THIS WEEK

EDITORIALS

POPULARITY The legacy and enduring appeal of two research giants **p.586**

WORLD VIEW United Nations rethinks to face the green future **p.587**

DOGS First domesticated canines were no farm animals **p.589**

Support our buoys

An international effort is needed to restore an early-warning system for the vast warming of the Pacific Ocean that leads to extreme weather worldwide.

The numbers don't add up. When, in 2012, the US National Oceanic and Atmospheric Administration (NOAA) retired the *Ka'imimoana*, a former US Navy ship dedicated to maintaining an array of moored buoys that monitors the equatorial Pacific Ocean, administrators were able to chop roughly US$6 million from the annual NOAA budget. In 2013, the agency says, it spent up to $3 million chartering boats for the same purpose. Those charters have failed to keep pace with the rigorous maintenance requirements, however, and the Tropical Atmosphere Ocean (TAO) array has partially collapsed as a result (see *Nature* http://doi.org/q72; 2014). The upshot is that, to save a few million dollars, NOAA has left the world partially blind to a phenomenon that can cause tens of billions of dollars in damage.

The TAO array exists as a direct result of that phenomenon: an intense warming of surface waters in the eastern equatorial Pacific, known as El Niño. In 1982–83, scientists did not see it coming, and could only watch as its effects rippled through the global weather system to wreak havoc around the world. NOAA researchers responded with a moored array that could be used to monitor both the upper layer of the ocean and the atmosphere above. The agency partnered with the international community to test and deploy the instruments in the 1980s, and by 1994 nearly 70 moorings were in place. That helped scientists to give advance warning several months before the epic El Niño of 1997–98, which nonetheless contributed to extreme weather that killed thousands of people and caused massive amounts of damage.

Working in concert with computer models and satellite observations, the TAO array remains an integral component of a system to give early warning of events in the tropical Pacific. It has also helped researchers to advance the science surrounding El Niño and its sister effect La Niña, which is defined by a cooling in the same region. Progress in this field has laid the foundation for long-range forecasts, and the array provides crucial data for seasonal weather models released by the United States and other governments.

Those are reasons enough to maintain a viable monitoring system in the equatorial Pacific, but the array's value extends well beyond weather forecasting and into basic climate research. It also provides baseline data for researchers studying the effects of global warming on El Niño cycles. For instance, an analysis published this month suggests that the frequency of major El Niño events — such as those in 1982–83 and 1997–98 — are likely to double this century (W. Cai *et al. Nature Clim. Change* http://doi.org/q4c; 2014). And as discussed two weeks ago in these pages, the equatorial Pacific is also a focal point of research into the current global-warming hiatus (see *Nature* **505,** 261–262; 2014).

Budget pressures are understandable, and difficult funding decisions are made every day at agencies such as NOAA. But there can be no doubt that the decision to cut the costs of array maintenance was a mistake. The question now is what to do about it.

To discuss potential solutions, a group of researchers from around the world is meeting this week at the Scripps Institution of Oceanography in La Jolla, California. Although few seem to expect an immediate fix for the array, NOAA promised extra resources for it last week, and all involved must hope that the agency delivers. Further afield, and keeping fiscal constraints in mind, researchers must look at all the available technologies and identify what they need to maintain a viable monitoring system in the Pacific. The burden of implementation need not fall solely on NOAA, and could be shared among government agencies in other countries that benefit from these data, from South Korea to the United Kingdom.

"The benefits of this system are truly global."

Also on the agenda in La Jolla are the bureaucratic barriers hindering the international cooperation that would ensure scientists have the funds and ships they need to maintain the array. These obstacles must be overcome, and a look at the array's own past provides reason for hope. Six countries took part in its initial testing and deployment, and since 2000, Japan has maintained a dozen of the original moorings in the western Pacific, called the TRITON array. The benefits of this system are truly global. It makes sense for the international community to come together on a long-term solution. ■

Open invitation

Europe's proposed climate targets fire the starting gun on the long build-up to Paris 2015.

When European leaders agreed on three climate and energy targets in 2008, and established a set of policies by which to achieve them, the European Union (EU) was widely acknowledged as the world's first major economic power to tackle the climate-change problem in earnest.

Those landmark '20-20-20 targets' for 2020 aimed for a 20% reduction in greenhouse-gas emissions below 1990 levels while setting a mandatory 20% target for the share of electricity consumption coming from renewable energy sources and a 20% improvement in energy efficiency by that time.

With EU emissions now down by some 18% relative to levels in 1990, Europe is well on its way to exceeding the first and crucial goal. Against that background, the new mid-term emissions target — a 40% reduction on 1990 levels by 2030 — proposed by the European Commission last week has received a lukewarm response from environmental groups, scientists and green-minded politicians (see page 597).

The commission wants to scrap binding national renewable-energy targets and introduce a mere aspirational goal for the EU as a whole. This has led some critics to infer a Brussels-conspired counter-revolution in climate policies, which they say will deal a blow to Europe's emerging renewable industry and open the door to a renaissance of nuclear power on the continent. But the commission's proposal has more teeth than its critics would like to admit.

According to state-of-the-art energy-economy models, 40% emissions cuts by 2030 are achievable at reasonable cost and, provided sound investment is made in energy research, do keep Europe on track to cut emissions by at least 80% by mid-century.

Announced just as Europe is surfacing from the most severe economic downturn since the Great Depression, the cost efficiency of the plan is essential to its chances of success. To burden member countries with excessive environmental measures at this time could do more long-term harm than good. An economically weak, socially struggling region is unlikely to produce the wealth and creative power that will be needed to achieve the great transformation to a low-carbon civilization.

That transformation is a global task. With the EU accounting for little more than 10% of global greenhouse-gas emissions, the bulk of the effort will need to be accomplished elsewhere. But although the focus of global climate policies is increasingly shifting to the world's rising economies — and to China in particular — both the timing and the content of Europe's latest promise on global warming could be essential to building political momentum.

With a view to the United Nations climate talks next year in Paris, where nations hope to replace the underachieving 1997 Kyoto Protocol with a more stringent global climate agreement, the EU's bid is a clear and unambiguous signal. What Brussels has dished up well in advance of the Paris climate gala is a polite but firm invitation to the rest of the world, and one that governments from Beijing to Washington cannot lightly afford to ignore. By the end of the year, at the latest, the EU's main economic competitors will be expected to lay on the table solid offers for that crucial round of negotiations.

In terms of the magnitude of emissions cuts, the EU's unilateral proposal is an indication of the minimum level of commitment other developed nations can be expected to make if they take their climate-change responsibilities remotely seriously. But governments — including those of EU member states — must be reminded that gentle pathways to decarbonization such as the EU hopes to follow are by no means a guarantee of a benign future climate. In fact, even the more optimistic scenarios currently under debate would give the world at best a 50% chance of staying below 2 °C of warming, the often-cited threshold to dangerous climate change. The science strongly suggests that reducing this probability to a tolerably small value would require global emissions cuts at least twice as high as those proposed in Brussels last week.

"Europe's latest promise on global warming could be essential to building political momentum."

The question of how the substantial global cuts that might be required to safely stay below 2 °C of warming should be apportioned between rich and poor countries is one that science alone cannot answer. This issue requires input from ethics and the theory of justice as much as it does from science and empirical economics. The EU's latest climate aspirations, whether or not one considers them sufficient, are a timely reminder of the intricacies of the issues at stake. ■

Crystal clear

Celebrating the many achievements of crystallography.

In one of the more bizarre examples of science outreach, the website starnostar.com gives readers the chance to vote on who should win a popularity fight between the physicists Max von Laue and Paul Dirac (see go.nature.com/fw1omn). To the non-expert, there is not much to go on; the website biographies offer brief details on the physicists' birth places and their sign of the zodiac, but nothing on their achievements, popular or otherwise. (Dirac currently leads, with 69% of the vote, but don't despair, von Laue fans; the contest remains open, and a surge in support could yet tip the balance.)

Itching to pitch in to help choose between two of the greatest minds of the twentieth century, but unsure about their true credentials? Read on.

"In the right corner, Max." A friend of Albert Einstein and a student of Max Planck, von Laue (a Libra) is the rugged outdoors type. He discussed his Nobel-prizewinning idea that X-rays passing through a crystal would bounce around to form an identifiable signature while skiing. Skiing! He was brave as well — he stood up to the Nazis in his native Germany and helped Jewish colleagues to escape the country. He won a Nobel prize, and earns bonus points for the rip-roaring boys'-own tale of how the gold award was dissolved to hide it during the war, and then later recast.

"In the left corner, Paul." An awkward man and a sensitive soul, Dirac lived for his work and had little time for small talk, or for much else. But what work it was. His mathematical wizardry unlocked the secrets of quantum mechanics and quantum electrodynamics. He won a Nobel prize too, aged just 31! And for all you anti-establishment British types, he refused a knighthood. (He did not want to be known by his first name.)

Still undecided? Then take a look at a special collection of articles that begins on page 601, and a research paper on page 657. More than a century since von Laue's moment of inspiration on the slopes, and exactly a century since his Nobel prize, 2014 is the International Year of Crystallography. There are a lot of such celebratory years these days. But indulge us, and the organizers, who want to shout about the achievements and contributions of X-ray crystallography. Crystallographers deserve the chance — too often in the background when the spotlight falls on scientific accomplishment, like one of their refraction patterns, it is worth piecing together their separate successes to build a coherent image of the whole.

Such anniversaries and commemorations inevitably cast the eye and the mind backwards in time. But as this week's special collection makes clear, crystallography remains a cutting-edge field, and one that, if harnessed properly, could contribute as much in the next 100 years as it did in the previous 100. The development of the X-ray free-electron laser, for example, is a monumental technical achievement, and one that seems more suited to the world of 2114 than 1914, or even 2014.

Dirac's work continues as well. On page 657, physicists describe the first creation of something he predicted in 1931 — a magnet with a single pole: the Dirac monopole. A triumph of a growing research field called quantum simulation, which exploits real quantum systems to model others that are difficult to achieve, the research shows that not all magnets need have opposing 'north' and 'south' poles. Now that they know such a thing is possible (see the News & Views article on page 627 for more), physicists will continue to search for them with a spring in their step. As Dirac said: "one would be surprised if Nature had made no use of it."

Back to starnostar. To choose between Dirac and von Laue, of course, is to be forced to select either the north pole or the south pole of a magnet. As Dirac and von Laue, and later physicists, show us, we don't need to do that. Each can stand on his own. And much else rests on both. ■

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

Quiet green revolution starts to make some noise

The formation of the UN Scientific Advisory Board is an important step towards integrating global sustainability efforts, says **Owen Gaffney**.

This week sees the first meeting of a board of scientific experts set up to advise UN secretary-general Ban Ki-moon. In a modest way, it is a historic move — never before has the head of the United Nations had what amounts to a team of chief scientific advisers. Furthermore, the meeting in Berlin marks one of the first outward signs of a quiet international revolution that is building new bridges between science and policy.

Each member of the board will serve for two years, and is supposed to act independently, rather than lobbying for his or her nation. Among the 26 scientists are Abdul Hamid Zakri, science adviser to the prime minister of Malaysia and chair of the Intergovernmental Platform on Biodiversity and Ecosystem Services; Brazilian Earth-system scientist Carlos Nobre; and Bulgarian global environmental governance expert Maria Ivanova.

The inclusion of political scientists is a bold move reflecting a growing awareness that the governance arrangements of the twentieth century are struggling to cope with the challenges of the twenty-first. That failing was highlighted repeatedly at the annual meeting of the World Economic Forum in Davos, Switzerland, last week.

The board has its origins in the UN report *Resilient People, Resilient Planet*, published for the Rio+20 conference on sustainable development in 2012, which recommended that "the Secretary-General should consider naming a chief scientific adviser or establishing a scientific advisory board with diverse knowledge". But it can also be seen as a response to another 2012 UN report, the damning *21 Issues for the 21st Century*, which highlighted what it called broken bridges between science and policy. It identified a lack of "meeting points" between scientists and politicians that is causing knowledge to remain locked in silos. As a result, the link between science and society becomes strained and public confidence — in climate science for example — is weakened.

Partly because of the size of the UN and partly because of how it has evolved, myriad commissions, programmes and organizations work on what can be grouped under the heading of sustainable development. This makes it difficult to coordinate policies. Worse, some are in direct conflict. The World Bank, for example, has invested in energy projects that fly in the face of efforts to reduce carbon emissions.

Reform will take time, and the problems run deeper than the links between science and policy. Greater change is under way: Ban announced the scientific advisory board last September at the first meeting of the UN High-Level Political Forum on Sustainable Development, the flagship that he hopes will bring about much-needed coordination.

Political fragmentation has a knock-on effect on international science programmes: when researchers work with the UN, we are forced to deal with issues in the same silos it does. But when it comes to global sustainability, the environment can no longer be separated from economic growth, nor can action on food security be separated from action on biodiversity.

A significant strength of the new advisory board is that it will form a bridge between the UN and international research. The timing is good. The landscape of international Earth-system and sustainable-development research is itself undergoing major reform, spearheaded by the ten-year research programme Future Earth, which is bringing together all the major players. As Future Earth develops its science plan, the advisory board has within its remit to identify "knowledge gaps" that could be addressed by "international research programs, e.g., the emerging 'Future Earth'". The scene is set for these two initiatives to lock together like a jigsaw puzzle.

WHEN IT COMES TO GLOBAL SUSTAINABILITY, THE ENVIRONMENT CAN NO LONGER BE SEPARATED FROM ECONOMIC GROWTH.

Future Earth integrates networks including the International Geosphere-Biosphere Programme (IGBP), the DIVERSITAS biodiversity programme and the International Human Dimensions Programme. The latter two will close this year and, after 28 years, the IGBP is scheduled to close its doors in 2015.

It is early days for Future Earth, but the ambition is clear: its architects argue that there needs to be an urgent shift in international science, from a focus on understanding the Earth system and how humans interact with it to meeting the needs of 10 billion people as Earth's life-support system is transformed. This is not so much bridge repair as construction of an entirely new bridge.

As such, planning is detailed, negotiations protracted, the lag between idea and implementation drawn out. Traditionally, international science programmes have had few links with engineering, technology and business, but this is where the solutions to modern problems will be found. Whole new networks need to emerge.

This, too, is happening. Immediately following Rio+20, Ban set up the Sustainable Development Solutions Network, led by US economist Jeffrey Sachs. This is a global network of research centres, universities and businesses tasked with innovative problem-solving. With a direct line to the secretary-general's office and Future Earth, it has already built much momentum.

Taken together, these initiatives and the appointment of the UN scientific advisory board will inject energy into a tired system. This is worth celebrating — not least because it creates a mechanism for ongoing reform, rather than having to wait 20 years for the next Earth summit. ■

➲ NATURE.COM
Discuss this article online at:
go.nature.com/7tdxae

Owen Gaffney *is director of communications at the International Geosphere-Biosphere Programme in Stockholm.*
e-mail: owen.gaffney@igbp.kva.se

RESEARCH HIGHLIGHTS

Selections from the scientific literature

MICROBIOLOGY

Immobile bacteria hitchhike on rafts

Bacteria that are unable to move on their own can hitch a lift on their mobile neighbours.

Yael Helman of the Hebrew University of Jerusalem and her team found that, on agar plates, the bacterium *Xanthomonas perforans* — which does not move on solid surfaces — triggered *Paenibacillus vortex* to move closer to it, and then used this travelling species as transport. This interaction occurred even when the two species were separated by a plastic barrier, suggesting that *X. perforans* releases an airborne substance to signal for a lift. Electron-microscope images revealed single *X. perforans* cells on 'rafts' of *P. vortex*.

This hitchhiking also occurred on leaves, and between other xanthomonads and motile bacteria, suggesting that the behaviour could be widespread.

ISME J. **http://doi.org/q67 (2013)**

E. LACASA-MARQUINA/L. MARTÍN-FRANCÉS

PALAEOANTHROPOLOGY

Broken teeth point to rough diet

Teeth from a 1.8-million-year-old human fossil show signs of disease and are extremely worn — possibly from eating hard and fibrous foods.

In 2000, researchers uncovered a jaw bone (pictured) at a site in Dmanisi, Georgia, which has produced the oldest human fossils outside Africa. Laura Martín-Francés at the National Research Centre on Human Evolution in Burgos, Spain, and her team examined the fossil, dubbed D2600, including its teeth. Most of the teeth had no protective enamel left, and the roots and interior showed signs of infection.

The wear patterns — which are unlike those of other human specimens of a similar age — could have been caused by a diet of abrasive and fibrous plants and fruits, similar to that of apes, the researchers say.

Comptes Rendus Palevol **http://doi.org/q5t (2014)**

GEOLOGY

Water dives deep inside Earth

Slabs of Earth's crust that are plunging deep into the planet could be carrying much larger amounts of water into the planet's mantle than previously thought.

In the northwest Pacific Ocean, where the Pacific plate sinks beneath Japan, Tom Garth and Andreas Rietbrock of the University of Liverpool, UK, studied earthquakes originating from within the diving slab. Modelling indicated that the quakes occur along water-rich faults that form as the plate bends before diving below.

