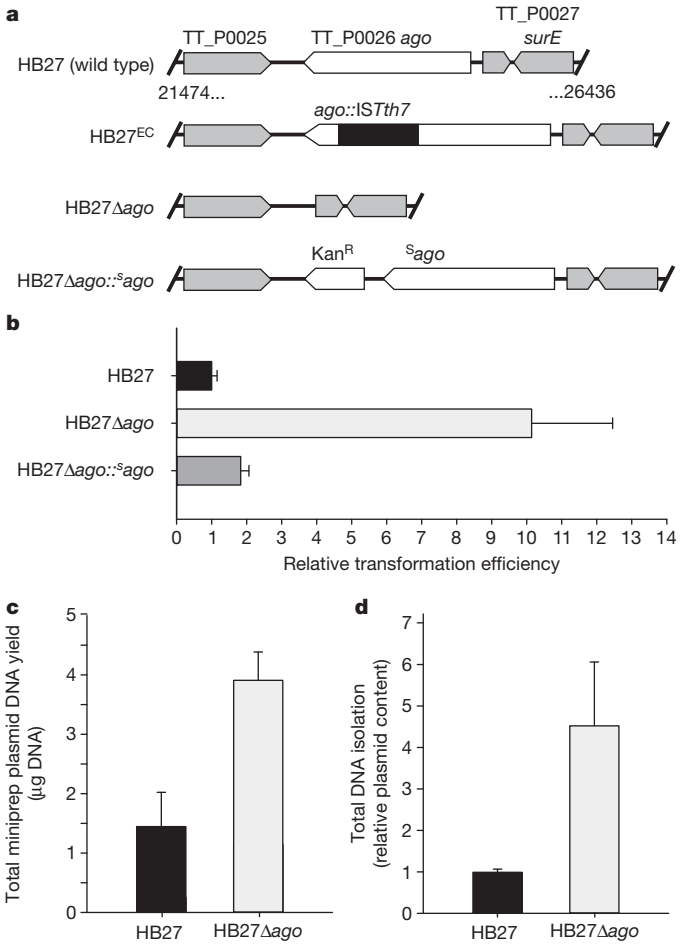


Figure 1 | TtAgo interferes with plasmid DNA. a, Overview of *ago* gene loci of *T. thermophilus* strains: HB27 (wild type), HB27^{EC} (spontaneous derivative with enhanced competence), HB27Δ*ago* (knockout), and HB27Δ*ago*::^{s}*ago* (HB27Δ*ago* complemented with a *strep(II)-tag-ago* gene fusion insert). Kan^{R}, kanamycin resistance marker. **b**, Transformation efficiency of *T. thermophilus* strains on transformation with the plasmid pMHPnqosGFP (Extended Data Table 5). Error bars indicate standard deviations of biological duplicates. **c**, Yield of pMHPnqosGFP plasmid mini preparation (miniprep) of HB27 and HB27Δ*ago*. Error bars indicate standard deviations of biological triplicates. **d**, Plasmid content of total DNA purified from HB27Δ*ago* relative to that from HB27, as quantified by Genetools (Syngene) after resolving the DNA on a 0.8% agarose gel. Error bars indicate standard deviations of biological triplicates.

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, the Netherlands. [2]Clare Hall Laboratories, Cancer Research UK, London Research Institute, South Mimms EN6 3LD, UK. [3]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. [4]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [5]Centro de Biología Molecular Severo Ochoa, UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain.
*These authors contributed equally to this work.

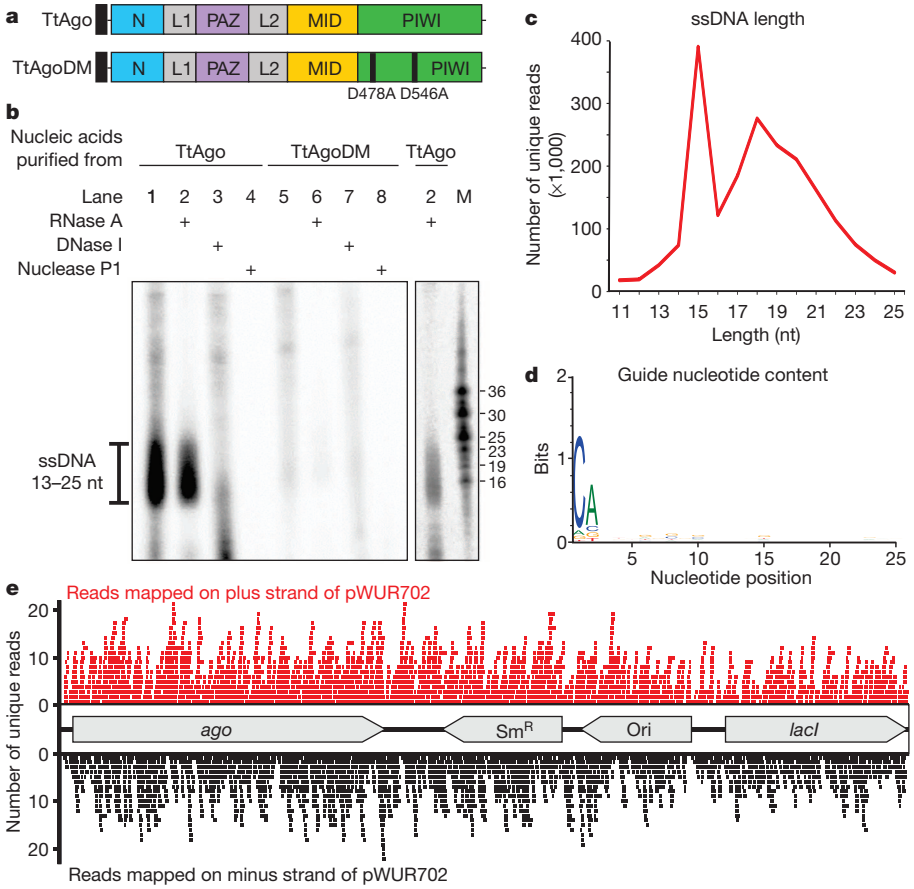


Figure 2 | TtAgo guides are 5′-phosphorylated DNA molecules. **a**, Schematic representation of TtAgo and TtAgoDM proteins used for all experiments (N, PAZ, MID, and PIWI are structural domains, L1 and L2 are linkers[8]). The amino-terminal Strep(II)-tag is indicated as a black square. **b**, Co-purified nucleic acids from TtAgo and TtAgoDM are labelled with [γ-^{32}P]ATP after phosphate exchange by PNK from bacteriophage T4, and treated with enzymes as indicated. M, custom ssDNA marker; nt, nucleotides. **c**, Length distribution of unique ssDNA sequences co-purified with TtAgo. **d**, Nucleotide composition of unique ssDNA sequences co-purified with TtAgo. **e**, Unique reads of TtAgo co-purified ssDNA molecules mapped on the TtAgo expression vector pWUR702.