Over Earth's lifetime, the Pacific plate could have taken the equivalent of 3.5 oceans into the mantle. Some of that water is released and rises upward, fuelling volcanoes; the rest plunges deeper into the planet.

Geology **http://doi.org/q7p (2014)**

ATMOSPHERIC SCIENCE

Shifting winds freeze China

Not only has climate change been responsible for frequent bouts of record-breaking summer heat in China since 2000, but it could also be the cause of the unprecedented winter cold that has plagued northern parts of the country in several recent years.

Xueyuan Kuang and her team at Nanjing University in China analysed the distribution of record-breaking high and low temperatures observed between 1951 and 2010 at nearly 1,900 weather stations across China. Records for summer highs were set more frequently between 2000 and 2010 than in the previous two decades. Record winter lows seemed to cluster in northern China in the 2000s, whereas in the 1990s they were spread across most of the country.

This clustering seems to be a result of air-pressure anomalies and shifting jet streams over Eurasia in autumn and winter since the late 1990s. These changes can cause cold Siberian air to flow into and persist over northern China, the team found.

J. Geophys. Res. **http://doi.org/q5k (2014)**

METABOLISM

Mother's fatty diet hurts offspring

Female mice that eat a high-fat diet while nursing their pups predispose them to obesity and diabetes by altering the pups' brain wiring.

Tamas Horvath at Yale University in New Haven, Connecticut; Jens Brüning at the Max Planck Institute for Neurological Research in Cologne, Germany; and their team discovered that mice that

ate a fatty diet during lactation had pups that were fatter, had higher insulin levels and were less sensitive to insulin than the offspring of mothers that ate a normal diet. In the fat pups, fewer fibres from specific neurons branched into regions of the brain's hypothalamus that regulate energy metabolism.

This circuitry is established in mice shortly after birth, but in humans it develops during the last trimester of pregnancy. The authors suggest that a mother's diet during this period could have long-term health effects for the child.
Cell **http://doi.org/q7k (2014)**

ENGINEERING

Phone device detects mercury

A smartphone attachment can detect low levels of mercury in water samples, opening the door to on-site, low-cost environmental monitoring.

Inorganic mercury is harmful to the kidneys, and can be converted by bacteria into its neurotoxic, organic forms. The device (**pictured**), developed by Aydogan Ozcan and his colleagues at the University of California, Los Angeles, can measure inorganic mercury at levels of 3.5 parts per billion (p.p.b.) — good enough to detect the maximum acceptable level of 6 p.p.b. advised by the World Health Organization. The attachment shines green and red light through tiny test tubes, which contain the water sample and a few reagents. The mobile phone's camera detects the light, which shifts towards green wavelengths if mercury is present. A custom-made app provides the measurement.

The researchers tested their device by creating a mercury-contamination map of 50 locations in California.
ACS Nano **http://doi.org/q6n (2014)**

OZCAN RESEARCH GROUP/UCLA

IMMUNOLOGY

Poor diet boosts innate immunity

Vitamin A deficiency enhances the immune system's response to parasitic worm infections in mice.

Malnutrition typically impairs the body's ability to fight infection. But Yasmine Belkaid at the US National Institutes of Health in Bethesda, Maryland, and her team found that depriving mice of vitamin A boosts an arm of the immune system that protects the body's barriers, such as the gut. Animals lacking this vitamin had a much higher level of ILC2 cells — immune cells that are active in barrier defence — in the gut than mice on a normal diet, and were better able to fend off infection by a nematode worm.

Vitamin A deficiency is common in areas where worm infection is also prevalent. The findings suggest a way that the immune system has adapted to promote survival even in the face of malnutrition.
Science 343, **432–437 (2014)**

PUBLIC HEALTH

Mutations toughen up tuberculosis

A genomic analysis of the tuberculosis bacterium in a Russian population reveals that the microbe is not only evolving resistance to multiple drugs, but also retaining its ability to survive and spread.

Francis Drobniewski at Queen Mary University of London and his colleagues sequenced the genomes of 1,000 *Mycobacterium tuberculosis* isolates from people in western Russia. Two-thirds of the isolates belonged to a lineage that first emerged in Asia and is prone to developing drug resistance. More than 60% of the isolates had drug-resistance mutations. Such mutations typically hinder bacteria's ability to spread, but the team found new 'compensatory' mutations that might maintain transmissibility in more than 400 isolates with resistance to the antibiotic rifampicin.

The findings suggest that biological factors, and not just weak public-health measures, are behind the high incidence of tuberculosis in Russia.
Nature Genetics **http://dx.doi.org/10.1038/ng.2878 (2014)**

PHOTOVOLTAICS

Hot solar cells make more power

A photovoltaic device that converts sunlight into heat to generate power has achieved greater efficiency than previous such devices, thanks to the design of nanomaterials in the light-absorbing layer.

Thermophotovoltaics contain a layer that absorbs a wider spectrum of wavelengths than conventional solar cells. This layer radiates heat that is used to generate electricity. Evelyn Wang and her team at the Massachusetts Institute of Technology in Cambridge designed their absorber–emitter material by growing an array of carbon nanotubes, which turn light into heat, onto a layer of photonic crystals, which they engineered to emit energy of the optimum levels for power generation.

The researchers' device reached an energy conversion efficiency of 3.2%, three times greater than in previous experiments. The authors say that with further improvements, efficiency could exceed 20%.
Nature Nanotech. **http://doi.org/q6j (2014)**

COMMUNITY CHOICE

The most viewed papers in science

GENOMICS

Dogs domesticated before farming

HIGHLY READ on plosgenetics.org 23 Dec–22 Jan

Dogs became companions for humans long before the advent of agriculture, according to a genome-sequencing study.

A team led by Robert Wayne at the University of California, Los Angeles, and John Novembre now at the University of Chicago, Illinois, analysed the genomes of three wolves (*Canis lupus*) from regions where dogs are thought to have first been domesticated. The authors also studied the genomes of two dog breeds, including Australian dingoes (pictured), and of a golden jackal. The researchers determined that dogs were probably domesticated from now-extinct wolves between 11,000 and 16,000 years ago — before humans began farming around 10,000 years ago.

The findings contradict a previous genome study, which argued that dog domestication was associated with farming.
PLOS Genetics 10, **e1004016 (2014)**

STEFFEN AND ALEXANDRA SAILER/ARDEA.COM

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

SEVEN DAYS *The news in brief*

POLICY

Drilling setback

The US government failed to assess potential environmental impacts adequately when it opened the Chukchi Sea off Alaska to oil drilling in 2008, a federal appeals court ruled on 22 January. In its analysis, the US Department of the Interior used a production estimate of 1 billion barrels of oil, but the court sided with the argument of environmentalists and Native American groups that actual production could be much higher. The ruling could further delay exploration in the region by companies such as Royal Dutch Shell (see *Nature* **495,** 11; 2013).

EU climate package

The European Commission unveiled a package of climate and energy proposals on 22 January, with targets for 2030. European Union member states are to reduce their collective greenhouse-gas emissions by 40% relative to 1990 levels (see page 597). The package also includes recommendations for managing shale-gas extraction by fracking, but not the binding environmental regulation that had been under consideration. Instead, the commission will weigh up over the next 18 months whether further legislation is needed.

Pharma patent flap

Advocates of affordable medicines expressed outrage last week after leaked documents revealed a proposed public-relations campaign by a lobbying firm in Arlington, Virginia, to stymie drug-patent reform in South Africa. The country is considering loosening patent protections to improve access to cheaper, generic drugs, in line with moves in India and Brazil in recent years (see *Nature* **500,** 266; 2013). The Innovative Pharmaceutical Association South Africa, a trade group based in Randburg, acknowledged receipt of the campaign proposal, but said that the plans had been reviewed and rejected.

ADAM BLOCK/MOUNT LEMMON SKYCENTER/UNIV. ARIZONA

Supernova seen in nearby galaxy

Astronomers have spotted one of the closest supernovae in years — in the galaxy M82, about 3.5 megaparsecs (11.4 million light years) away. Students and staff at the University of London Observatory discovered the exploding star in the Ursa Major constellation during a telescope lesson on 21 January. Other astronomers quickly combed through archive data, unearthing earlier, fainter images of the event. Designated SN 2014J, the supernova (pictured) is expected to reach peak brightness in early February. It belongs to the type Ia class of supernovae, formed when one star loses enough mass to a companion white-dwarf star to cause the white dwarf to explode. Owing to their predictable brightness, type Ia supernovae played a key part in the discovery of the Universe's accelerating expansion. See go.nature.com/wmeet2 for more.

Help for headaches

Britain's National Institute for Health and Care Excellence has approved the treatment of migraine headaches with a magnetism-based procedure applied through the scalp. Guidelines issued on 22 January said that transcranial magnetic stimulation (TMS) could be used to reduce headache severity or frequency. However, the agency warned that TMS is not a cure, and that evidence for its efficacy and long-term safety is limited. Last December, regulators in the United States approved the country's first commercial TMS device to relieve migraine pain.

BUSINESS

Biotech blues

Biotechnology company Advanced Cell Technology (ACT) has lost its chief executive, Gary Rabin, who resigned on 22 January. The firm, based in Marlborough, Massachusetts, faces the possibility of bankruptcy after a series of financial missteps. ACT is running the only trials approved by the US Food and Drug Administration to test therapies involving embryonic stem cells. See *Nature* http://doi.org/q8f (2014) for more.

Google thinks deep

Google has purchased the London-based artificial-intelligence company DeepMind, which uses human neuroscience to inspire computer algorithms. Google, of Mountain View, California, confirmed the deal this week; in the past few years it has hired several big names in artificial intelligence, including futurist Ray Kurzweil and computer scientist Geoffrey Hinton (see *Nature* **505,** 146–148; 2014). The company may use artificial intelligence

to improve picture tagging, voice recognition and search engines.

RESEARCH

Antihydrogen made

Physicists have produced a stream of antihydrogen atoms for the first time. Members of the Atomic Spectroscopy And Collisions Using Slow Antiprotons experiment at CERN, Europe's high-energy physics laboratory near Geneva in Switzerland, reported on 21 January detecting 80 of the antiatoms 2.7 metres from their source (N. Kuroda *et al. Nature Commun.* **5,** 3089; 2014). The researchers hope that by isolating the antiatoms from the strong magnetic fields used to create and trap the particles, they can characterize small differences between antihydrogen and hydrogen. These differences could help to explain why the Universe contains more matter than antimatter.

River dolphin found

A new species of river dolphin, found in Brazil's Araguaia River basin, is the first such discovery in almost 100 years, researchers reported on 22 January (T. Hrbek *et al. PLoS ONE* **9,** e83623; 2014). The species, *Inia araguaiaensis* (**pictured**), was identified through genetic testing and

NICOLE DUTRA

probably diverged from similar South American river species more than 2 million years ago. About 1,000 individuals may live in the Araguaia River basin, the scientists estimate.

Who's who

More than half a million researchers have now registered for a scheme to provide authors of scientific publications with a unique identifier. The Open Research and Contributor ID (ORCID) group, a non-profit organization based in Bethesda, Maryland, announced on Twitter on 21 January that it had hit the membership milestone. Organizers of the ORCID database hope to link researchers' identities across publications, grant applications, patents and other activities (see *Nature* **485,** 564; 2012).

EVENTS

Massive ivory burn

Hong Kong is set to incinerate a huge stockpile of about 30 tonnes of seized ivory, following a unanimous decision by the Endangered Species Advisory Committee on 23 January. The government is still working on details of the plan, but destruction of the ivory is expected to begin by mid-2014, and to be complete within two years. Hong Kong's announcement follows recent, high-profile examples of ivory destruction in the United States and China. See go.nature.com/ib2fpa and *Nature* http://doi.org/q8g (2014) for more.

Rabbit rescue

China's moon rover has run into major trouble, according to a report on 25 January from state-run news agency Xinhua. The Yutu ('Jade Rabbit') rover experienced a "mechanical control abnormality" as it prepared to hibernate over its second lunar night (roughly equivalent to 14 days on Earth) since landing on the Moon last month (see *Nature* **504,** 336; 2013). Scientists are working to resolve the problem, but have released few other details.

Pig virus spreads

Canada confirmed its first case of porcine epidemic diarrhoea virus on 23 January. The virus, which causes diarrhoea and vomiting in pigs, was detected on a farm in Middlesex County, Ontario. First identified in the United Kingdom in 1971, the virus can kill 80–100% of infected piglets. It caused mass epidemics in Europe in the 1970s and 1980s. Last spring, the United States reported its first case (see *Nature* **499,** 388; 2013), and the virus has since spread to 23 states.

COMING UP

30–31 JANUARY
In Berlin, the United Nations convenes the first meeting of its newly formed Scientific Advisory Board, which consists of 26 international scientists (see page 587 and go.nature.com/4ts2qb). The group will advise the agency on sustainable development, including issues of food and water security, and climate change.

3 FEBRUARY
The World Health Organization releases its third World Cancer Report, six years after its previous publication. The latest report will include updated trends in cancer incidence, prevalence and mortality.
go.nature.com/x39hvk

PEOPLE

Fraudster punished

Biotech investor David Blech is heading for prison after unsuccessfully appealing against a four-year sentence for fraud. Blech, who helped to set up the biopharmaceutical firm Celgene in Summit, New Jersey, pleaded guilty in May 2012 to manipulating the stock of two other companies. On 21 January, a US appeals court upheld the prison term, as well as an order for Blech to forfeit US$1.3 million.

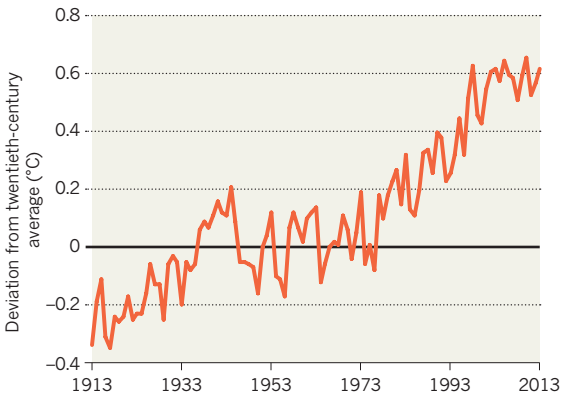
➲ NATURE.COM
For daily news updates see:
www.nature.com/news

TREND WATCH

The average global temperature in 2013 was 0.62 °C above the twentieth-century average, according to an analysis released on 21 January by the US National Oceanic and Atmospheric Administration (see chart). Overall precipitation was near average, but the year was characterized by extreme drought and flooding events. Brazil, Angola and Namibia experienced their worst droughts in decades, and other African nations and Europe's Alpine region saw intense rains and floods.

TAKING THE GLOBE'S TEMPERATURE

Average annual temperatures over land and ocean have exceeded the twentieth-century average each year since 1977.



SOURCE: NOAA

NEWS IN FOCUS

KIM KYUNG-HOON/REUTERS/CORBIS

Woo Suk Hwang's human-cloning research was deemed fraudulent by Seoul National University in 2006.

MISCONDUCT

Whistle-blower breaks his silence

South Korean researcher reveals the fallout he faced from his tip-offs about former cloning fraudster Woo Suk Hwang.

BY DAVID CYRANOSKI

The whistle-blower who played a key part in exposing the fraud of South Korean cloning specialist Woo Suk Hwang has spoken for the first time about his role in the scandal — and the suffering he endured as a result.

Young-Joon Ryu, who was a key figure in Hwang's laboratory for several years, kept his silence for eight years. But in a blog post in December 2013 and a subsequent interview with *Nature*, he revealed that he was responsible for initiating the investigation that uncovered one of the biggest frauds in science. He has since received both support and abuse, highlighting just how divided South Korean society still is over the legacy of its fallen hero.

"The nature of the Hwang scandal is the abuse of other people's sacrifice and other people's lives for personal success," Ryu, now in the pathology department at Kangwon National University in Chuncheon, told *Nature*.

Hwang claimed in 2004 to have cloned a human embryo and produced stem cells from it, potentially opening the way for new disease treatments. In 2006, he admitted fabricating his findings, but despite being convicted of fraud, has since made a controversial comeback (see *Nature* **505,** 468–471; 2014).

Ryu joined Hwang's laboratory at Seoul National University in 2002, and that year led the team that attempted to create cloned human embryos and stem-cell lines from them. He wrote the first manuscript of an article on the work, which was published with great fanfare in February 2004 (W. S. Hwang *et al. Science* **303,** 1669–1674; 2004).

While Hwang basked in its glory, Ryu started to have misgivings about Hwang's tendency to seek publicity. He also felt that human cloning had little potential for clinical applications. In April 2004, he left the laboratory and soon began work at the Korea Cancer Centre Hospital.

When Hwang's group published a dazzling follow-up the next year that suggested that the previous proof-of-principle was almost ready for the clinic (W. S. Hwang *et al. Science* **308,** 1777–1783; 2005), Ryu was suspicious. He knew that important lab members had left, yet the team had pumped out 11 embryonic stem-cell lines in a short time. "I knew how difficult it was," he says. "It wasn't logical."