(PNK) forward reaction, indicating the presence of 5′ hydroxyl groups (Extended Data Fig. 1c). By contrast, co-purified DNA has a more defined length (13–25 nucleotides), and is preferentially labelled in a PNK exchange reaction, indicating phosphorylated 5′ ends (Fig. 2b). A 5′ phosphate group is a general feature of Ago guides[7–11].

Whereas eukaryotic Ago proteins exclusively use ssRNA guides, some prokaryotic Ago proteins have a higher affinity for single-stranded DNA (ssDNA) guides[9,10]. Moreover, the characteristics of the small DNAs that associate with TtAgo *in vivo* are in agreement with previously described *in vitro* guide requirements[8,12,13]. TtAgo catalyses cleavage of ssDNA targets *in vitro* when supplied with complementary 5′-phosphorylated 21-nucleotide ssDNA guides, but not when supplied with analogous ssRNA guides[8,12,13] (Extended Data Fig. 3). During isolation of an active site double mutant, TtAgoDM (TtAgo(D478A,D546A); Fig. 2a), only RNAs co-purify (10–150 nucleotides; Extended Data Fig. 1c). This suggests that active site residues are involved in processing and/or binding of the ssDNA molecules.

Cloning and sequencing of TtAgo-bound DNA molecules resulted in 70.6 million sequences, of which 65% can be mapped on the TtAgo expression plasmid pWUR702, 3% on the plasmid pRARE, and 32% on the chromosome of *E. coli* K12 (Extended Data Table 3). Remarkably, when normalized for the DNA content in each cell, TtAgo predominantly co-purifies with guides complementary to pWUR702 and pRARE (approximately 54 and 8.8 times more frequently, respectively), rather than with guides complementary to the *E. coli* K12 chromosome (Extended Data Table 3).

More detailed analysis of unique guide sequences revealed two populations of DNA guides: one 15-nucleotides long, and the other ranging from 13 to 25 nucleotides in length (Fig. 2c). No obvious bias towards specific regions of the plasmids or the chromosome was detected: the guides target coding and non-coding regions on both strands independent of GC content (Fig. 2e). Some guides map on one of the plasmids as well as on the chromosome of *E. coli* (for example, on *lacI* and *proL*). The fact that these guides do not seem to be under-represented compared with other plasmid-targeting guides indicates that there is no selection against chromosome-targeting guides, but rather that the differential guide loading (Extended Data Table 3) is a result of preferential acquisition of guides from plasmids.

Interestingly, 89% of the DNA guides have a deoxycytidine (dC) at the first position at the 5′ end and 72% have a deoxyadenosine (dA) at the second position (Fig. 2d). Despite this bias, identical TtAgo cleavage activities are observed with DNA guides containing a 5′ dC, dT, dA or dG (Extended Data Fig. 4a–d). The 5′ dC preference may result from specific guide processing, or from preferential 5′ nucleoside selection by TtAgo. A bias for specific 5′ nucleosides also occurs in certain eukaryotic Ago proteins[14,15].

We performed activity assays to investigate whether the *in vivo* plasmid-derived ssDNAs are functional guides that enable TtAgo to cleave double-stranded DNA (dsDNA) targets (expression plasmid pWUR702). Purified TtAgo linearizes or nicks pWUR702, resulting in linear or open circular plasmid DNA, respectively (Fig. 3a, lane 4), whereas TtAgoDM does not show this activity (Fig. 3a, lane 3). The cleavage activity of TtAgo is strongly temperature dependent: whereas ssDNA is cleaved at temperatures ≥20 °C, plasmid DNA is only cleaved at temperatures ≥65 °C (Extended Data Fig. 4e, f). This agrees with the observation that during TtAgo expression in *E. coli* at 20 °C, plasmid concentrations are not decreased (Extended Data Fig. 1b). Purified TtAgo is unable to cleave plasmids that have no sequence similarity to pWUR702 or pRARE (for example, pWUR708; Fig. 3b, lane 4). However, when supplied with two synthetic 5′-phosphorylated ssDNA guides that target both strands of the plasmid at the same locus (Fig. 4b), TtAgo was able to linearize or nick pWUR708 (Fig. 3b, lane 8). These findings, together with the guide sequence data, indicate that the *in vivo* acquired DNA molecules guide TtAgo to cleave dsDNA targets. We

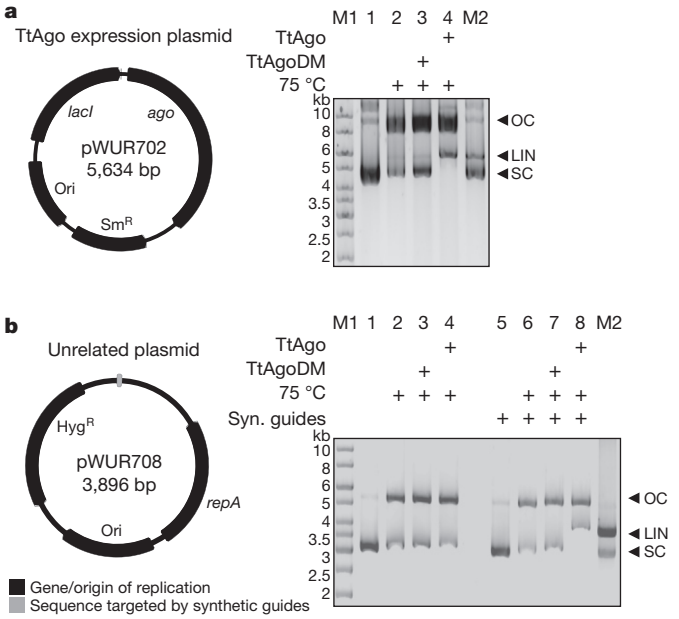


Figure 3 | TtAgo cleaves plasmids complementary to its guides. **a**, **b**, Untreated target plasmid (lane 1, 5), plasmid incubated at 75 °C in the absence of proteins (lane 2, 6), or in the presence of TtAgoDM (lane 3, 7) or TtAgo (lane 4, 8) purified from *E. coli*, resolved on 0.8% agarose gels. LIN, linear; M1, 1 kb Generuler marker (Fermentas); M2, linearized and untreated target plasmid; OC, open circular; SC, supercoiled plasmid. **a**, TtAgo expression vector pWUR702. **b**, Target plasmid pWUR708, which shares no sequence identity with expression vector pWUR702 or pRARE. Additionally, synthetic (Syn.) ssDNA guides were added to the reactions with pWUR708 (lane 5–8).

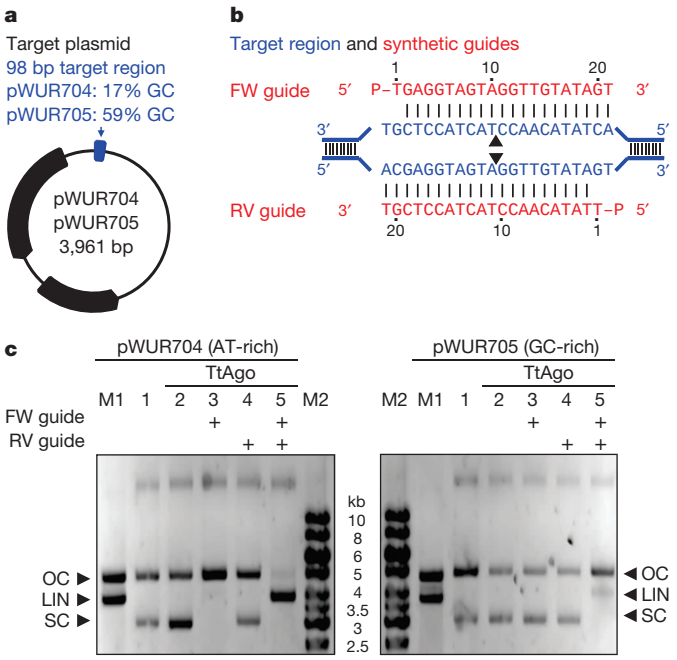


Figure 4 | TtAgo cleaves plasmids by nicking two strands. **a**, Plasmids pWUR704 and pWUR705 contain a 98 bp target region with a GC content of 17% or 59%, respectively, as indicated in blue (for details, see Extended Data Fig. 5a, b). **b**, Part of the pWUR704 and pWUR705 target site (indicated in blue) and complementary ssDNA guides used in this experiment (indicated in red). Black triangles indicate predicted cleavage sites. **c**, 0.8% agarose gels loaded with pWUR704 and pWUR705 plasmids that were incubated without proteins (lane 1), or with TtAgo (lane 2), TtAgo–forward (FW) guide complex (lane 3), TtAgo–reverse (RV) guide complex (lane 4), or TtAgo–FW and TtAgo–RV guide complexes. LIN, linear; M1, open circular and linear pWUR704 or pWUR705; M2, 1 kb Generuler marker (Fermentas); OC, open circular; SC, supercoiled plasmid.

propose to refer to these guides of TtAgo as small interfering DNAs (siDNAs).

To gain insight into the molecular mechanism of dsDNA cleavage by TtAgo, we performed additional *in vitro* plasmid cleavage assays using purified TtAgo loaded with synthetic siDNAs. Negatively supercoiled plasmids (isolated from *E. coli*) were used, because at least 95% of all plasmids isolated from *T. thermophilus* have a negatively supercoiled topology[16,17]. Negative supercoiling facilitates melting of the DNA duplex, especially at elevated temperatures[18–20]. Target plasmids pWUR704 and pWUR705 are identical except for the flanking regions of the target site (AT-rich or GC-rich; Fig. 4a). Both plasmids share no sequence similarity with TtAgo expression plasmid pWUR702, and they are not cleaved by TtAgo unless complementary siDNAs are added (Fig. 4c). When supplied with a single 21-nucleotide siDNA, TtAgo nicks the negatively supercoiled plasmid (Fig. 4c, lanes 3, 4), and when supplied with a mixture of two 21-nucleotide siDNAs that target both DNA strands at the same locus, TtAgo linearizes the plasmid (Fig. 4b, c, lane 5). Both nicking and dsDNA cleavage are more efficient when the target sequence is flanked by AT-rich regions (Fig. 4a, c and Extended Data Fig. 5a, b). Interestingly, the same TtAgo–siDNA complexes are not able to cleave linearized plasmids (Extended Data Fig. 5c, d). This suggests that cleavage of dsDNA by TtAgo depends on the negatively supercoiled topology of the target DNA.

Subsequent analysis revealed that the TtAgo–siDNA complex is able to linearize a relaxed, nicked plasmid if its target site is directly opposite the first nick (Extended Data Fig. 5e). If the nicked site is located further away (33 bp) from the target site, linearization of the nicked plasmid occurs only if the target region is AT-rich (Extended Data Fig. 5f, g). Thus, although the negatively supercoiled topology of the plasmid is lost after the primary nick, the nick facilitates local melting of the dsDNA (especially in AT-rich DNA), which allows TtAgo–siDNA complexes to nick the second strand, resulting in a dsDNA break. Like eukaryotic Ago proteins[21], the TtAgo–siDNA complex cleaves a phosphate ester bond between the target nucleotides that base pair with guide nucleotides 10 and 11 (ref. 22). Sequence analysis of a cleaved dsDNA target (Extended Data Fig. 5h) demonstrated that dsDNA breaks also result from nicking both strands at the canonical Ago cleavage site.