He then heard that Hwang was preparing a clinical trial for a 10-year-old with a spinal-cord injury, whom Hwang had promised to make walk again. Ryu had known the boy and worried that a trial could hurt him. "I was furious," he says. "I wanted to stop all of that."

Lacking evidence and worried that his identity might be revealed, Ryu baulked at approaching the university or police. Instead, on 1 June 2005, he e-mailed television network Munhwa Broadcasting Corporation (MBC) to recommend an investigation.

MBC producers were initially intimidated by Hwang's star status, but decided to work with Ryu to develop their case. A first programme on the subject, about ethical violations in the way that Hwang recruited egg donors, aired on 22 November 2005, and forced a confession from him. A storm of support for Hwang ensued. A second programme, concerning the fraudulent research, was postponed ▶

after sponsors withdrew support for the TV network and producers faced legal and physical threats. But suspicion was mounting. Posts on the website of the Biological Research Information Center (BRIC), in which volunteers noted errors in the papers, helped to force Seoul National University to open an investigation. By the time the second show aired, on 15 December, Hwang's fate was sealed.

Ryu's identity was leaked after MBC's first programme, and his worst fears about the militancy of Hwang's supporters were borne out. Ryu says that they hacked his blog and sent threatening e-mails to him, his employer and his wife, another former researcher in Hwang's laboratory. On 6 December 2005, Ryu resigned from his hospital job under pressure.

Ryu, his wife and their 8-month-old daughter went into hiding for the next six months. "We cried a lot," says Ryu. It was 2007 before the ostracized Ryu could find paid employment, as a pathology resident at Korea University in Seoul.

On 23 December 2013, Ryu posted a note on the BRIC site to thank those who supported him and signed off with his real name. Some 8,000 people viewed the post, which garnered a few dozen sympathetic comments. But then the story was picked up by a local newspaper and the tone changed. Of more than 1,000 comments on the popular Daum news-aggregator website, 90% have been negative. Online commenters have said that by "revealing a petty truth", Ryu caused South Korea to "fall behind in the stem-cell business". Another accuses him of "satisfying his arrogance" while "seriously injuring the nation" as the "entire project was stolen by other nations".

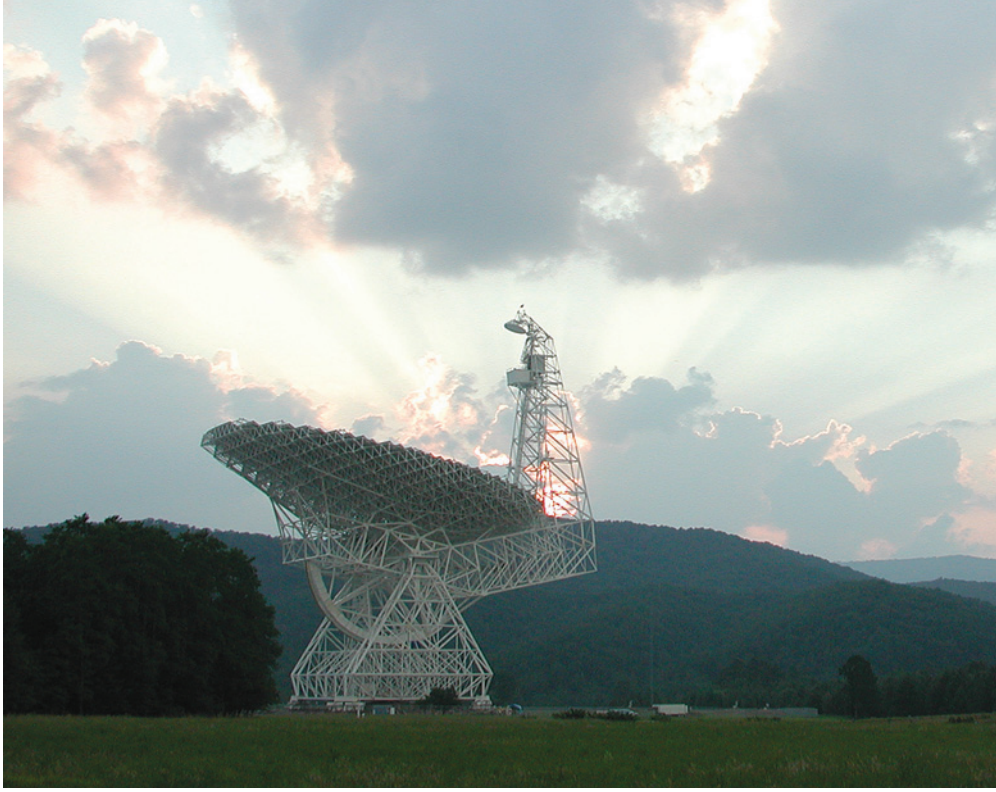
DAVID CYRANOSKI/*NATURE*

"The nature of the Hwang scandal is the abuse of other people's sacrifice."
Young-Joon Ryu

Ryu says that he has no regrets about what he did. The scandal did not ruin his faith in science either. He completed a PhD in bioethics and safe research in 2011 and is now pursuing a doctoral degree in animal reproductive biology at Seoul National University.

The episode shows how whistle-blowing still carries risks, especially for junior researchers, says Bernd Pulverer, head of scientific publications at the European Molecular Biology Organization in Heidelberg, Germany. "The Hwang case was a wake-up call for many journals to police [fraud] more seriously," he says. But he adds that "little has formally changed regarding the protection and encouragement of constructive whistle-blowing". ■

Additional reporting by Soo Bin Park.



West Virginia's Green Bank Telescope needs partners to pay half of its US$8-million operating costs.

ASTRONOMY

US struggles to offload telescopes

West Virginia radio observatory seeks money from partners to fend off closure by the National Science Foundation.

BY ALEXANDRA WITZE

Astronomer D. J. Pisano got to spread some good news last month. He and his colleagues at West Virginia University in Morgantown announced a US$500,000 grant from the National Science Foundation (NSF). The money will allow his team to build an antenna-like detector to speed up sky surveys at the Green Bank Telescope (GBT), the nearby 110-metre-wide radio dish that is the largest steerable radio telescope in the world.

There is just one problem. Even as the NSF funding goes towards improving the telescope, the agency is trying to get rid of it.

Following an independent 'portfolio review' in 2012 (see *Nature* **488,** 440; 2012), the NSF is exploring closing the GBT and nine other telescopes it operates (see 'Closing time'). The alternative is to find partners to share the cost. West Virginia University has already shelled out $1 million to buy time on the GBT to bolster its growing astronomy faculty — a first hint of what a future for these jettisoned telescopes might look like.

Still, such partnerships are frustratingly hard to achieve. Last month, the NSF reported that, thanks to the slow pace of discussions and the complex environmental reviews required to shut national facilities, it is not close to making any permanent decisions. That leaves the future of the telescopes in limbo — and puts the careers of astronomers such as Pisano on edge. "We were obviously upset by it, we were somewhat confused," he says.

For the NSF, there is some urgency to replace the old with the new. By offloading the old telescopes, the agency could free up about 10% of its $233-million astronomy budget. That would allow more money for research grants. More importantly, it would regain money for future telescopes, such as the Large Synoptic Survey Telescope, which astronomers are slated to begin building in Chile this year (see *Nature* **505,** 461–462; 2014). "Our job is to foster frontier science," says James Ulvestad, who heads the NSF's astronomy division. "Within a constrained budget there is nothing you can do that isn't going to hurt somebody."

NRAO/AUI/NSF

CLOSING TIME
The US National Science Foundation is seeking partners to take over its share of operations for ten sets of telescopes. (The first four are closest to divestment.)

Telescope	Location
Arecibo Observatory (radio)	Puerto Rico
Green Bank Telescope (radio)	West Virginia
Very Long Baseline Array (radio)	10 locations across the United States
NOAO 2.1-metre telescope (optical)	Arizona
Mayall 4-metre telescope (optical)	Arizona
WIYN 3.5-metre telescope (optical)	Arizona
McMath–Pierce Solar Telescope (solar)	Arizona
SOAR 4.1-metre telescope (optical and near-infrared)	Chile
Dunn Solar Telescope (solar)	New Mexico
NSO Integrated Synoptic Program (solar)	Multiple locations worldwide

NOAO, National Optical Astronomy Observatory; NSO, National Solar Observatory; SOAR, Southern Astrophysical Research; WIYN, Wisconsin–Indiana–Yale–NOAO.

Among the possible closures, the GBT stands out: it first saw light only in 2000 and still produces high-quality science. "It's too early to be considering shutting this instrument down," says Anthony Beasley, director of the National Radio Astronomy Observatory (NRAO) in Charlottesville, Virginia, which operates the facility. "The GBT hasn't hit strong middle age yet."

SOUND OF SILENCE

The GBT is in the mountains of West Virginia, near the heart of a federally designated 'radio quiet zone', where radio broadcasts and similar transmissions are banned. Visitors must use old-fashioned pay phones to make calls out of the rural valley. The area, known as Green Bank Observatory, is ideal for radio astronomy: in 1960, astronomer Frank Drake conducted the first search for extraterrestrial intelligence using the observatory's Howard Tatel telescope.

When the GBT was completed 14 years ago, it became the premier telescope at the observatory. But it took some time to hit its stride. The circular track on which the 7,300-tonne dish sits deteriorated faster than expected, and had to be replaced in 2007. Soon afterwards, scientists extended the high-frequency end of the telescope's observing range to 100 gigahertz, at which it can probe dense gas in galaxies and interstellar molecules.

Today, the GBT is known for its wide range of wavelengths, its high angular resolution and its ability to point to 85% of the sky. Pisano uses it to map hydrogen gas within and between galaxies (see S. A. Wolfe *et al. Nature* **497,** 224–226; 2013), and pulsar astronomers use it to clock the millisecond flashes coming from spinning neutron stars. This month, a team led by the NRAO's Scott Ransom reported discovering such a millisecond pulsar. It is accompanied by two white dwarf companion stars, a rare triple system that could allow scientists to test a particular aspect of general relativity (see S. M. Ransom *et al. Nature* **505,** 520–524; 2014).

Ransom is part of a worldwide consortium of astronomers who hope that, by looking for tiny variations in the rotation rate of pulsars, they can be the first to detect long-sought gravitational waves rippling through the fabric of space and time. The North American part of that hunt relies on the GBT and the 305-metre-wide Arecibo radio telescope in Puerto Rico. Arecibo has a larger dish than the GBT but it is fixed, so it sees a smaller fraction of the sky, and follows fewer pulsars. "In my opinion, the GBT is the best pulsar telescope in the world," Ransom says.

Still, the NSF says that it cannot afford the roughly $8 million required to operate the GBT annually. It needs partners to contribute at least half of that cost, says Ulvestad — otherwise, observing time on the telescope could be cut back, or the facility could be mothballed or even dismantled.

In theory, West Virginia University could become a senior partner in managing the Green Bank site, or even take over operations itself. The state has powerful political advocates in its two Democratic senators Jay Rockefeller and Joe Manchin, who helped to steer the university's $1 million towards the GBT. More money may eventually be forthcoming, but discussions with the NSF are essentially on hold, says physicist Earl Scime, the interim associate vice-president for research at West Virginia University. The university is waiting to see which organization bids to manage the NRAO after the current agreement expires in 2015; an announcement for that competition is expected in the coming months.

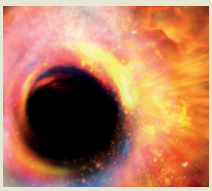
"It's too early to be considering shutting this instrument down."

For Pisano, there is little to do but wait and see. Since joining West Virginia University five years ago from the NRAO, he has spent an average of 350 hours a year observing with the GBT. He says: "Having done some great science with that telescope, I would hate to see it go." ■

VICTOR HABBICK VISIONS/SPL/GETTY

MORE ONLINE

TOP STORY



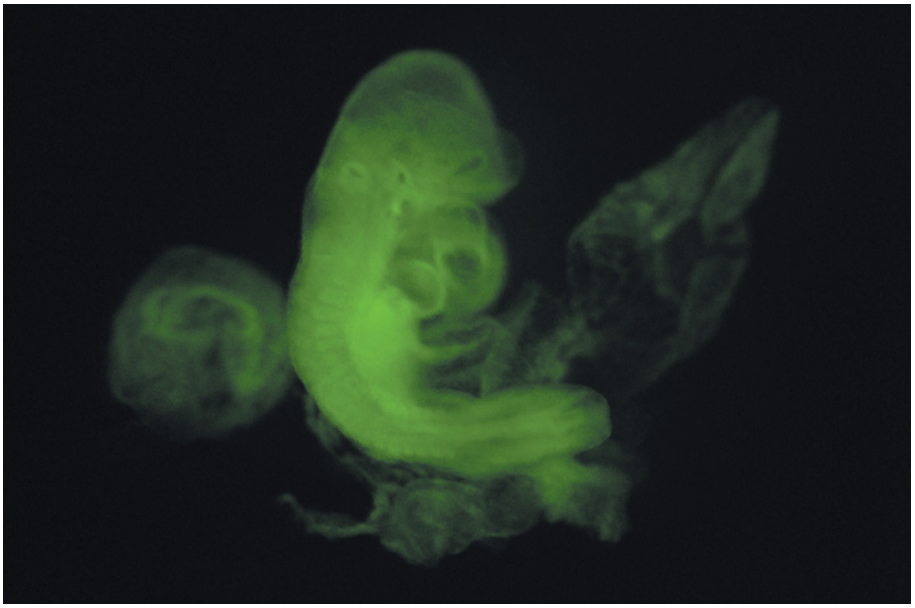
Stephen Hawking questions nature of black holes go.nature.com/iqdgti

MORE NEWS

- Contagious dog tumour traced back to its roots go.nature.com/2raakm
- Mantis shrimp's unusually complex eyes are mediocre at discerning colours go.nature.com/xxqtdw
- Drug-resistant tuberculosis evolves to divide quickly go.nature.com/i28ux5

NATURE PODCAST

Genome of a hunter-gatherer, reprogramming cells, and magnetic monopoles in cold atoms nature.com/nature/podcast



HARUKO OBOKATA

A mouse embryo injected with cells made pluripotent through stress, tagged with a fluorescent protein.

REGENERATIVE MEDICINE

Acid bath offers easy path to stem cells

Just squeezing or bathing cells in acidic conditions can readily reprogram them into an embryonic state.

BY DAVID CYRANOSKI

In 2006, Japanese researchers reported[1] a technique for creating cells that have the embryonic ability to turn into almost any cell type in the mammalian body — the now-famous induced pluripotent stem (iPS) cells. In papers published this week in *Nature*[2,3], another Japanese team says that it has come up with a surprisingly simple method — exposure to stress, including a low pH — that can make cells that are even more malleable than iPS cells, and do it faster and more efficiently.

"It's amazing. I would have never thought external stress could have this effect," says Yoshiki Sasai, a stem-cell researcher at the RIKEN Center for Developmental Biology in Kobe, Japan, and a co-author of the latest studies. It took Haruko Obokata, a young stem-cell biologist at the same centre, five years to develop the method and persuade Sasai and others that it works. "Everyone said it was an artefact — there were some really hard days," says Obokata.

Obokata says that the idea that stressing cells might make them pluripotent came to her when she was culturing cells and noticed that some, after being squeezed through a capillary tube, would shrink to a size similar to that of stem cells. She decided to try applying different kinds of stress, including heat, starvation and a high-calcium environment. Three stressors — a bacterial toxin that perforates the cell membrane, exposure to low pH and physical squeezing — were each able to coax the cells to show markers of pluripotency.

But to earn the name pluripotent, the cells had to show that they could turn into all cell types — demonstrated by injecting fluorescently tagged cells into a mouse embryo. If the introduced cells are pluripotent, the glowing cells show up in every tissue of the resultant mouse. This test proved tricky and required a change in strategy. Hundreds of mice made with help from mouse-cloning pioneer Teruhiko Wakayama at the University of Yamanashi, Japan, were only faintly fluorescent. Wakayama, who had initially thought that the project would probably be a "huge effort in vain", suggested stressing fully differentiated cells from newborn mice instead of those from adult mice. This worked to produce a fully green mouse embryo.

Still, the whole idea was radical, and Obokata's hope that glowing mice would be enough to win acceptance was optimistic. Her manuscript was rejected multiple times, she says.

To convince sceptics, Obokata had to prove that the pluripotent cells were converted mature cells and not pre-existing pluripotent cells. So she made pluripotent cells by stressing T cells, a type of white blood cell whose maturity is clear from a rearrangement that its genes undergo during development. She also caught the conversion of T cells to pluripotent cells on video. Obokata called the phenomenon stimulus-triggered acquisition of pluripotency (STAP).

The results could fuel a long-running debate. For years, various groups of scientists have reported finding pluripotent cells in the mammalian body, such as the multipotent adult progenitor cells described[4] by Catherine Verfaillie, a molecular biologist then at the University of Minnesota in Minneapolis. But others have had difficulty reproducing such findings. Obokata started the current project in the laboratory of tissue engineer Charles Vacanti at Harvard University in Cambridge, Massachusetts, by looking at cells that Vacanti's group thought to be pluripotent cells isolated from the body[5]. But her results suggested a different explanation: that pluripotent cells are created when the body's cells endure physical stress. "The generation of these cells is essentially Mother Nature's way of responding to injury," says Vacanti, a co-author of the latest papers[2,3].

One of the most surprising findings is that the STAP cells can also form placental tissue, something that neither iPS cells nor embryonic stem cells can do. That could make cloning dramatically easier, says Wakayama. Currently, cloning requires extraction of unfertilized eggs, transfer of a donor nucleus into the egg, *in vitro* cultivation of an embryo and then transfer of the embryo to a surrogate. If STAP cells can create their own placenta, they could be transferred directly to the surrogate. Wakayama is cautious, however, saying that the idea is currently at "dream stage".

Obokata has already reprogrammed a dozen cell types, including those from the brain, skin, lung and liver, hinting that the method will work with most, if not all, cell types. On average, she says, 25% of the cells survive the stress and 30% of those convert to pluripotent cells — already a higher proportion than the roughly 1% conversion rate of iPS cells, which take several weeks to become pluripotent. She now wants to use these results to examine how reprogramming in the body is related to the activity of stem cells. Obokata is also trying to make the method work with cells from adult mice and humans.

"The findings are important to understand nuclear reprogramming," says Shinya Yamanaka, who pioneered iPS cell research. "From a practical point of view toward clinical applications, I see this as a new approach to generate iPS-like cells." ■

1. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
2. Obokata, H. *et al. Nature* **505,** 641–647 (2014).
3. Obokata, H. *et al. Nature* **505,** 676–680 (2014).
4. Jiang, Y. *et al. Nature* **418,** 41–49 (2002).
5. Obokata, H. *et al. Tissue Eng. Part A* **17,** 607–615 (2011).

POLICY

EU climate targets under fire

Critics fear that Europe's proposed energy goals and emissions cuts are too soft.

BY QUIRIN SCHIERMEIER

José Manuel Barroso wore a winning smile as he took to the stage in Brussels last week to announce a plan intended to shape Europe's low-carbon energy future and revitalize the stalling international climate debate. But scientists warn that the European Commission's proposal is at the low end of what is needed to curb climate change and might burden the next generation with soaring costs.

Europe is to cut its greenhouse-gas emissions to 40% below 1990 levels by 2030, according to the commission's proposed energy and climate framework. This would double the previous 2020 ambitions of the European Union (EU), argues commission president Barroso (see 'Slow decline'). But with emissions already almost 20% down compared with 1990 levels — thanks largely to the collapse of former Eastern bloc industries and the global economic crisis — little additional effort will be required to meet that goal, say critics. And because the commission wants to replace binding national targets for renewable energy with a soft EU-wide aspirational goal, many feel that the package is regressive rather than a leap ahead in climate policies.

"A sole 40% reduction target is neither in line with what is economically feasible nor with what the science says is needed to avoid dangerous climate change," says Rebecca Harms, co-chair of the Green group in the European Parliament. The Parliament must still approve the package in a plenary vote expected next month. If it does, the measures could become binding EU legislation before the end of the year.

Although achievable at modest economic cost, the proposed 40% target might not be enough to meet the EU's longer-term ambitions of reducing emissions by at least 80% by 2050, according to a multi-model study of energy transformation pathways published last month (B. Knopf *et al. Clim. Change Econ.* **4** (suppl. 1), 1340001; 2013).

"By 2030, it is possible to achieve a 40% emission reduction using existing technologies," says Brigitte Knopf, an energy specialist at the Potsdam Institute for Climate Impact Research in Germany, who led the study. "But our models suggest that costs might rise sharply after 2040 if we do not incentivize technological innovation by clearly pricing greenhouse-gas emissions. Otherwise, the last step, from 70% or so to almost full decarbonization of the entire economy, could be hard to achieve."

The study finds that Europe can accomplish its climate ambitions without relying on controversial carbon capture and storage technology. However, substantial progress across a range of other energy technologies, such as organic solar cells or even nuclear fusion, will be needed to meet the EU's long-term goal while keeping costs in check, Knopf says.

Obstacles on the road to a low-carbon future include unresolved problems with both energy storage and long-distance power transmission, says Claudia Kemfert, a climate and energy economist at the German Institute for Economic Research in Berlin. Future efforts, she says, must also include the development of low-carbon fuels and research into possible new energy sources, such as ocean energy.

The proposed framework leaves it to EU member countries to decide how to achieve their national emissions-reduction targets in the most cost-efficient way. Some fear that this will limit the speed of transition to a renewable-energy-based economy and thwart the creation of hundreds of thousands of 'green' jobs.

> ***"It feels as if political elites are suffering the climate debate rather than engaging in it."***

"The science shows that a renewable energy target would go a long way in creating jobs and economic growth in Europe without increasing the costs of the energy system," says Jacopo Moccia, a policy director at the European Wind Energy Association in Brussels.

Indeed, a leaked internal impact assessment by the European Commission finds that a robust 2030 renewable-energy target would create substantially more jobs and economic growth than will a sole greenhouse-gas-emissions target as proposed last week.

Some EU governments share that view. In a letter sent to Brussels last month, ministers from eight countries — Austria, Belgium, Denmark, France, Germany, Ireland, Italy and Portugal — urged the commission not to abandon the successful system of mandatory national renewable targets. Firm targets, which are crucial to ensure cost-effective investments in energy systems, will, they argue, "lead to decreased dependency on energy imports and … will pave the way for an efficient planning and expansion of the European grid".

Others — including the United Kingdom, Poland and the Czech Republic — oppose that view, arguing instead that EU member states need flexibility as they try to move towards a low-carbon economy.

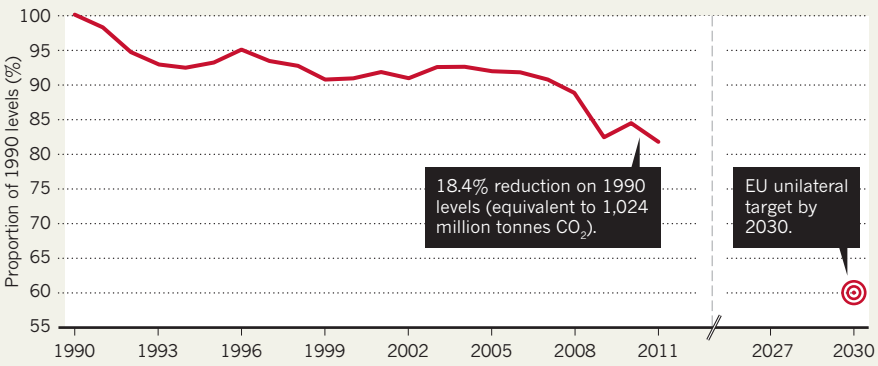
Unlike Germany, which is phasing out the use of nuclear energy by 2022, the United Kingdom, France and other EU countries do see a viable nuclear future. Poland, which relies more heavily on electricity from coal-powered plants than any other EU country, would rather abstain from greenhouse-gas reduction obligations altogether.

Even so, Barroso is confident that Europe can take a leading role in the negotiation process towards a new global climate agreement at a United Nations conference in Paris next year. But critics fear that the climate discussion has passed its sell-by date.

"Back in 2008, climate change was a top-priority issue among world leaders," says Harms. "Six years on, it feels as if political elites are suffering the climate debate rather than engaging in it." ■ SEE EDITORIAL P.585

SLOW DECLINE

The European Commission has proposed cutting EU greenhouse-gas emissions by 40% of 1990 levels by 2030.



EEA/UNFCCC

PATRICK BOYLE/GINKGO BIOWORKS

Yeast has been engineered by Ginkgo BioWorks to produce a fragrance ingredient.

BIOENGINEERING

Synthetic-biology firms shift focus

Switch to food and fragrances risks consumer rejection.

BY ERIKA CHECK HAYDEN

Plain old vanilla doesn't impress Neil Goldsmith, chief executive of Evolva, a synthetic-biology company based in Reinach, Switzerland. This year, his company will release a product that has been created by genetically modified yeast that converts sugars to vanillin. It will be the first major synthetic-biology food additive to hit supermarkets.

The product marks a shift for the industry, which has typically focused on the synthesis of drugs and commodities such as biofuels and rubber. Now, synthetic-biology companies are turning to 'fine chemicals': food and fragrance ingredients that command high prices in small batches. "The products take less time to develop, they take less money to develop, and they're much less risky," says Goldsmith.

But the products may carry a different type of hazard: consumer rejection. By creating products designed to be ingested or put on the body, synthetic-biology companies are starting to attract the attention of groups that oppose the use of genetically modified (GM) organisms. But regulations governing the use and labelling of GM organisms do not apply to fermented ingredients, because the organisms used to make them are not present in the final products.

Synthetic-biology companies are already marketing a few fine chemicals: engineered yeast has been used to make valencene and nootkatone, which provide the aroma of oranges and grapefruits, respectively, in perfumes and cosmetics. And at least five high-profile fine chemicals are scheduled to be released this year. Biofuels and commodity materials are still a mainstay, but firms are moving quickly to tap into an estimated US$20-billion market for fine chemicals, says Mark Bünger, research director at Lux Research, headquartered in Boston, Massachusetts, which tracks the industry. "We're barely scratching the surfaces of the chemicals for which we already know there are markets," he says.

"We're using a process very similar to that used to make beer."

Synthetic-biology companies have found it hard to break into established commodity markets with new biofuels and petroleum-based products, because businesses trade in high volumes and low prices. Also, the price of oil has not risen as high as some biofuels advocates had predicted. "The big challenge with making commodity chemicals is that those things are really cheap, and you have to straight-up compete on price," says Reshma Shetty, co-founder of Ginkgo BioWorks in Boston, which has signed deals with unnamed partners to make six fine-chemical ingredients. These ingredients can command prices of the order of $10–10,000 per kilogram, compared with around $1 per kilogram for biofuels.

There are other pluses. Synthetic biologists can fine-tune their product profiles to be more palatable. That is a big draw for products such as stevia, a no-calorie sweetener extracted from a leafy green plant native to South America. The sweetness comes mainly from rebaudioside compounds such as Reb A and Reb D. But the most abundant of these — Reb A — becomes bitter in large quantities, whereas the sweeter ones, such as Reb D, are present in such small amounts that it would be too expensive to extract them from stevia plants in the mass quantities needed, for example, to sweeten soft drinks. So Evolva is trying to engineer a yeast that would ferment a better-tasting stevia based on the sweeter Rebs. "What we hope this means is that you can go to having a cola product based on, let's say, Reb D, where you can get the taste right and the economics in units affordable to the consumer," Goldsmith says.

Another advantage of the bioengineering route is that these additives can be swapped for those extracted from nature and still legally be called natural because they are made by living organisms (typically, yeast). And because it is added to food after the yeast has been removed, the ingredient itself need not be labelled in any particular way. As long as it is equivalent to one of the many used in the food industry that are generally recognized as safe, it can be added to foods without any regulatory review.

How consumers will respond to these products is unclear. Already, Friends of the Earth US, an environmental group based in Washington DC, is asking consumers to sign an online petition calling for food companies not to use synthetic-biology-derived vanillin in ice cream.

Some companies are positioning themselves for the coming battle. Solazyme, based in South San Francisco, California, modifies algae to produce oils that are added to cosmetics sold by the international beauty chain Sephora. A spokesperson says that its products "are made naturally by microalgae".

Carolyn Fritz, chief executive of Allylix in San Diego, California, takes a different tack in trying to head off concerns. Her company uses yeast to make terpenes — organic chemicals that can be extracted from plants for use in fragrances and foods. She points out that one of the main synthetic-biology processes, using the fermentation powers of yeast, is something that should be familiar to thirsty consumers. "We're using a process very similar to that used to make beer, wine and lots of other products," she says. ■

THE CRYSTAL CENTURY

From aiding drug design to analysing soil samples on Mars, crystallography has helped to propel much of modern science over the past 100 years.

VIKTOR KOEN

While skiing in the Alps over Easter in 1912, the German physicist Max von Laue told his colleagues about an innovative idea: he posited that X-rays passing through a crystal would reflect off atomic centres in the lattice and interfere with each other to create a diffraction pattern. His skiing partners were sceptical, thinking that the thermal jiggling of atoms in a crystal would ruin any pattern. By June, however, von Laue's idea had been proved right; and in 1914 he was awarded the Nobel Prize in Physics "for his discovery of the diffraction of X-rays by crystals", a technique that not only elucidated the behaviour of X-rays but also allowed chemists to deduce the placement of atoms in a crystal.

Since then, X-ray crystallography has gone on to inform almost every branch of science by providing a means to understand the structure of complex molecules and materials. In this special issue, *Nature* celebrates the International Year of Crystallography by examining the impact of von Laue's method and its descendants. A graphic on page 602 summarizes the highlights and evolution of the field over the past century. Looking forwards, a News Feature on page 604 describes how a number of countries have invested in expensive X-ray free-electron lasers to crack some of the most difficult problems in crystallography. And a News & Views Forum on page 620 compares these with synchrotron X-ray sources for applications in structural biology.

Despite the enormous successes of crystallographic research, there are concerns about its future. In a Comment on page 607, physicist Paulo G. Radaelli calls for a governing body to steer the development of large international X-ray and neutron facilities. And a Careers Feature on page 711 explores how jobs in the field are evolving in ways that demand diverse skills.

Taking a historical perspective can point the way to future development. In a Comment on page 609, author Georgina Ferry reflects on how women have had leading roles in crystallography over the past century. "The features of this field that have attracted, retained and encouraged women," she writes, "have lessons to offer for the future of women's progress in science more generally." ■

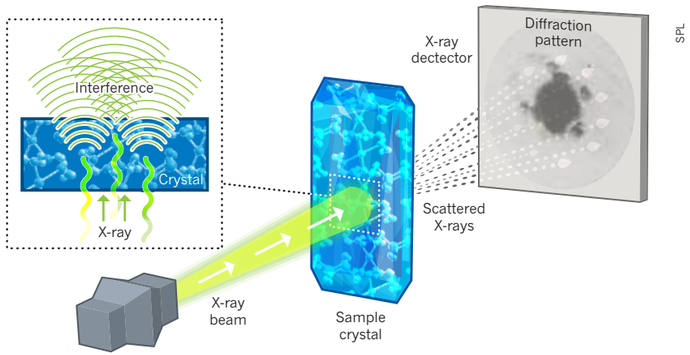
CRYSTALLOGRAPHY AT 100
A *Nature* special issue
nature.com/crystallography

ATOMIC SECRETS

100 YEARS OF CRYSTALLOGRAPHY

BY NICOLA JONES

In 1914, German scientist Max von Laue won the Nobel Prize in Physics for discovering how crystals can diffract X-rays: a phenomenon that led to the science of X-ray crystallography. Since then, researchers have used diffraction to work out the crystalline structures of increasingly complex molecules, from simple minerals to high-tech materials such as graphene and biological structures, including viruses. With improvements in technology, the pace of discovery has accelerated: tens of thousands of new structures are now imaged every year. The resolution of crystallographic images of proteins passed a critical threshold for discriminating single atoms in the 1990s, and newer X-ray sources promise images of challenging proteins that are hard or impossible to grow into large crystals.



SPL

BIRTH OF AN IDEA

Von Laue hit on the idea that when X-rays passed through a crystal, they would scatter off the atoms in the sample and then interfere with each other like waves passing through a breach in a shore wall. In some places, the waves would add to each other; in others, cancel each other out. The resulting diffraction pattern could be used to back-calculate the location of the atoms that scattered the original X-rays. Von Laue and his colleagues proved his theory in 1912 with a sample of copper sulphate.

1914 1916 1918 1920 1922 1924 1926 1928 1930 1932 1934 1936 1938 1940 1942 1944 1946 1948 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970

1913

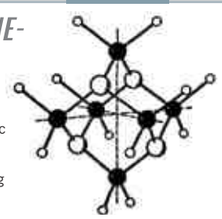
DIAMOND

Diffraction image allowed researchers to confirm the tetrahedral structure of carbon atoms in this famous crystal.

1923

HEXAMETHYLENE-TETRAMINE

The first organic molecule to be imaged, chosen because of its simple cubic symmetry. It proved that molecules, not just atoms, can make up the repeating elements of a crystal.



1925

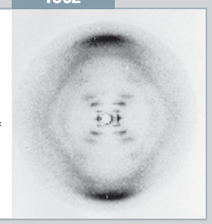
QUARTZ

The determination of the structure of silicate minerals was fundamental to the field of mineraology.

1952

DNA

Rosalind Franklin's X-ray image of DNA, known as photo 51, helped James Watson and Francis Crick to create their famous model of the double helix. An atomic-resolution image of the structure proposed in 1953 was not taken until 1980.



1958

MYOGLOBIN

The irregular folds seen in the structure of the first imaged protein were a huge surprise.

1965

LYSOZYME

The first enzyme to be imaged, sourced from hen egg whites.