While this manuscript was under revision, a characterization of a prokaryotic Ago protein from *Rhodobacter sphaeroides* (RsAgo) was published[23]. Despite similarities in the overall domain architecture of TtAgo and RsAgo, there are major functional differences between these proteins. RsAgo acquires mRNA-derived RNA guides with a 5′ uridine (U), whereas TtAgo acquires DNA guides with a 5′ dC. In both proteins, guides complementary to plasmids are over-represented. However, RsAgo lacks a functional catalytic site and functions by target-binding alone. TtAgo, on the other hand, harbours a functional catalytic site allowing cleavage of both single- and double-stranded targets.

On the basis of our findings, we propose a model for DNA interference by TtAgo. On the entry of plasmid DNA into the cell, TtAgo acquires siDNA guides (13–25 nucleotides in length) from the invader. Although the mechanism of guide acquisition by TtAgo is unknown, the requirement of an intact catalytic site suggests involvement of the nuclease itself. TtAgo is loaded with siDNAs that are preferentially derived from plasmids; as such, single guides may allow for neutralization of multi-copy invaders. Combining our *in vivo* and *in vitro* data, we speculate that TtAgo uses siDNA guides to specifically cleave ssDNA targets, such as DNA taken up by the natural competence system[5] or replication intermediates. The siDNA–TtAgo complex also targets negatively supercoiled dsDNA, which results in plasmid nicking. Especially in the case of plasmid DNA, single-strand breaks will result in loss of the supercoiled topology and, as such, in decreased transcription levels[24]. Furthermore, if the nick site is located in an AT-rich region, TtAgo loaded with an siDNA that targets the opposite strand may generate

a dsDNA break, potentially leading to degradation of the plasmid by other nucleases. The observation that invading DNA elements generally have a lower GC content than their hosts[25] may explain self/non-self discrimination by TtAgo. Whereas the eukaryotic Ago protein is a key component of sophisticated multi-enzyme systems for RNA-guided RNA interference, we reveal the biochemical activity and functional importance of an evolutionarily related enzyme in prokaryotes that protects its host against mobile genetic elements through DNA-guided DNA interference.

METHODS SUMMARY

T. thermophilus HB27, HB27^{EC}, and two derivatives of the HB27 strain, HB27Δ*ago* and HB27Δ*ago*::^{s}*ago*, were used for plasmid transformation experiments. Plasmid pMHPnqosGFP was isolated from HB27 and HB27Δ*ago*. RNA for RNA-seq analysis was purified from HB27 and HB27Δ*ago* during mid-log-phase growth. Strep(II)-tagged TtAgo was heterologously produced from a plasmid in *E. coli* KRX (Promega) and purified by affinity purification before analyses of co-purified nucleic acids. Guides co-purified with TtAgo or synthetic guides were used in *in vitro* TtAgo cleavage assays using synthetic ssDNA or dsDNA plasmid as targets.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 July; accepted 19 December 2013.

Published online 16 February 2014.

1. Ketting, R. F. microRNA biogenesis and function: an overview. *Adv. Exp. Med. Biol.* **700,** 1–14 (2011).
2. Joshua-Tor, L. & Hannon, G. J. Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb. Perspect. Biol.* **3,** a003772 (2011).
3. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4,** 29 (2009).
4. Koyama, Y., Hoshino, T., Tomizuka, N. & Furukawa, K. Genetic transformation of the extreme thermophile *Thermus thermophilus* and of other *Thermus* spp. *J. Bacteriol.* **166,** 338–340 (1986).
5. Averhoff, B. Shuffling genes around in hot environments: the unique DNA transporter of *Thermus thermophilus*. *FEMS Microbiol. Rev.* **33,** 611–626 (2009).
6. Gregory, S. T. & Dahlberg, A. E. Transposition of an insertion sequence, IS*Tth7*, in the genome of the extreme thermophile *Thermus thermophilus* HB8. *FEMS Microbiol. Lett.* **289,** 187–192 (2008).
7. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305,** 1437–1441 (2004).
8. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456,** 209–213 (2008).
9. Yuan, Y. R. *et al.* Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* **19,** 405–419 (2005).
10. Ma, J. B. *et al.* Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434,** 666–670 (2005).
11. Nakanishi, K., Weinberg, D. E., Bartel, D. P. & Patel, D. J. Structure of yeast Argonaute with guide RNA. *Nature* **486,** 368–374 (2012).
12. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456,** 921–926 (2008).
13. Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461,** 754–761 (2009).
14. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465,** 818–822 (2010).
15. Frank, F., Hauver, J., Sonenberg, N. & Nagar, B. *Arabidopsis* Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *EMBO J.* **31,** 3588–3595 (2012).
16. Collin, R. G., Morgan, H. W., Musgrave, D. R. & Daniel, R. M. Distribution of reverse gyrase in representative species of eubacteria and archaebacteria. *FEMS Microbiol. Lett.* **55,** 235–240 (1988).
17. Charbonnier, F. & Forterre, P. Comparison of plasmid DNA topology among mesophilic and thermophilic eubacteria and archaebacteria. *J. Bacteriol.* **176,** 1251–1259 (1994).
18. Duguet, M. The helical repeat of DNA at high temperature. *Nucleic Acids Res.* **21,** 463–468 (1993).
19. Westra, E. R. *et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46,** 595–605 (2012).
20. Bates, A. D. & Maxwell, A. *DNA Topology* 2nd edn (Oxford Univ. Press, 2005).
21. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20,** 6877–6888 (2001).
22. Sheng, G. *et al.* Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl Acad. Sci. USA* **111,** 652–657 (2013).
23. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial Argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell* **51,** 594–605 (2013).
24. Travers, A. & Muskhelishvili, G. DNA supercoiling—a global transcriptional regulator for enterobacterial growth? *Nature Rev. Microbiol.* **3,** 157–169 (2005).
25. Rocha, E. P. C. & Danchin, A. Base composition bias might result from competition for metabolic resources. *Trends Genet.* **18,** 291–294 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We want to thank A. Hidalgo, C. E. César, M. Davids and R. H. J. Staals for advice on experimental procedures. Furthermore, we would like to thank R. Engelhart, B. van Genugten, G. Göertz and R. Stolk for experimental contributions. This work was financially supported by grants from the Netherlands Organization of Scientific Research (NWO) to J.O. (NWO-TOP, 854.10.003), and to S.J.J.B. (NWO Vidi , 864.11.005), and by project BIO2010-18875 from the Spanish Ministry of Science and Innovation, and an Institutional Grant from the Fundación Ramón Areces to CBMSO (J.B.).