1970

SYNCHROTRON

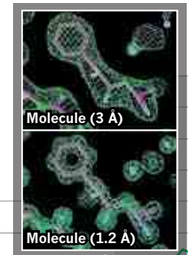
A study of insect muscle at the German Electron Synchrotron (DESY) in Hamburg was the first to use X-rays generated by a synchrotron. The use of these machines caused a boom in crystallography studies.

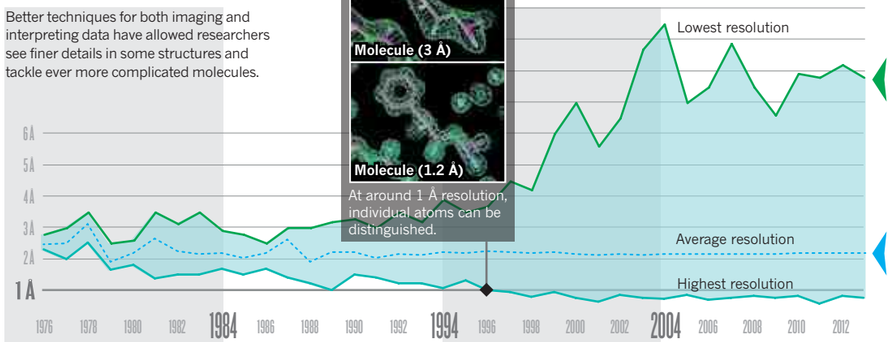
HEXAMETHYLENETETRAMINE: AM. CHEM. SOC.; DNA: KING'S COLLEGE LONDON

CRYSTALLOGRAPHY AT 100
A *Nature* special issue
nature.com/crystallography

GETTING CLEARER

Better techniques for both imaging and interpreting data have allowed researchers see finer details in some structures and tackle ever more complicated molecules.





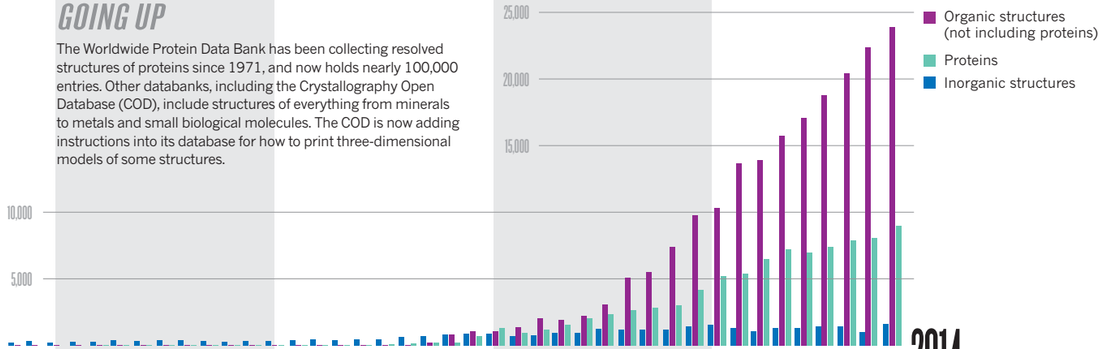
Resolution suffers in images of some complex structures, often because of variations or motion within a crystal.

The average resolution of proteins imaged through crystallography has not changed much; advances in resolution are balanced by attempts to image more complex structures.

IMAGES: BERNHARD RUPP/GARLAND SCIENCE/TAYLOR & FRANCIS
GRAPH: H. M. BERMAN *PROTEIN SCI.* **21**, 1587–1596 (2012), WITH UPDATES FROM WORLDWIDE PROTEIN DATA BANK

GOING UP

The Worldwide Protein Data Bank has been collecting resolved structures of proteins since 1971, and now holds nearly 100,000 entries. Other databanks, including the Crystallography Open Database (COD), include structures of everything from minerals to metals and small biological molecules. The COD is now adding instructions into its database for how to print three-dimensional models of some structures.



SOURCE: WORLDWIDE PROTEIN DATA BANK/ CRYSTALLOGRAPHY OPEN DATABASE

1978

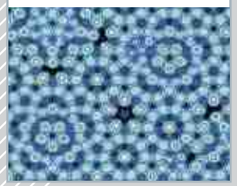
TOMATO BUSHY STUNT VIRUS

First atomic-scale image of a complete virus: in this case, a plant virus. It revealed structural rules that were found to hold true in human pathogens a few years later.

1984

QUASICRYSTALS

The first crystals were identified with atomic arrangements that do not repeat exactly, defying general wisdom about crystals.



2000

RIBOSOME

The molecular machine that assembles proteins from instructions encoded in DNA.

2009

X-RAY FREE-ELECTRON LASER

The Linac Coherent Light Source at the SLAC National Accelerator Laboratory in Menlo Park, California, went into operation, opening up a new world of imaging possibilities (see page 604).

2013

HIV TRIMER

An X-ray crystallographic image of the hook that HIV uses to bind to human cells helped to resolve a debate about what this important protein looks like.

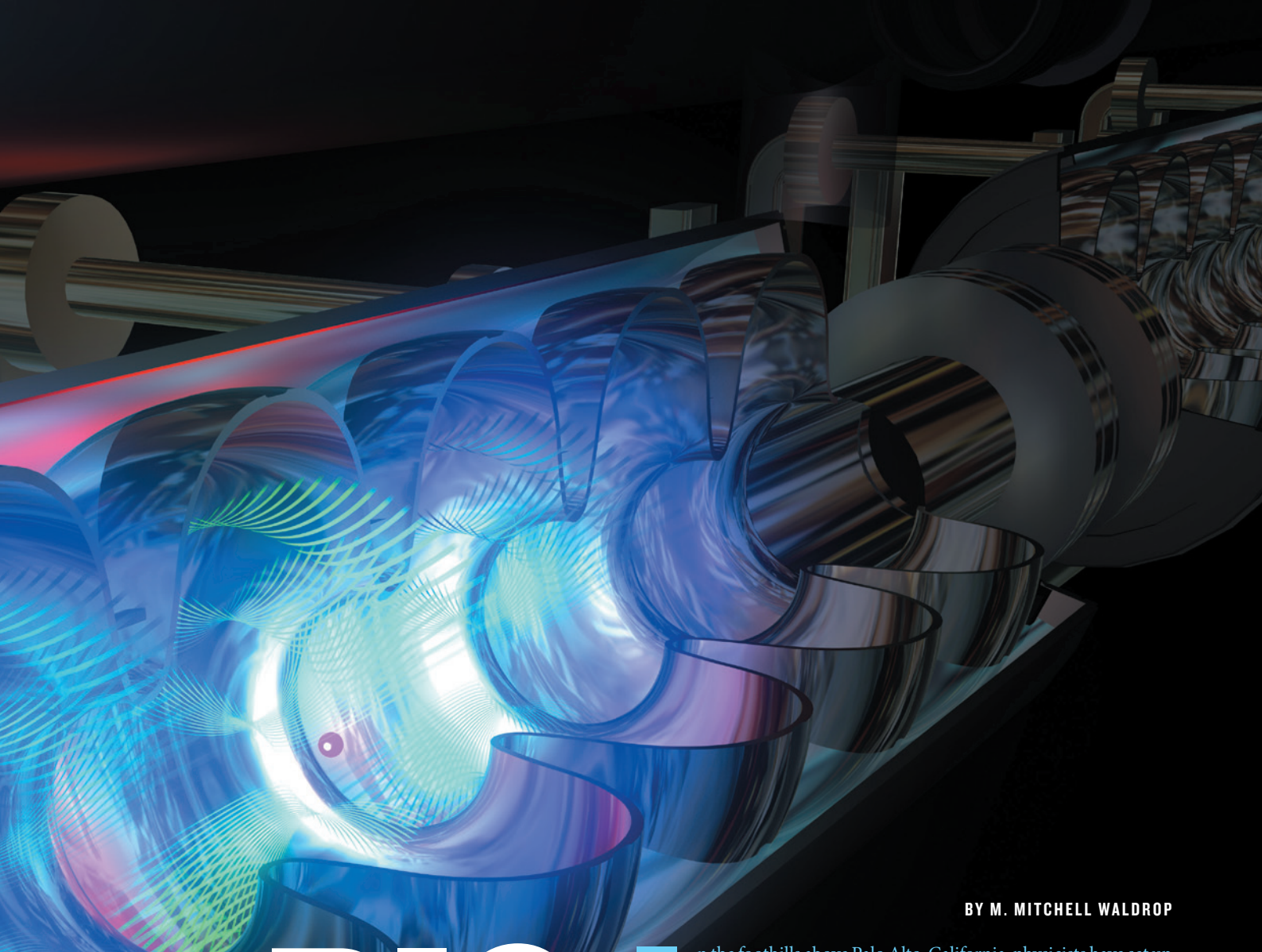
QUASICRYSTALS: US DEPT OF ENERGY/AFP/GETTY; RIBOSOME: V. RAMAKRISHNAN & D. E. BRODERSEN/MEDICAL RESEARCH COUNCIL; HIV: AAAS

THE FUTURE

The 'most wanted' list of proteins that remain to be imaged includes the massive spliceosome, which helps to organize and edit messenger RNA, and the even larger nuclear-pore complex, which serves as a nucleus's gatekeeper.

These structures can contain hundreds of proteins, making them hard to crystallize or keep still for an image.

One strategy is to crystallize bits of these structures and piece them together like a jigsaw; the use of X-ray free-electron lasers should also help.



DESY

THE BIG GUNS

Powerful X-ray lasers are getting to the heart of matter.

BY M. MITCHELL WALDROP

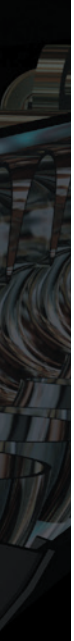
In the foothills above Palo Alto, California, physicists have set up an extreme obstacle course for some of the world's fastest electrons. First the particles are accelerated through a 3-kilometre vacuum pipe to almost the speed of light. Then they slam through a gauntlet of magnets that forces them into a violent zigzag. They respond with a blast of X-rays so fierce it could punch through steel.

But the scientists at the SLAC National Accelerator Laboratory have no interest in weaponry. Their machine, one of the world's most powerful X-ray free-electron lasers (XFELs), is a tool for studying challenging forms of matter, whether compressed to the kind of pressures and temperatures found deep inside a star, or folded into the complex tangle of a protein molecule.

Structural biologists, in particular, stand to benefit greatly from XFELs. With X-ray pulses short enough to capture strobe-like images of molecular motions, and intense enough to image the multitude of biomolecules that have defied conventional methods, XFELs are giving biologists new ways to scan for potential drug targets, to probe the mechanics of photosynthetic molecules, and more.

"XFELs, without any doubt, are disruptive technology," says Keith Moffat, a crystallographer at the University of Chicago in Illinois who has served on the SLAC machine's scientific advisory board — "an advance that is so far beyond what has gone before that it alters the way you do things".

But XFELs have also been controversial technology — especially the



Electrons are accelerated before entering an X-ray free-electron laser.

SLAC machine, known as the Linac Coherent Light Source (LCLS), which was one of the first and biggest. It was given the go-ahead by the US Department of Energy (DOE) in 2002 in the face of frequent criticism from researchers, many of whom doubted whether its scientific benefits would ever be worth its US$414-million cost — assuming that the unproven technology worked at all.

Those concerns have ebbed since the LCLS began operation in 2009, says Moffat: "This thing worked, pretty much as advertised, pretty much right out of the box, on schedule, on budget." In its wake, Japan has built its own XFEL; Europe is following with an even more capable version set to open in 2015; and others are being planned for Switzerland and South Korea. Global investments in XFELs over the next few years will total billions of dollars. But to reach their full potential, these machines will have to surmount many more technical hurdles, from boosting their power and brightness to handling the deluge of data they produce.

"Physicists, biologists, laser scientists, high-energy-density scientists — a completely new community is being formed, because you have to understand all the processes involved," says Janos Hajdu, a molecular biophysicist at Uppsala University in Sweden. "There are lots of developments that have to come together to make this work."

CORRALLING X-RAYS

The path towards XFELs began just over 100 years ago, when pioneering physicists including Max von Laue recognized the power of X-rays for studying matter (see page 602). Only photons with extremely short wavelengths can image molecules or materials at the atomic scale — roughly 0.1 nanometres, or 1 ångström.

But getting images from X-rays is tricky. There is no X-ray equivalent of a visible-light microscope, mainly because there are no good lenses for focusing the rays. So for the past century, physicists have relied on X-ray crystallography, in which they fire a beam of X-rays through a crystal lattice of identical molecules and record the resulting 'diffraction pattern' of scattered X-rays. They then work backwards from the pattern to mathematically reconstruct the original structure.

In recent decades, this has been done mostly at synchrotrons: accelerators that generate X-rays by whipping electrons around in a circle. Dozens of these light sources have grown up around the world, and they have been a boon to structural biology: the international Protein Data Bank repository currently has nearly 100,000 structures on file, most obtained from synchrotrons.

Unfortunately, many of the most scientifically interesting biomolecules, such as some membrane-bound protein complexes that mediate molecular traffic in and out of the cell, are still out of the reach of synchrotrons because they do not grow into crystals that are large enough and perfect enough to produce a usable diffraction pattern.

Yet even the most crystallization-resistant macromolecules will often form nanocrystals a few dozen molecules across. Because the beams from synchrotrons are not bright enough to get usable diffraction patterns from such structures, researchers have turned to XFELs, which are at least a billion times brighter than synchrotrons.

The basic principles of XFELs were worked out in the 1980s, building on an earlier generation of free-electron lasers that produced photons much less energetic than X-rays. In both types of laser, a beam of unconfined electrons passes through magnets that force it into an undulating path, and the beam emits photons along its line of flight. But at X-ray energies, the photons interact with the electrons in a manner that produces ferociously bright X-ray laser pulses lasting only a few femtoseconds (10^{-15} seconds) each — short enough to essentially freeze the motion of molecules in the target (see 'X-ray vision').

In 1992, Claudio Pellegrini, a physicist at the University of California, Los Angeles, and the idea's leading champion, proposed to build one of these machines at SLAC, arguing that the facility's soon-to-retire 50-GeV electron beam could be adapted to make an XFEL operating at wavelengths of 1–40 ångströms.

To the idea's many sceptics, Pellegrini admits, this was a fool's errand: no one had ever demonstrated a free-electron laser at these energies. "There was a lot of scepticism that you could really reach 1 ångström," he says.

RESEARCHERS HAD NEVER CONTEMPLATED A COMPUTATIONAL CHALLENGE OF THIS MAGNITUDE

Still, says Pellegrini, there were also many physicists around the world who thought that the idea was worth pursuing. And through experiments and simulations during the 1990s, advocates systematically built a persuasive argument that XFELs would work.

By the early 2000s, that case was solid enough for the DOE to commit to building the SLAC machine. Germany had already started to build the Free-Electron Laser in Hamburg (FLASH), a lower-energy 'soft' XFEL at the German Electron Synchrotron (DESY); and Japan and a group of European countries were initiating studies that would, a decade later, lead to their own machines.

BEFORE THE EXPLOSION

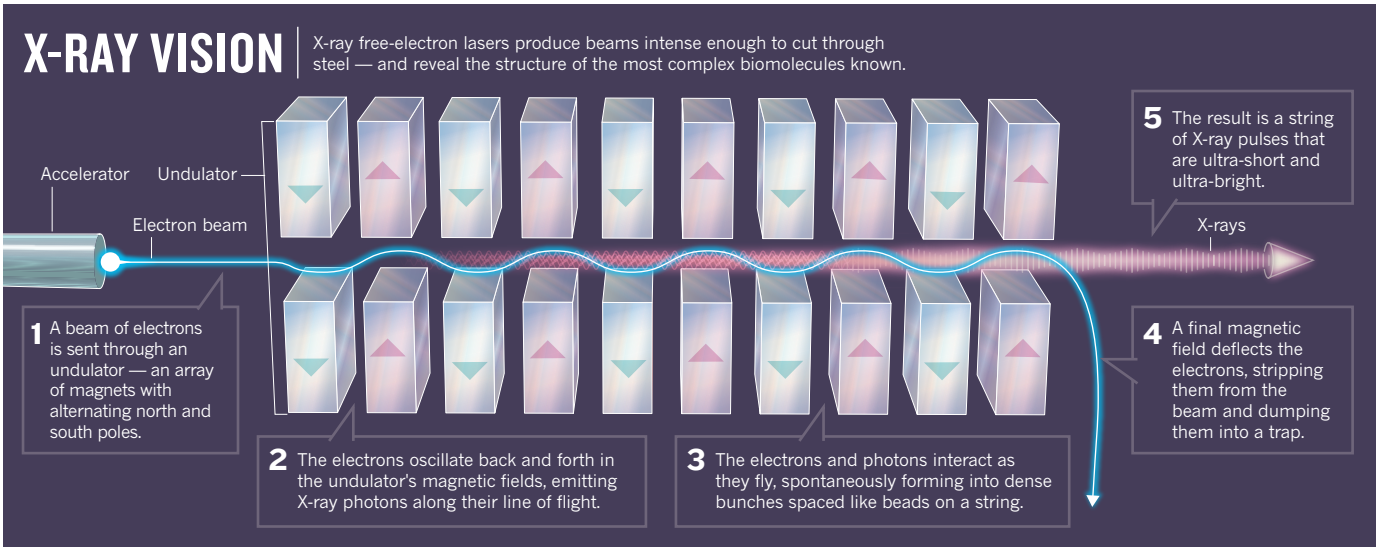
Even as the first XFELs were taking shape, however, would-be users were grappling with a seemingly intractable problem — such bright beams would destroy any sample in their path. Only in 2000 did Hajdu and his team demonstrate an escape[1]: on a femtosecond timescale, even molecular explosions unfold slowly. It takes roughly 10 femtoseconds for photons to be absorbed, molecular bonds to break and atoms to start moving from their original positions. But all the while, the photons that are not absorbed — the ones that scatter off the individual atoms and produce the diffraction pattern — are racing through the crystal at the speed of light.

The team's simulations confirming this idea, called diffract-before-destruction, were published just in time to help the DOE to make the science case for the LCLS. But that left the question of how to implement it. Unlike at synchrotrons, where large crystals of a sample can be mounted at a precise angle and measurements taken at leisure, repeatedly, at the LCLS researchers would somehow have to take nanocrystals too small to see or touch, and position them in front of X-ray pulses that would make them explode — with the machine firing 120 pulses per second.

John Spence, a physicist at Arizona State University in Tempe, took up this challenge in collaboration with Henry Chapman, a physicist now at the University of Hamburg. "Because every sample is destroyed, you have to provide new ones," says Spence. The team's solution was a device that functioned much like an ink-jet printer: it would fire tiny droplets of water across the beam in a continuous stream with the nanocrystals in suspension.

Furthermore, says Spence, because the beam would be zapping those drops and producing new diffraction patterns so often, "a few days would give you 100 terabytes of data". And each pulse would catch its nanocrystal in some unknown, random orientation, he says, so you would need to process every terabyte to reconstruct the original molecule. "This was a shocking thing to the crystallography community," says Spence: such researchers had never contemplated a computational challenge of this magnitude. Only in 2008 did Spence's student Richard Kirian work out the algorithms required to do it[2].

In late 2005, Chapman had led a team that demonstrated the technique using FLASH's longer-wavelength soft X-rays[3]. But that did not convince sceptics that it would work in a 'hard' XFEL, says Petra Fromme, a biochemist at Arizona State who was contributing her expertise in nanocrystals to the effort. "By this time, we had submitted



ten different grant proposals to investigate big membrane complexes in XFELs," she says — and had received ten rejections.

So the group, with SLAC and the DOE, had a lot of credibility at stake in December 2009, when XFEL technology, the injector and diffract-before-destruction all came together: their membrane-complex experiment would be one of the first at the newly operational LCLS. And when the computer monitors lining the walls of the tiny, underground LCLS control room suddenly started flashing twice per second with diffraction patterns, the dozens of scientists and technicians crowded inside erupted into cheers, applause and hugs. "There is extraordinary excitement that is building up around this," Chapman wrote in an e-mail that evening.

BIGGER AND BETTER

With this experiment[4] and the many that have followed, says Moffat, "the gamble was absolutely validated." Indeed, "thousands of people have been coming out of the woodwork, salivating to use this machine". In 2013 alone, the published output of the LCLS ranged from a femtosecond-scale study of how matter is affected by an intense shock wave[5] to the previously unknown structure of cathepsin B, an enzyme (and potential drug target) found in the sleeping-sickness parasite *Trypanosoma brucei*[6]. Demand for time on the machine is so high that the DOE is planning an upgrade dubbed LCLS-II, which would triple the number of simultaneously operating experimental stations by 2018.

Last November, the US National Science Foundation committed $25 million over the next five years to fund a centre for Biology with X-Ray Free Electron Lasers (BioXFEL) at the University at Buffalo in New York. With Spence as scientific director, the centre will push the technology on multiple fronts, from improving the preparation of nanocrystals to watching proteins in action as they react with other compounds.

Still, says John Tainer, a structural biologist at the Lawrence Berkeley National Laboratory in Berkeley, California, "we haven't yet shown XFEL's full potential". For example, biologists are interested in exploring structures including protein–RNA complexes, proteins that can take on many different shapes, and highly flexible functional regions that allow one molecule to interact with others. "We haven't figured out how to use XFELs to solve those problems," he says.

The good news is that the LCLS-II and a flurry of other new machines will give researchers plenty of opportunities. Since 2011, for example, Japan has been operating its SACLA XFEL in Harima. Utilizing a specially built compact accelerator, SACLA is six times brighter and slightly higher in energy than the SLAC machine. In 2015, a consortium of European research institutions expects to finish construction of the €1.15-billion (US$1.6-billion) European XFEL in Hamburg, which will be just as bright as SACLA, and a little more energetic still.

Fromme is particularly excited about the European machine's pulse rate. The LCLS's 120 pulses per second sound like a lot, she says. But the machine struggles to keep up with the nanocrystal injector, which spits out 10,000 drops per second. The European XFEL will produce 27,000 pulses per second. Not only will this allow researchers to avoid wasting more than 99% of the expensive, hard-to-make nanocrystals, but it will also allow the machine to accommodate many more users. "You could get millions of diffraction patterns in five or ten minutes, instead of five or ten hours," says Fromme.

That would allow researchers to make movies of molecular motion; in a day, they could capture images of 10,000 time steps. Right now, she says, because each frame would require looking at thousands of nanocrystals to get a full structural determination, "you'd have to go all day for each time step".

But the increase in pulse speed will work only if the system can capture and process the tsunami of data, says Fromme. The current top speed for detectors is about 3,000 diffraction patterns per second; that will have to be improved. And so will the computers, says Hajdu. "Currently, in a single experiment, one comes home with 100 terabytes of data," he says. With the European XFEL, which will produce about 2 billion pulses per day, it will be 1,000 times that. "We'll have to develop methods to reduce data on the fly to allow us to deal with it," he says.

Eventually, researchers hope to be able to get diffraction patterns from individual molecules, allowing them to watch biomolecules moving and interacting in a completely natural setting, surrounded by water, instead of trapped in the artificial environment of a crystal. "That's my future vision for crystallography," says Fromme. "Get away from being a coroner imaging dead molecules, and instead get molecular movies."

What makes this hard is that an isolated molecule does not have a host of identical twins to help it to scatter the incoming photons, as happens in a crystal. The only way to compensate is to hit it with a lot more photons to produce a stronger diffraction pattern — a flux between 1,000 and 10,000 times brighter than the current LCLS.

The European XFEL will be only about a factor of ten brighter, says Fromme. "So there are new challenges on the physics side to increase beam brightness." Still, the LCLS upgrade is intended to get close, boosting brightness by a factor of 1,000. Fromme sees the goal in sight: "I'm optimistic that we could get there in ten years." ■

M. Mitchell Waldrop *is an editor with* Nature *in Washington DC.*

1. Neutze, R., Wouts, R., van der Spoel, D., Weckett, E. & Hajdu, J. *Nature* **406,** 752–757 (2000).
2. Kirian, R. A. *et al. Opt. Express.* **18,** 5713–5723 (2010).
3. Chapman, H. N. *et al. Nature Phys.* **2,** 839–843 (2006).
4. Chapman, H. N. *et al. Nature* **470,** 73–77 (2011).
5. Milathianaki, D. *et al. Science* **342,** 220–223 (2013).
6. Redecke, L. *et al. Science* **339,** 227–230 (2013).

COMMENT

HENNING LARSEN ARCHITECTS/COB ARCHITECTS

The European Spallation Source will be built in Lund, Sweden (artist's impression).

Crystallography needs a governing body

Planning for large facilities should incorporate the views of all crystallographers, says **Paolo G. Radaelli**.

There is much to cheer as the celebrations for the International Year of Crystallography begin. Since modern crystallography dawned with X-ray diffraction experiments on crystals by Max von Laue in 1912 and William and Lawrence Bragg (a father and son team) in 1913, and was recognized by Nobel prizes in physics for von Laue in 1914 and the Braggs in 1915, the discipline has informed almost every branch of the natural sciences.

Aeroplanes fly safely because crystallography tests computer models of materials under stress. Drugs are more potent because crystallographers can see and modify how molecules interact with target sites in cells. An X-ray diffraction instrument on NASA's Curiosity rover is now even studying the mineralogy of Mars.

Yet the very strengths of the field — its size and diversity — could prove to be its downfall within a decade. Crystallography is increasingly focusing its resources on large multidisciplinary facilities, such as powerful X-ray and neutron sources. And too few researchers are involved in making decisions about these. As a result, national and local interests are being put ahead of science.

Crystallographers should take a lesson from particle physicists and create a body run by scientists for the governance of large international X-ray and neutron facilities. It should be guided by input from regular meetings of researchers from across the scientific community. This will ensure that the next generation of infrastructure will have the strongest possible scientific case, articulated clearly.

Crystallographers have a raft of methods at their disposal. Von Laue scattered X-ray photons from atoms. Now experimenters can also bombard crystal lattices with electrons and neutrons, and exploit properties such as the polarization of photons and neutrons and their interactions with magnetic fields.

It is still possible to conduct world-class crystallographic research in the laboratory. Materials scientist Dan Shechtman's 2011 Nobel-prizewinning discovery of quasicrystals — metallic alloys that organize themselves in a way that was thought to be forbidden by crystallographic theory — required only an electron microscope of the kind found in most physics and chemistry departments. But increasingly, large national and international synchrotron and neutron source facilities are used to produce the powerful photon or particle beams needed for the most demanding experiments, such as detailed studies of complex macromolecules.

Structural biology has seen particularly exciting progress in the past decade, culminating in the structure of the ribosome, the complex molecular machine that builds proteins from DNA. Driven by the desire to establish the structures of proteins and other biological molecules that are difficult to crystallize, many countries are busy building a new generation of intense X-ray sources — free-electron lasers.

The enormous investments involved ▶

— more than €1.1 billion (US$1.5 billion) for the European X-ray Free Electron Laser (XFEL) — have been justified by the hope of illuminating protein nanocrystals, or even single molecules, with intense femtosecond pulses, to capture diffraction patterns before the radiation destroys the structures[1]. Such 'molecular movies' will benefit other fields, such as the study of matter at extreme temperatures and pressures, and high-temperature superconductivity.

UGLY POLITICS

But the enormous political and financial stakes attached to this infrastructure are not matched by the scientific governance necessary to define clear research priorities. As a result, individual crystallographers are disenfranchised and have little control over their future 'means of production'.

Particle physicists, by contrast, chart the future of their discipline[2] through an open process designed to get maximum input from the community on a frequent basis. The May 2013 European Strategy for Particle Physics, for example, was drafted at an open symposium in Krakow, Poland, in September 2012 and was coordinated by CERN, Europe's high-energy physics laboratory near Geneva, Switzerland. Such processes allow particle physicists to present a common road map that is scientifically robust enough to withstand political pressures and adverse funding decisions.

Crystallography's lack of an international equivalent to CERN has left national institutes and councils pushing their vested interests. The community's priorities are unclear and its calls for funding are fragmented.

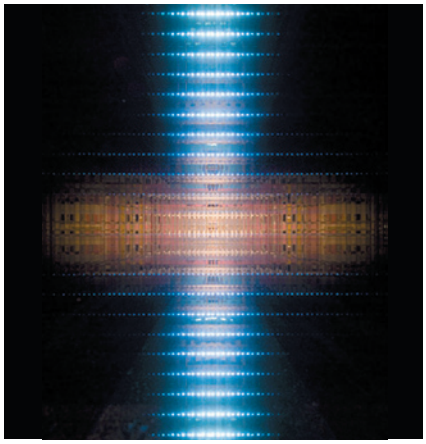
The effects of politics trumping science are being felt everywhere. In the United States, rivalry between national laboratories, state politics and a tendency by the Department of Energy to underfund instrumentation are widely believed to have hampered flagship facilities such as the Advanced Photon Source X-ray synchrotron in Argonne, Illinois, and the Spallation Neutron Source in Oak Ridge, Tennessee.

In the United Kingdom, the ISIS neutron source in Oxfordshire — one of the most innovative and productive neutron facilities worldwide — is run as a national facility, albeit with 25% of the beam time allocated to international proposals. In recent years, UK research councils have left it idle for one-third of the available operating time to save 10% of its energy costs[3]. This is a disgrace. An international review of ISIS concluded that a 10% increase in its budget would increase productivity by almost 50% (see go.nature.com/3xmjab). Overseas colleagues rarely push their funding bodies for further internationalization of ISIS, instead muttering about the United Kingdom's poor record as a European partner.

For future large-scale facilities, the spotlight is firmly on Europe. The European XFEL, under construction near Hamburg, Germany, will from 2015 deliver short X-ray pulses of around 100 femtoseconds with wavelengths of 0.05–6 nanometres. XFEL will be perfect for studying extremely small crystals of macromolecules and perhaps even single biological molecules. It will also be able to investigate the mechanisms underlying high-temperature superconductivity by taking snapshots of these materials at femtosecond timescales that are relevant for their electronic and magnetic excitations (see pages 604 and 620).

The €1.5-billion European Spallation Source (ESS), for which construction is about to begin near Lund, Sweden, is set to become the most powerful spallation neutron source in the world when it opens in 2019. Neutrons are insightful probes of many material properties, and are complementary to X-rays, but they are in short supply compared to photons, and neutron experiments take much longer. An increase in flux and brilliance over existing sources, coupled with adequate resolution, would enable neutron crystallography on small enzyme crystals, for example.

These large, international facilities are funded by multilateral agreements between governments. The European XFEL will be run as a non-profit company, with the German Electron Synchrotron centre (DESY) as the only shareholder, and 12 European nations currently contribute to the construction and operation costs. The ESS is a public company owned by the Swedish and Danish governments with, at present, 17 partner countries but no final commitment to the construction phase.



STFC

A view down the beam guides at the ISIS neutron source, UK.

European large-infrastructure road maps are drawn up by organizations — such as the European Commission's European Strategic Forum for Research Infrastructures — that are populated by political appointees[4]. Although these delegates are usually excellent scientists, they mainly represent their governments rather than the research community.

For example, when funding difficulties plagued the European XFEL, following the withdrawal of the United Kingdom in 2009 and reductions in the contributions of other partners[5], structural scientists had no way to rush to the project's defence despite the fact that the vast increase in brilliance, by several orders of magnitude, guarantees that XFEL will facilitate discoveries. Fortunately, other countries, including the Russian Federation, have plugged XFEL's €150-million financial hole, and signs are good for a renewed UK participation.

By comparison, the future of European neutron scattering looks precarious. With many smaller sources scheduled to close in the next few years, a lot of budgetary eggs are being put in the ESS basket. But, unlike XFEL, the record power of the ESS is no guarantee of exceptional performance.

The power gain of the ESS, driven mostly by longer pulses (of around 2 milliseconds instead of 10–300 microseconds), improves on existing sources by less than a factor of five in most cases. Longer pulses mean more complex instruments, so scientists will need ingenuity to translate its potential from paper to reality.

Breakthroughs will follow only if the ESS gets the source and instrument design, sample environment, software, support model and staff profile right. Yet the site and scope of the ESS were finalized by a small team long after the last open scientific debate of the technical and scientific cases (ref. 6).

Long pulses are good for many applications in soft matter and biology. But they would be hard pressed to deliver the high resolutions necessary to determine the structures of complex biological molecules, for example. With the running cost of the ESS estimated at €150 million per year, any serious blunder will be a disaster from which European neutron science may never recover.

To maximize scientific output, it is imperative that the ESS and other neutron sources in Europe are used round the clock and fitted with the most potentially ground-breaking instruments. A string of early breakthroughs will persuade the paymasters that neutron science is worth the high investments. And success will encourage them to fund upgrades to ISIS and the Laue–Langevin Institute neutron source in Grenoble, France.

A SCIENTIFIC ROAD MAP

Particle physicists know what the 'next big things' are for them — an upgrade of the Large Hadron Collider (LHC) at CERN and an electron–positron linear collider. Crystallographers should state their priorities with the same confidence.

The first step is for users of multidisciplinary facilities to muster existing

bodies, such as the International Union of Crystallography (IUCr) and the European Neutron Scattering Association (ENSA), to establish and present the community view. These organizations should commission independent scientific and technical reviews, similar to the US Astronomy and Astrophysics Decadal Survey, and make recommendations for future projects.

Although this approach may be adequate to coordinate road maps for national facilities of the scale of ISIS, higher-level political power play is necessary for multinational facilities such as the European XFEL and the ESS. An international organization of facility users, with the political muscle of CERN, should be set up urgently to provide governance, mediate with national and international political bodies, and implement community decisions.

In fact, it is questionable whether the multilateral funding model for the largest international facilities is still fit for purpose. With its reputation for excellence, the European Research Council could become the primary funder for the next generation of European facilities, with a suitable increase in its budget (currently €13.1 billion for 2014–20). Extra contributions would come from the host nations, as for the LHC, and other international partners. Such a radical change will not happen immediately, but these ideas should be discussed ahead of the renewal of the European Union Framework Programme for Research and Innovation in 2020.