Author Contributions M.M.J. and J.H.J. made genomic *T. thermophilus* mutants under the supervision of J.v.d.O. *T. thermophilus* experiments were performed by D.C.S., M.M.J. and J.H.J. under the supervision of J.B., S.J.J.B. and J.v.d.O. D.C.S. and E.R.W. purified RNA for RNA-seq, and D.C.S. analysed RNA-seq data under the supervision of S.J.J.B. and J.v.d.O. D.C.S. and A.P.S. performed experiments in which TtAgo expression in *T. thermophilus* was shown using mass spectrometry. D.C.S., M.M.J. and J.H.J. made all plasmid constructs under the supervision of S.J.J.B., J.B. and J.v.d.O. D.C.S., E.R.W. and Y.Z. purified and analysed TtAgo guides. *In vitro* activity assays were designed and analysed by D.C.S., S.J.J.B., Y.W., D.J.P. and J.v.d.O., and performed by D.C.S. and Y.Z. under the supervision of S.J.J.B. and J.v.d.O. All authors read and approved the submitted manuscript.

Author Information The RNA-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus under accession number GSE52738. The siDNA sequence data discussed in this publication have been deposited in NCBI s BioSample database and are accessible under accession number SAMN02593821. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.v.d.O. (john.vanderoost@wur.nl).

CORRIGENDUM

doi:10.1038/nature13125

Corrigendum: Crystal structure of a nitrate/nitrite exchanger

Hongjin Zheng, Goragot Wisedchaisri & Tamir Gonen

Nature **497,** 647–651 (2013); doi:10.1038/nature12139

While revising our manuscript prior to formal acceptance, we accidentally removed two references from the main text. The footnote to Table 1 of this Letter should cite ref. 1 (not ref. 24) for footnote symbol † and ref. 2 (not ref. 21) for footnote symbol ‡. We apologise for any inconvenience.

1. Unkles, S. E. *et al.* Two perfectly conserved arginine residues are required for substrate binding in a high-affinity nitrate transporter. *Proc. Natl Acad. Sci. USA* **101,** 17549–17554 (2004).
2. Unkles, S. E. *et al.* Alanine scanning mutagenesis of a high-affinity nitrate transporter highlights the requirement for glycine and asparagine residues in the two nitrate signature motifs. *Biochem. J.* **447,** 35–42 (2012).

CORRIGENDUM

doi:10.1038/nature13126

Corrigendum: An environmental bacterial taxon with a large and distinct metabolic repertoire

Micheal C. Wilson, Tetsushi Mori, Christian Rückert, Agustinus R. Uria, Maximilian J. Helf, Kentaro Takada, Christine Gernert, Ursula A. E. Steffens, Nina Heycke, Susanne Schmitt, Christian Rinke, Eric J. N. Helfrich, Alexander O. Brachmann, Cristian Gurgui, Toshiyuki Wakimoto, Matthias Kracht, Max Crüsemann, Ute Hentschel, Ikuro Abe, Shigeki Matsunaga, Jörn Kalinowski, Haruko Takeyama & Jörn Piel

Nature **506,** 58–62 (2014); doi:10.1038/nature12959

One of the accession numbers for this Article was listed as AZHXW01000000 instead of AZHX01000000. It has been corrected in the online versions of the paper.

CORRIGENDUM

doi:10.1038/nature13143

Corrigendum: Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs

Maurilio Sampaolesi, Stephane Blot, Giuseppe D'Antona, Nicolas Granger, Rossana Tonlorenzi, Anna Innocenzi, Paolo Mognol, Jean-Lauren Thibaud, Beatriz G. Galvez, Ines Barthélémy, Laura Perani, Sara Mantero, Maria Guttinger, Orietta Pansarasa, Chiara Rinaldi, M. Gabriella Cusella De Angelis, Yvan Torrente, Claudio Bordignon, Roberto Bottinelli & Giulio Cossu

Nature **444,** 574–579 (2006), doi:10.1038/nature05282 and corrigendum *Nature* **494,** 506 (2013); doi:10.1038/nature11976

In Fig. 4b of this Article, the gel for the loading control MyHC for the dog Varus was an unintentional duplication of the loading controls for the dog Vampire (which is correct). The correct gel is shown below in Fig. 1. The error does not affect any of our results and a University College London (UCL) committee appointed by the Vice Provost for Research has investigated the data presented in this Article and is satisfied of its authenticity.

Correspondence should be addressed to G.C. (giulio.cossu@manchester.ac.uk).

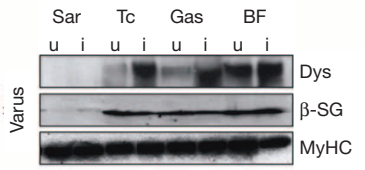


Figure 1 | This is the corrected lower-left panel of Fig. 4b of the original Article.

CAREERS

TURNING POINT Behavioural researcher aims to transform funding models **p.265**

CAREER PROGRESSION Childcare breaks reduce wages for female doctors **p.265**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

SHANNON STAPLETON/REUTERS/CORBIS

A fracking site in Williston, North Dakota. The state has suffered a housing shortage as a result of thousands of workers flocking to join the oil boom.

GEOSCIENCE

Fracking fundamentals

Scientists in the United States who are looking to ride the gas-exploration boom can find a variety of options for employment, from chemical research to environmental monitoring.