The 23rd IUCr Congress and General Assembly in Montreal, Canada, in August will provide plenty of opportunities to celebrate the past triumphs of crystallography. It would also be wise for the community to use the occasion to start discussions about securing the field's future. ■

Paolo G. Radaelli *is professor of experimental philosophy and head of condensed-matter physics at the University of Oxford, UK.*
e-mail: p.g.radaelli@physics.ox.ac.uk

1. Cavalleri, A. *Nature* **448,** 651–652 (2007).
2. Lockyer, N. *Nature* **504,** 367–368 (2013).
3. House of Lords Select Committee on Science and Technology *2nd Report of Session 2013–14: Scientific Infrastructure* (Authority of the House of Lords, 2013).
4. European Commission *ESFRI Roadmap for Research Infrastructures* (EC, 2006, 2008 2010); available at http://go.nature.com/ip6afc.
5. Banks, M. 'Funding shortfall hits European X-ray laser facility' *Physics World* (29 July 2011); available at http://go.nature.com/ojuhdg.
6. Kjems, J., Taylor, A. D., Finney, J. L., Lengeler, H. & Steigenberger, U. *ESS: A Next-Generation Neutron Source for Europe* (ESS Council, 1997).

CHARLES HEWITT/PICTURE POST/GETTY

Pioneer: Kathleen Lonsdale was one of the first women to be elected to the Royal Society.

Women in crystallography

Georgina Ferry celebrates the egalitarian, collaborative culture that has so far produced two female Nobel prizewinners.

"It takes a very special breed of scientist to do this work … it is an area of science in which women dominate." So said the professor introducing distinguished British crystallographer Judith Howard in 2004 as she received an honorary degree from the University of Bristol, UK.

Some 15 years previously, Howard had received an invitation to apply for a new chair in structural chemistry at Durham University, UK, framed in similarly irksome terms: "because aren't women supposed to be good at that sort of thing?" Her former PhD supervisor, the Nobel prizewinner Dorothy Hodgkin, encouraged Howard not to let such comments get in her way. Howard got the job, established one of the world's leading laboratories for low- and variable-temperature structural chemistry, served as head of the department of chemistry, was elected a Fellow of the Royal Society and became the founding director of Durham's interdepartmental Biophysical Sciences Institute.

Whatever their level of distinction, female crystallographers have always in fact been in the minority. But there is a relationship between the outstanding achievements of some of them and the reputation and culture of the field that is worth examining as we celebrate the International Year of Crystallography. I would argue that the features of this field that have attracted, retained and encouraged women have lessons to offer for the future of women's progress in science more generally. ▶

Women were among crystallography's earliest pioneers. William Bragg, co-discoverer of X-ray crystal analysis with his son Lawrence a century ago, recruited Kathleen Lonsdale to his laboratory in 1922. Working at the Royal Institution in London, she confirmed the structure of the benzene ring, carried out studies of diamond, was one of the first two women to be elected to the Royal Society (in 1945), and was appointed the first female tenured professor at University College London.

Hodgkin was one of several women who joined the lab of the physicist John Desmond Bernal (a former Bragg student) in Cambridge, UK, in the 1930s, and with him she took the first X-ray photographs of crystalline proteins. Her solutions of the structures of penicillin and vitamin B12 won her the Nobel Prize in Chemistry in 1964. Of the four women who have won the chemistry Nobel, two were crystallographers: Hodgkin and the Israeli scientist Ada Yonath, who was awarded the prize in 2009.

Rosalind Franklin is chiefly remembered for taking the X-ray photograph of a DNA fibre that proved instrumental to James Watson and Francis Crick's Nobel-prizewinning discovery of the double helix. In her short life (she died of cancer in her 30s), she also carried out important structural studies of carbon in coal and graphite, and of plant and animal viruses.

Isabella Karle of the United States Naval Research Laboratory developed an experimental approach to using 'direct methods' of structural analysis for the solution of molecules smaller than 1,000 atoms. Her application of this statistically based technique for estimating the phases of the X-ray reflections enormously expanded the range of substances that could be tackled. Yet only her husband Jerome shared the 1985 Nobel Prize in Chemistry with Herbert Hauptman, for developing the theoretical underpinnings of the method. Other prize-giving bodies have showered Isabella with awards in her own right.

FIRST AMONG EQUALS

Women's names adorn many of the textbooks and research resources in the field. Lonsdale edited the *International Tables for Crystallography* for many years. These volumes provide information on crystal lattices, symmetry and space groups, as well as mathematical, physical and chemical data on structures. Olga Kennard of the University of Cambridge founded and ran the Cambridge Crystallographic Data Centre, an internationally recognized source of structural data on small molecules, from 1965 until 1997. Jenny Pickworth Glusker of the Fox Chase Cancer Center in Philadelphia, Pennsylvania, co-authored *Crystal Structure Analysis: A Primer*, first published in 1971 and now in its third edition (2010). Eleanor Dodson of the University of York, UK, who began as Hodgkin's technician, was the main instigator behind CCP4, the collaborative computing project that currently shares more than 250 software tools with protein crystallographers worldwide.

But the widespread assumption that these illustrious figures reflect a predominance of women in the field is false. More than two decades ago, the US mathematical crystallographer Maureen Julian of the Virginia Polytechnic Institute and State University (Virginia Tech) in Blacksburg tallied the entries in the *World Directory of Crystallographers* and found that the proportion of women was 14% internationally (and slightly lower in the United States)[1]. At the time, only 2% of the members of the American Physical Society were women; Julian concluded that a percentage in double figures gave the impression that the field was "saturated with women".

"Women's names adorn many of the textbooks and research resources in the field."

Today, the International Union of Crystallography's online list of eminent crystallographers (go.nature.com/g5iarg) is more than 90% male. Its prestigious Ewald Prize, awarded triennially since 1987, has had one female recipient (Dodson) out of 14 (7%).

A COLLABORATIVE ETHOS

There are grounds, however, for believing that the field of crystallography was unusually welcoming to women at its foundation a century ago, at least by comparison with other branches of physical science. In her 1990 study[1], Julian also traced a scientific genealogy starting with the Braggs, through colleagues both male and female, to a total of 50 female crystallographers. Bragg protégés such as Lonsdale and Bernal and their students fostered diverse and egalitarian lab cultures.

That pedigree could now be greatly extended. For example, the British protein crystallographer David Phillips worked with Lawrence Bragg at the Royal Institution from 1955 to 1966. Phillips recruited Louise Johnson as a PhD student, and when he moved to the University of Oxford, UK, in 1966, she went with him. There, the Phillips group worked alongside Hodgkin and her international, gender-balanced and left-leaning team. In 1990, Johnson succeeded Phillips as professor of structural biology, and from 2003 to 2008 was also director of life sciences at the Diamond Light Source, the United Kingdom's national synchrotron facility.

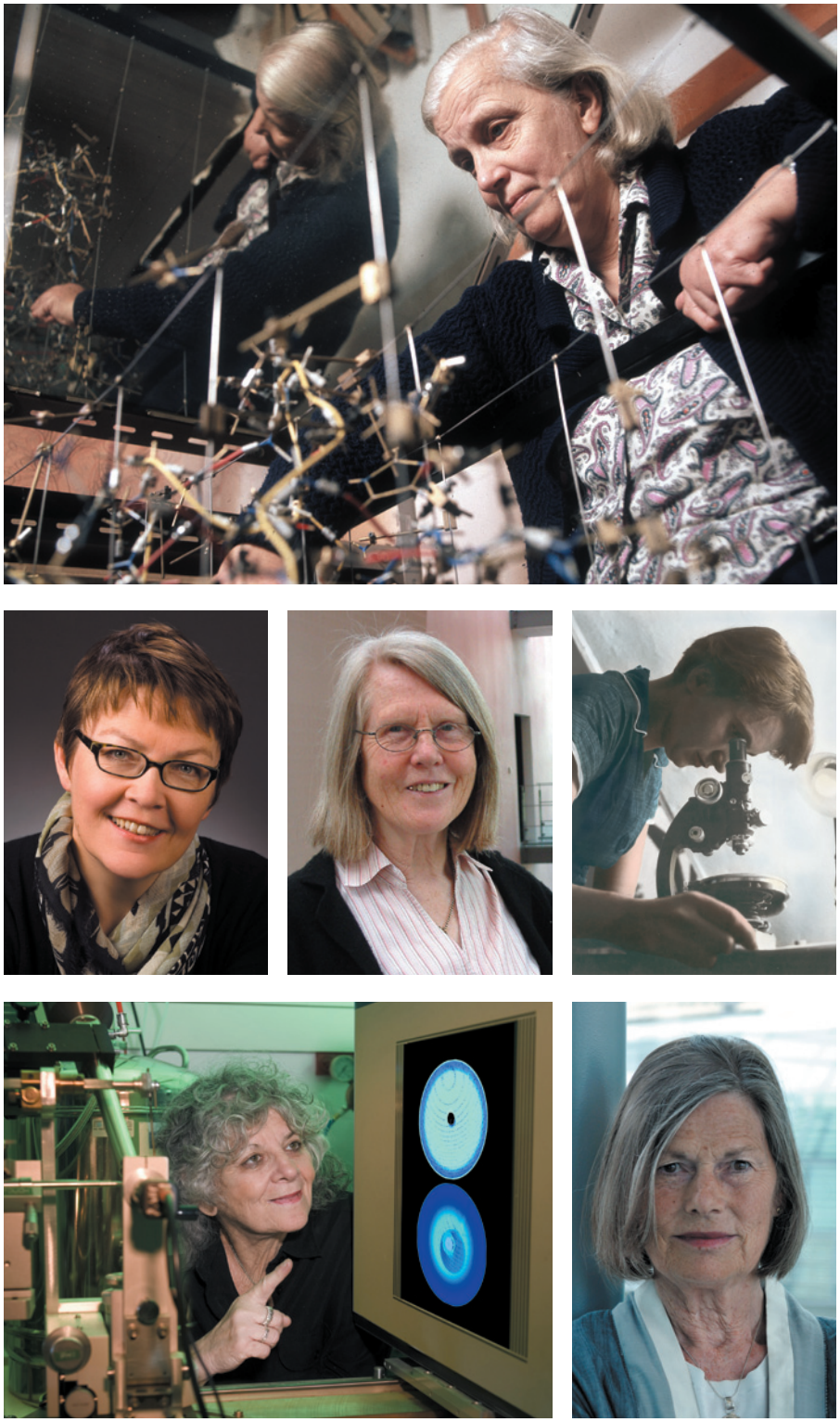
Susan Lea is professor of microbiology in the Sir William Dunn School of Pathology at Oxford. She did her PhD there in the late 1980s with structural biologist Dave Stuart. It did not occur to her to look for a female role model, because she was surrounded by them. "Louise [Johnson] was the head of structural biology, and there were a lot of women in biophysics at Oxford," says Lea, "so I never really thought about it." She remembers "a very good atmosphere", citing the normality of children being around in the lab: "It was expected that if you were bright you would get the job done." In 1995, having completed her PhD on structural studies of the foot-and-mouth disease virus, Lea was one of the first to receive a Dorothy Hodgkin fellowship from the Royal Society, designed to allow some flexibility around family and other commitments for early-career scientists. Her first child was born a year later.

That collaborative ethos owes as much to the nature of the science as to the benevolent legacy of the Braggs. "It's a science that, when practised well, tends to involve six to eight disciplines," explains Lea. "One minute I'm talking to a virologist, the next a crystallographer, the next an immunologist." For Hao Wu, who uses structural techniques to study innate immunity at Harvard Medical School in Boston, Massachusetts, this interdisciplinarity was a prime attraction. "I had no clue about other women crystallographers when I started," she says. "What attracted me was that it had mathematics, physics and biology in it."

As she neared her graduation in medicine in Beijing, Wu was fascinated by a lecture from Michael Rossmann, visiting from Purdue University in West Lafayette, Indiana. Rossmann is a mathematical crystallographer who met Lonsdale as a schoolboy, did a PhD with J. M. Robertson (a former Bragg student), and worked with molecular biologist Max Perutz (a student of Bernal's) on the solution of the haemoglobin structure. Wu subsequently secured a PhD place in Rossmann's lab. Structural analysis is "like a detective story", she says. "There is no direct path from diffraction to structure," so it appealed to her in requiring a broad range of skills, from growing and mounting crystals to computer analysis.

EVOLVING FIELD

One downside of crystallography's reputation as a technical discipline, and one sometimes perceived to be 'women's work', is that for a while, other scientists (particularly chemists) saw it as a laboratory service, and not a science in its own right. When Hodgkin's team at Oxford finally solved the complex structure of vitamin B12 in 1955 (ref. 2), the result was trumpeted in *The New York Times* as the work of Alexander Todd at the University of Cambridge, UK, whose chemical analyses of B12 were published[3] in *Nature* back to back with an earlier paper by the Oxford team[4]. Todd also gave the first



D.H.: MONDADORI PORTFOLIO/GETTY; I.S.: HEIDELBERG UNIV.; E.D.: JOHN OLIVE/UNIV. YORK; R.F.: SCIENCE SOURCE/SPL; A.Y.: MICHELINE PELLETIER/CORBIS; L.J.: ANNE PURKISS/THE ROYAL SOCIETY

Top: Dorothy Hodgkin. Middle, left to right: Irmgard Sinning; Eleanor Dodson; Rosalind Franklin. Bottom: Ada Yonath (left); Louise Johnson (right).

talk on the structure at the 1955 meeting of the Chemical Society at the University of Exeter, UK — Hodgkin stood up at the end to make clear exactly who had done what.

Glusker, who as Hodgkin's postdoc undertook the analysis of a key derivative that broke the back of the problem, remembers how indignant they all were (see go.nature.com/o74dse). "We thought he viewed us just as technicians and did not realize the amount of thought that went … into devising which electron-density maps to draw, which parameters to refine and how to do this."

Modern crystallography is now very different. Much of the trial-and-error process has gone, because almost all the stages of X-ray crystal analysis have been automated, turning the spotlight onto the meaning and relationships of structural features, rather than the structures themselves. "Nowadays it's not possible to publish a structure on its own in a high-impact journal," says Irmgard Sinning, professor of biochemistry and structural biology at the Heidelberg University Biochemistry Center in Germany.

Sinning studies protein-targeting systems, and has just been announced as a 2014 recipient of Germany's top research award, the Gottfried Wilhelm Leibniz Prize. "Crystallography has developed tremendously in the past two decades," she says, "and solving the structure often takes less time than working out the molecular mechanism."

She now recruits more biochemists than chemists to her lab, about half of them women. Overall, the number of women in crystallography is climbing. I analysed speaker lists from various science meetings, and found that at the European Crystallography Meeting in August 2013, 27% of the speakers were women. This compares with around 21% at the 2013 European Physical Society Conference on High Energy Physics, and 43% at the 2013 International Congress of Immunology. The numbers of women entering research careers are increasing across the physical and life sciences, with most in biomedical fields. But recent evidence suggests that they still have a harder time than their male colleagues in making it to the top (see, for example, www.nature.com/women).

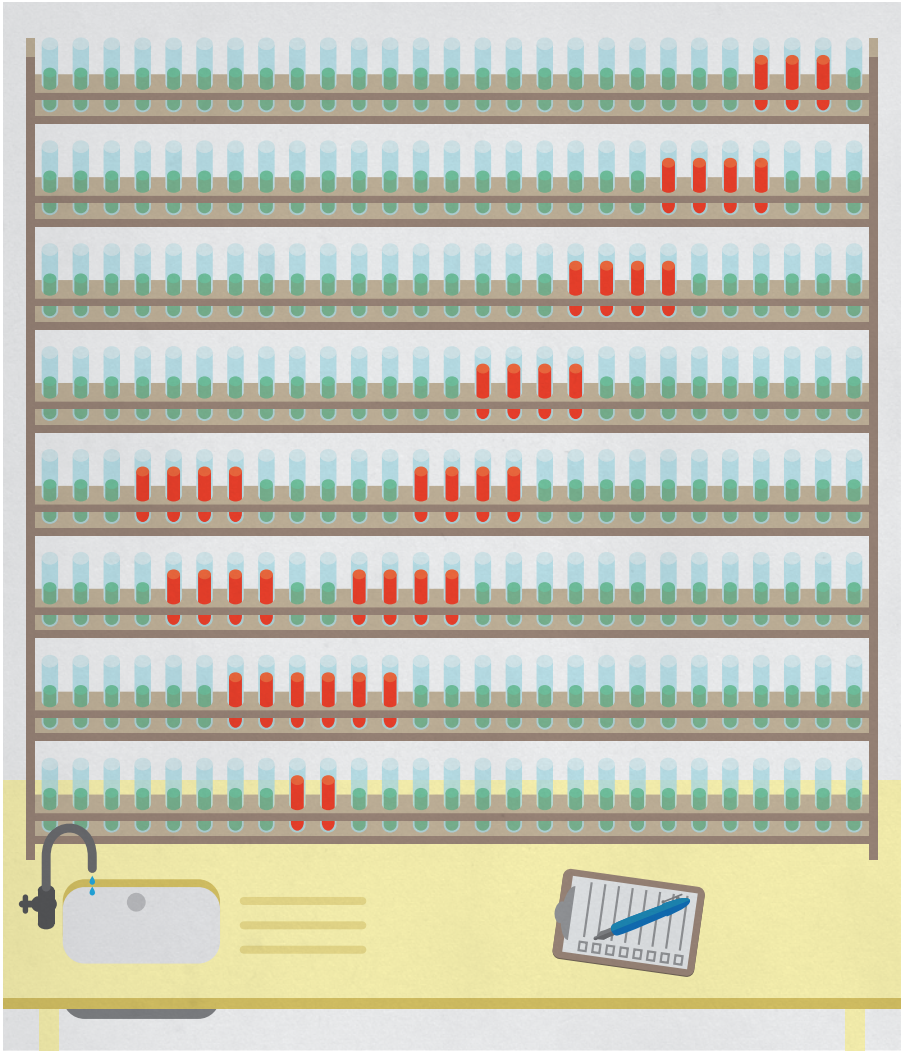
BROKEN SYMMETRY

"Today it's more demanding to balance a family and a career," says Wu, who has recently been appointed to a chair. Howard, too, is concerned: "There's a drop-off at postdoc level and beyond." Sinning urges younger women to have more confidence in themselves when applying for promotions: "Just do it! A guy would never say 'Am I good enough?' — they automatically think they are."