BY SID PERKINS

The practice of hydrofracturing (commonly called 'fracking') is booming in North America. The United States is the world's largest producer of shale gas, second only to Canada, and US shale-gas production increased by tenfold between 2006 and 2013. And despite concerns about the sustainability of fracking (see J. D. Hughes *Nature* **494,** 307–308; 2013) and its environmental impact, job opportunities in the industry — many of them science-related — are flourishing.

Fracking involves pumping large amounts of chemical-laden water and sand into subterranean shale formations to shatter rock and then prop open the resulting fissures, which frees up the oil and natural gas entombed there. The increase in hydrofracturing is driving a need for field geologists and petroleum engineers, as well as opening up job prospects for a wide variety of scientists, including chemists and environmental engineers. Many of the posts are related to the need to treat, recycle or dispose of the millions of litres of wastewater that a hydrofractured well can generate.

WELLSPRING OF OPPORTUNITY

The oil-and-gas industry consists of a wide range of companies, all the way from major producers (such as BP in London and Exxon-Mobil in Irving, Texas) and the subcontractors that provide services to them (such as Halliburton in Houston, Texas) down to consulting firms. Hence, the job opportunities are widespread and varied, with different companies often requiring different sets of technical skills and levels of experience. Most hire people holding bachelor's degrees and then train them in-house, says Michael Webber, deputy director of the Energy Institute at the University of Texas in Austin. But there are also plenty of slots for applicants with advanced degrees.

Preliminary figures from the American Geosciences Institute (AGI) in Alexandria, Virginia, show that about 75% of last year's US graduates in geology and geophysics went into the oil-and-gas industry. Furthermore, about 46% of those earning master's degrees and about 33% of those gaining PhDs in the United States also headed for the sector. This is "a very fundamental change", says Christopher Keane, the AGI's director of technology and communications; three years ago, the AGI reported that only some 10% of recently minted PhDs went into the private sector. This move towards industry may stem in part from a relatively limited academic market.

Salaries in the oil-and-gas industry, including the fracking sector, are attractive ▶

compared with most starting academic positions. According to the US Bureau of Labor Statistics, the median income of US geoscientists was just under US$91,000 in 2012. And the Bureau predicts that the number of geoscientist positions will leap by 16% (an increase of about 6,000 posts) by 2022, a full five percentage points higher than the average job growth in the United States during the same period.

North America is currently the hotbed for fracking-related jobs for scientists, but oil-and-gas-rich shales elsewhere will be tapped at increasing rates over the coming decades. For now, shale-gas production in Europe is almost zero but is expected to rise to nearly 85 billion cubic metres by 2040. Likewise, China is expected to produce 141.5 billion cubic metres of shale gas by 2040, making up 50% of the country's natural-gas production.

DANIEL J. SOEDER

Fracking in Greene County, Pennsylvania.

PLUGGING THE GAPS

Poll data underscore the requirement for particular technical skills related to fracking. A survey of oil-and-gas industry professionals by the Society of Petroleum Engineers, based in Richardson, Texas (*J. Petrol. Technol.* **65,** 82–85; 2013), identified a need for people with skills in the recycling, disposal and treatment of wastewater.

Keane notes that prospective employers cite two potential skills gaps in particular among new recruits: a lack of quantitative skills (such as expertise in fluid dynamics) and a lack of field experience (only 40% of recent graduates had attended at least one 6-week-long field camp, the equivalent of an internship). Filling these gaps would boost a job candidate's desirability. Although gaining quantitative skills is fairly standard in a geoscience or petroleum-engineering degree, getting pre-graduation field experience is a bit harder, says J. Foster Sawyer, an exploration geologist at South Dakota School of Mines & Technology in Rapid City. Most of the schools offering petroleum-engineering degrees reserve their field camps for students at their schools, so those seeking to maximize their chances to secure this credential should consider attending these programmes.

GREGORY DANIEL

"People with a keen knowledge of rock mechanics and petrophysics are in very short supply right now."
Scott Tinker

People with a keen knowledge of rock mechanics (how rocks respond to force) and petrophysics (how rocks and fluids interact) "are in very short supply right now", says Scott Tinker, a subsurface geologist and associate dean at the University of Texas at Austin. Before returning to academia 15 years ago, he spent 17 years in the oil-and-gas industry gaining such skills — interpreting rock, seismic and borehole sensor data to search for and develop oil- and gas-rich deposits, and then using those data to build three-dimensional models of oil and gas reservoirs. Such experience is mostly gained through on-the-job training, although some institutions offer classes in such areas. Interested scientists should carefully investigate programmes to ensure that they are offering marketable skills.

PROBLEMS CREATE OPPORTUNITIES

The problems associated with fracking wastewater can be attacked on several fronts, opening up niches for a range of scientists. The search is on for alternatives to the current cocktail of chemicals that is added to water used for fracking: these chemicals (which are often noxious by themselves) have ample opportunity to react with each other in the hot, high-pressure environment deep within the well, spawning potentially even more unpleasant by-products. And the potential for pollution will increase: although about 34% of today's US natural gas production comes from fracking, that fraction will rise to 50% in 2040, according to the US Energy Information Administration.

Chemists could play a major part in reducing wastewater problems, says David Alleman of ALL Consulting in Tulsa, Oklahoma. For one thing, he notes, researchers — whether in the oil-and-gas industry or in academia — are trying to mitigate environmental impacts by designing greener chemicals that either degrade more quickly or are less toxic. Moreover, chemists are looking to design blends in which the ingredients do not react detrimentally with each other within wells. "There's a lot of room for 'downhole' chemists to figure that out," says Webber.

There is also a demand for civil engineers tasked with projects such as designing better surface ponds for storing wastewater. That is because much of the risk from fracking comes from leakage into aquifers, and this occurs mostly from the surface rather than from wells below ground. Environmental engineers and wastewater-treatment specialists could also use their expertise to alleviate problems after a well has been fracked — either by treating the water or by developing no- or low-water technologies.

Petroleum engineer Mukul Sharma heads a research group at the University of Texas at Austin that is trying, among other things, to develop alternative fracking fluids. In Sharma's team, which includes about 27 graduate and five undergraduate students, most are pursuing degrees in chemical, mechanical or civil engineering. But the team also boasts students pursuing degrees in applied maths, geology or geophysics. "This is a very interdisciplinary problem, so it requires people who have a wide variety of backgrounds," he says.

Sharma and his colleagues, both within the industry and in academia, face a tough challenge. Possible alternatives to chemical-laden water include foams based on nitrogen or carbon dioxide. Such fluids would reduce the volume of waste generated during the fracking process because the gas could be removed from the foam after use. But a downside might be the need to inject chemicals to break up the thick foam (potentially creating chemical waste of a different sort). Issues such as these provide ample research opportunities.

RETIRING TYPE

The 16% rise in the number of geoscientist jobs in the United States by 2022 doesn't take into account the sizeable number of positions that will open owing to retirements or attrition. According to the most recently published AGI data, about 12% of the geoscientists working in 2011 are expected to retire by 2018.

The retirements will mean a big loss of technical knowledge in the federal and state regulatory and safety agencies, says Keane, so expert environmental engineers will be needed to monitor air quality and chemical use. For people joining the field, he notes that "it's going to be a tough transition", but early-career scientists will be well positioned to quickly advance into managerial positions, and experienced scientists also have cause for optimism. "People who have been in the field five to ten years will have incredible opportunities."

Whether mid-career scientists or freshly minted graduates, geoscientists interested in the oil-and-gas industry have plenty of options — and in some cases better prospects than in academia for well-paying posts with advancement potential. "Right now, the job market is strong, and the future for young geoscientists is very bright," says Sawyer. ■

Sid Perkins *is a freelance writer based in Crossville, Tennessee.*

TURNING POINT
Johan Bollen

Johan Bollen caused a stir in January when he and his colleagues proposed an alternative science-funding model (J. Bollen et al. EMBO Rep. *http://doi.org/f2pz34; 2014). Bollen, an informatician at Indiana University Bloomington, explains how the proposal developed, and how the idea of resource allocation became part of his research agenda.*

What got you thinking about funding models?
A lot of people are unhappy with the current system. When you submit a proposal, you are like a contractor, but science does not work like that — it works best by generating ideas and gifting them to society and other scientists.

How did your idea take shape?
Some friends and colleagues had a Christmas party in 2012, and as soon as alcohol started to flow, so did commiseration. Guests talked about reviewer comments on proposals, marvelling that one person can have that much power. The disgruntlement is a by-product of how the review system works. I started by saying, "Why not just take all that money and distribute it evenly?". The goal was to see if we could, with as little administration as possible, distribute funding so that researchers have the freedom to explore the topics that they think matter most.

Briefly, what is your plan for science funding?
All scientists would receive a base amount — for example, US$100,000, which roughly corresponds to the US National Science Foundation's 2010 budget divided by the number of senior researchers funded that year. Each scientist would be required to distribute a predetermined percentage of their funding to the researchers whom they believed would make best use of the money.

How did the proposal evolve?
A big concern emerged: some scientists who do not deserve funding will get it. But, we thought, what if every scientist had to distribute some of their funding to others on the basis of their track records? The more we thought about it, the more viable it seemed.

What kind of feedback have you had?
The feedback has been mostly positive, but the proposal is generally regarded as too crazy to work. The main critique is that this is a form of collusion: giving money to a colleague sounds like nepotism. But it would be easy to have conflict-of-interest rules. We could, for example, use funding databases to see if donors were former advisers or at the same institution as recipients. The problem is that the system has no top-down control, which doesn't work for some people.

How do you respond to critics who say that your proposal is anti-peer review?
Peer review is a valuable tool, but funding panels can be costly and have wildly different outcomes. Reviewing one-project proposals is not the best way to allocate funding — I think we should fund people rather than projects.

Your research models public mood using social media. Are you modelling the response to your proposal?
Not scientifically. I have been on Twitter, mostly to answer questions. It sounds callous, but I do not care if people like the proposal. I want them to reconsider their allegiance to the existing system.

Will you continue to push the concept?
Absolutely. My colleagues and I are talking to funders to see if we can run some experiments, including ones with actual funding being distributed and ones involving social choice and funding in selected communities.

How has the idea of innovative resource allocation bled into your research?
I have become enamoured with the idea that society could allocate resources by crowdsourcing rather than assembling panels of experts. I plan to focus more on how resource-allocation algorithms could be applied to societal problems such as poverty alleviation. The decisions of the few may not always be better than the decisions of the many. ■

INTERVIEW BY VIRGINIA GEWIN

DEMOGRAPHICS

UK science workforce

Ethnic-minority workers are most highly represented in the most senior and junior positions of the UK scientific workforce, says a 7 March report. *A Picture of the UK Scientific Workforce*, from the Royal Society in London, examines gender, ethnicity and other factors, and is the most comprehensive analysis of its type, says Julia Higgins, chair of the report's steering group. The report finds that black researchers are slightly under-represented in the most senior roles, whereas scientists from China are statistically over-represented in those positions. It also finds that although women comprise just over half of the scientific workforce, they account for less than one-quarter of those in the highest-level positions.

GENDER

Skewed rankings

Female full professors are less likely than men to co-author papers with assistant professors of the same sex, finds a study (J. F. Benenson *et al. Curr. Biol.* **24,** R190–R191; 2014). Study authors calculated the expected co-author combinations for papers published from 2008 to 2011 by psychologists at 50 US and Canadian universities. They found 14 pairings of senior and junior women, compared with the expected 29, and 76 pairings of senior and junior men, compared with the expected 61. Women's tendency to pair with another woman of the same rank impedes their academic mobility, says co-author Joyce Benenson of Emmanuel College in Boston, Massachusetts.