Crystallography has shining examples of successful women who inspire and support younger colleagues. But junior scientists still face too many obstacles in their progression through the ranks. Perhaps an important goal for this International Year of Crystallography would be to ensure that the Braggs' legacy of equal opportunities is replenished. ■

Georgina Ferry *is a science writer based in Oxford, UK.*
e-mail: mgf@georginaferry.com

1. Julian, M. M. in *Women of Science: Righting the Record* (eds Kass-Simon, G. & Farnes, P.) 335–383 (Indiana Univ. Press, 1990).
2. Hodgkin, D. C. *et al. Nature* **176,** 325–328 (1955).
3. Cannon, J. R., Johnson, A. W. & Todd, A. R. *Nature* **174,** 1168–1169 (1954).
4. Brink, C. *et al. Nature* **174,** 1169–1171 (1954).



CHRIS RYAN/*NATURE*

NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring[1,2]. As leaders of the US National Institutes of Health (NIH), we share this concern and here explore some of the significant interventions that we are planning.

Science has long been regarded as 'self-correcting', given that it is founded on the replication of earlier work. Over the long term, that principle remains true. In the shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised the ability of today's researchers to reproduce others' findings.

Let's be clear: with rare exceptions, we have no evidence to suggest that irreproducibility is caused by scientific misconduct. In 2011, the Office of Research Integrity of the US Department of Health and Human Services pursued only 12 such cases[3]. Even if this represents only a fraction of the actual problem, fraudulent papers are vastly outnumbered by the hundreds of thousands published each year in good faith.

Instead, a complex array of other factors seems to have contributed to the lack of reproducibility. Factors include poor training of researchers in experimental design; increased emphasis on making provocative statements rather than presenting technical details; and publications that do not report basic elements of experimental design[4]. Crucial experimental design elements that are all too frequently ignored include blinding, randomization, replication, sample-size calculation and the effect of sex differences. And some scientists reputedly use a 'secret sauce' to make their experiments work — and withhold details from publication or describe them only vaguely to retain a competitive edge[5]. What hope is there that other scientists will be able to build on such work to further biomedical progress?

Exacerbating this situation are the policies and attitudes of funding agencies, academic centres and scientific publishers. Funding agencies often uncritically encourage the overvaluation of research published in high-profile journals. Some academic centres also provide incentives for publications in such journals, including promotion and tenure, and in extreme circumstances, cash rewards[6].

Then there is the problem of what is not published. There are few venues for researchers to publish negative data or papers that point out scientific flaws in previously published work. Further compounding the problem is the difficulty of accessing unpublished data — and the failure of funding agencies to establish or enforce policies that insist on data access.

PRECLINICAL PROBLEMS

Reproducibility is potentially a problem in all scientific disciplines. However, human clinical trials seem to be less at risk because they are already governed by various regulations that stipulate rigorous design and independent oversight — including randomization, blinding, power estimates, pre-registration of outcome measures in standardized, public databases such as ClinicalTrials.gov and oversight by institutional review boards and data safety monitoring boards. Furthermore, the clinical trials community has taken important steps towards adopting standard reporting elements[7].

Preclinical research, especially work that uses animal models[1], seems to be the area that is currently most susceptible to reproducibility issues. Many of these failures have simple and practical explanations: different animal strains, different lab environments or subtle changes in protocol. Some irreproducible reports are probably the result of coincidental findings that happen to reach statistical significance, coupled with publication bias.

Another pitfall is overinterpretation of creative 'hypothesis-generating' experiments, which are designed to uncover new avenues of inquiry rather than to provide definitive proof for any single question. Still, there remains a troubling frequency of published reports that claim a significant result, but fail to be reproducible.

PROPOSED NIH ACTIONS

As a funding agency, the NIH is deeply concerned about this problem. Because poor training is probably responsible for at least some of the challenges, the NIH is developing a training module on enhancing reproducibility and transparency of research findings, with an emphasis on good experimental design. This will be incorporated into the mandatory training on responsible conduct of research for NIH intramural postdoctoral fellows later this year. Informed by this pilot, final materials will be posted on the NIH website by the end of this year for broad dissemination, adoption or adaptation, on the basis of local institutional needs.

Several of the NIH's institutes and centres are also testing the use of a checklist to ensure a more systematic evaluation of grant applications. Reviewers are reminded to check, for example, that appropriate experimental design features have been addressed, such as an analytical plan, plans for randomization, blinding and so on. A pilot was launched last year that we plan to complete by the end of this year to assess the value of assigning at least one reviewer on each panel the specific task of evaluating the 'scientific premise' of the application: the key publications on which the application is based (which may or may not come from the applicant's own research efforts). This question will be particularly important when a potentially costly human clinical trial is proposed, based on animal-model results. If the antecedent work is questionable and the trial is particularly important, key preclinical studies may first need to be validated independently.

Informed by feedback from these pilots, the NIH leadership will decide by the fourth quarter of this year which approaches to adopt agency-wide, which should remain specific to institutes and centres, and which to abandon.

The NIH is also exploring ways to provide greater transparency of the data that are the basis of published manuscripts. As part of our Big Data initiative, the NIH has requested applications to develop a Data Discovery Index (DDI) to allow investigators to locate and access unpublished, primary data (see go.nature.com/rjjfoj). Should an investigator use these data in new work, the owner of the data set could be cited, thereby creating a new metric of scientific contribution unrelated to journal publication, such as downloads of the primary data set. If sufficiently meritorious applications to develop the DDI are received, a funding award of up to three years in duration will be made by September 2014. Finally, in mid-December, the NIH launched an online forum called PubMed Commons (see go.nature.com/8m4pfp) for open discourse about published articles. Authors can join and rate or contribute comments, and the system is being evaluated and refined in the coming months. More than 2,000 authors have joined to date, contributing more than 700 comments.

COMMUNITY RESPONSIBILITY

Clearly, reproducibility is not a problem that the NIH can tackle alone. Consequently, we are reaching out broadly to the research community, scientific publishers, universities, industry, professional organizations, patient-advocacy groups and other stakeholders to take the steps necessary to reset the self-corrective process of scientific inquiry. Journals should be encouraged to devote more space to research conducted in an exemplary manner that reports negative findings, and should make room for papers that correct earlier work.

We are pleased to see that some of the leading journals have begun to change their review practices. For example, Nature Publishing Group, the publishers of this journal, announced[8] in May 2013 the following: restrictions on the length of methods sections have been abolished to ensure the reporting of key methodological details; authors use a checklist to facilitate the verification by editors and reviewers that critical experimental design features have been incorporated into the report, and editors scrutinize the statistical treatment of the studies reported more thoroughly with the help of statisticians. Furthermore, authors are encouraged to provide more raw data to accompany their papers online.

Similar requirements have been implemented by the journals of the American Association for the Advancement of Science — *Science Translational Medicine* in 2013 and *Science* earlier this month[9] — on the basis of, in part, the efforts of the NIH's National Institute of Neurological Disorders and Stroke to increase the transparency of how work is conducted[10].

Perhaps the most vexed issue is the academic incentive system. It currently overemphasizes publishing in high-profile journals. No doubt worsened by current budgetary woes, this encourages rapid submission of research findings to the detriment of careful replication. To address this, the NIH is contemplating modifying the format of its 'biographical sketch' form, which grant applicants are required to complete, to emphasize the significance of advances resulting from work in which the applicant participated, and to delineate the part played by the applicant. Other organizations such as the Howard Hughes Medical Institute have used this format and found it more revealing of actual contributions to science than the traditional list of unannotated publications. The NIH is also considering providing greater stability for investigators at certain, discrete career stages, utilizing grant mechanisms that allow more flexibility and a longer period than the current average of approximately four years of support per project.

"Efforts by the NIH alone will not be sufficient to effect real change in this unhealthy environment."

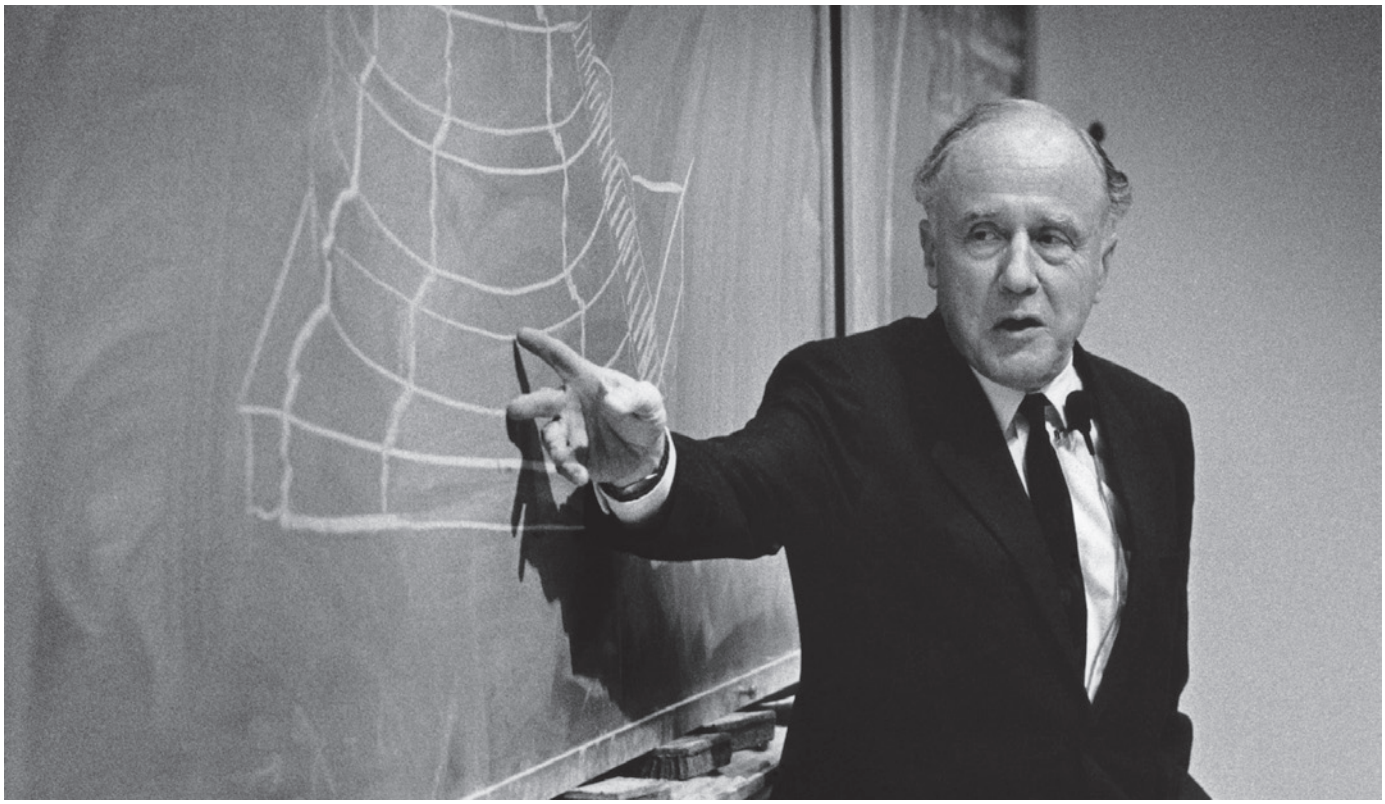
In addition, the NIH is examining ways to anonymize the peer-review process to reduce the effect of unconscious bias (see go.nature.com/g5xr3c). Currently, the identifiers and accomplishments of all research participants are known to the reviewers. The committee will report its recommendations within 18 months.

Efforts by the NIH alone will not be sufficient to effect real change in this unhealthy environment. University promotion and tenure committees must resist the temptation to use arbitrary surrogates, such as the number of publications in journals with high impact factors, when evaluating an investigator's scientific contributions and future potential.

The recent evidence showing the irreproducibility of significant numbers of biomedical-research publications demands immediate and substantive action. The NIH is firmly committed to making systematic changes that should reduce the frequency and severity of this problem — but success will come only with the full engagement of the entire biomedical-research enterprise. ■

Francis S. Collins *is director and* **Lawrence A. Tabak** *is principal deputy director of the US National Institutes of Health, Bethesda, Maryland, USA.*
e-mail: lawrence.tabak@nih.gov

1. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Disc.* **10,** 712–713 (2011).
2. *The Economist* 'Trouble at the Lab' (19 October 2013); available at http://go.nature.com/dstij3.
3. US Department of Health and Human Services, 2011 *Office of Research Integrity Annual Report 2011* (US HHS, 2011); available at http://go.nature.com/t7ykcv.
4. Carp, J. *NeuroImage* **63,** 289–300 (2012).
5. Vasilevsky, N. A. *et al. PeerJ* **1,** e148 (2013).
6. Franzoni, C., Scellato, G. & Stephan, P. *Science* **333,** 702–703 (2011).
7. Moher, D., Jones, A. & Lepage, L. for the CONSORT Group *J. Am. Med. Assoc.* **285,** 1992–1995 (2001).
8. *Nature* **496,** 398 (2013).
9. McNutt, M. *Science* **343,** 229 (2014).
10. Landis, S. C. *et al. Nature* **490,** 187–191 (2012).



EMILIO SEGRE VISUAL ARCHIVES/AIP/SPL

US theoretical physicist John Wheeler helped to bring general relativity into the mainstream.

PHYSICS

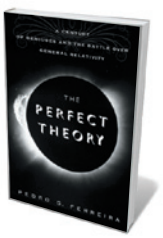
Einstein's curve ball

Graham Farmelo enjoys a 'biography' of the general theory of relativity.

The mathematical physicist Max Born remarked in 1955 that although his late friend Albert Einstein's general theory of relativity was a peerless scientific achievement, "its connections with experience [are] slender". The appeal of the theory for Born was similar to that of "a great work of art, to be enjoyed and admired at a distance".

Today, Born's comments seem quaint. In an age of precision astronomy, it is now possible to study consequences of the theory; the existence of gravitational waves, for instance, can be inferred from studying pulsars. With the theory's centenary only a year away, this is an opportune time to look back on its inception and its achievements, as astrophysicist Pedro Ferreira does in *The Perfect Theory*, a 'biography' of Einstein's brainchild for those with a smattering of science and next to no mathematics.

Einstein recalled that his crucial epiphany occurred in 1907. Sitting in the Swiss patent office in Bern, he realized that "if a person falls freely he will not feel his own weight". Using what Born described as "the most amazing combination of philosophical penetration, physical intuition and mathematical skill", Einstein developed his general theory of relativity — a new theory of gravity — and published it eight years later. In the final straight, the German mathematician David Hilbert was hot on his heels.



The Perfect Theory: A Century of Geniuses and the Battle over General Relativity
PEDRO G. FERREIRA
Houghton Mifflin Harcourt: 2014.

Ferreira outlines the theory, but I wish that he had tried a little harder to convey the surpassing beauty of Einstein's equations. As the great theoretician Steven Weinberg has stressed, this was the quality that persuaded his colleagues to take relativity seriously. I suspect that many readers would have tolerated a few moments' perplexity for a sense of its mathematical glory.

The theory was past its fiftieth birthday when it entered mainstream physics. As Ferreira describes, one of the most eminent of the physicists who brought the general theory back into the limelight was US theoretician John Wheeler. Wheeler was at first deeply uneasy about the theory's mathematical singularities — the point at which the quantities used to measure the strength of gravitational fields become infinite — and even wanted to remove them. In December 1963, he was one of the speakers at the first Texas Symposium on Relativistic Astrophysics, where the audience excitedly discussed the recently identified "