CAREER PROGRESSION

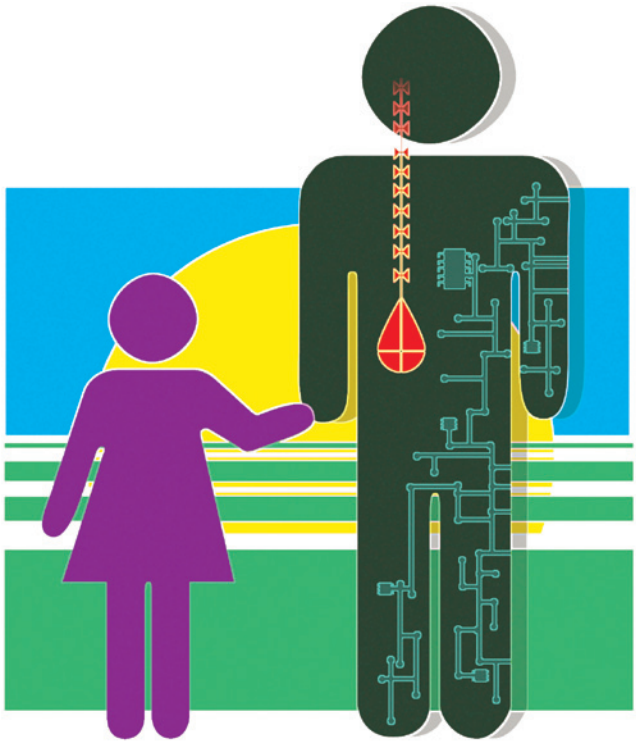
Costs of childcare

Career interruptions for childcare cost female physicians earning power, says a German study (A. Evers and M. Sieverding *Psychol. Women Q.* **38,** 93–106; 2014). The authors surveyed medical students in 1989, asking in part about attitudes towards medical school. A poll of the same cohort 15 years later revealed that earnings correlate with career absences, not with the respondents' earlier outlook. Some 87% of the 47 female respondents reported absences of an average of 1.8 years, mostly for childcare; just under two-thirds of the 52 men reported absences of an average of 7.2 months, mainly for non-employment. Roughly 90% of men were earning more than €36,000 (US$49,440) a year, compared with 55% of women.

A KITE FOR SARAH

In search of freedom.

BY DAVID G. BLAKE



JACEY

"What's it like when they shut one down, Papa?"

Sarah watched me, her emerald eyes cut with a purity I refused to blemish. I lied to her instead. "I don't know, sweetie."

"I bet it's a lot like being a kite that's lost and floating high in the sky."

The truth veered more towards being the one punished for losing the kite, yet I spared her again. "That sounds wonderful."

"I'll ask Mother. She knows *everything*."

Nothing could be further from what I wanted. Her mother would grind that beautiful innocence into a nub of ugly truth. I calmed myself by imagining I really was like a lost kite soaring high. It felt so … so free.

"You know she'll be tired when she gets home from work. Put on your pyjamas, brush your teeth and go to bed. We can talk tomorrow."

Her bottom lip jutted in an exaggerated pout and she stomped away; but it was not in her nature to stay mad for long. She poked her head out from the bathroom a few moments later. "I love you, Papa," she said, her smile an aureole of smeared toothpaste and happiness.

"I love *you*, sweetheart." A truth that magnified my suffering a hundredfold and made it somehow bearable at the same time.

Elizabeth stormed in around eleven.

I blocked the stairs and mustered what I hoped would not prove to be the last scrap of defiance left in me. "They shut her teacher off today. Right in front of the class."

Her green eyes — so much like Sarah's, yet so different — thinned. She tugged off her gloves one finger after another and tossed them onto the counter top. "She questioned you?"

My answering nod felt heavy, laden with betrayal.

"And how did you respond?"

"I lied."

She slapped me. Softly. Hard would have shown a measure of respect instead of cold indignation. "Don't be impertinent."

"I told her I didn't know."

"That'll have to do, I suppose. I'll make the necessary arrangements in the morning."

"Please, Elizabeth, I don't think she is ready."

"Are you sure it is *Sarah* who is not ready?" She held up her hand. "Don't bother giving an answer. It's not your place to think. You're to do as encoded until your usefulness has run its course. You remember what happens after that, don't you?"

"But sh —"

"You'll never replace Peter. Never. I don't care how much you look like him. Now get out of my sight. Your face disgusts me."

I snuck upstairs and watched Sarah sleep. Several times throughout the night, I almost woke her and confessed everything, but I could not so easily relinquish what little time we had left together. I also loathed the idea of my little girl becoming like her mother, and the quicker she learned the truth, the sooner that would be.

"Are we truly going to fly a kite, Papa?"

I nodded, afraid my voice would break to match my heart if I tried to speak; I had planned the day knowing it would be our last spent together.

She clutched the kite — just a simple red one shaped like a teardrop — with unrepressed enthusiasm. I marvelled at how such a thing could make her so utterly happy.

We walked to the field behind our house, and I showed her how to make the kite fly. She ran back and forth, gaze fixed on the whirling red teardrop. I joined her, arms spread wide as if I might also catch an updraft and soar away; it was a nice thought while it lasted. She laughed and spread her arms like mine, eyes reflecting the life of the summer grass. The kite circled above us with purpose, tethered only for so long as Sarah's grip remained true; in such we were more alike than I had imagined.

The sun scampered across the sky towards the frowning horizon. I stopped running in circles and twirled her around instead, losing myself in her laughter a little while longer. Just like that, our last day together started to end. "We should head back," I said. "Your mother's going to be home soon." I knew Elizabeth would not be late, not on this night.

"Do we have to, Papa?"

"I'm afraid so, honey." Such a truth made *nothing* bearable. "Want me to show you how to spool the kite?"

"No. I want to let it go free."

Free. What a powerful word. To hear it uttered by her, even once, was more than I could have hoped. I had decided — while I planned our last day, or perhaps later when I spun her in my arms and lost myself in her laughter — to tell her the truth; risk even more suffering, so that she would understand. But *this* was how I wanted her to remember me.

"Go ahead." It was nice to know I still had a crumb of defiance left in me. "Set it free."

Her hands opened like a summertime bloom. The kite rode the wind into the darkening sky, our similarities ending in one last flicker of red. "Will it go to heaven, Papa?"

I brushed away my tears before she could spot them. "I am sure it will float somewhere nice, sweetheart." ■

David G. Blake *lives in Pennsylvania with his girlfriend and their chocolate Labrador. In addition to* Nature, *his work has appeared in* Beneath Ceaseless Skies, Daily Science Fiction *and several other publications. For more info, visit his Facebook page.*

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm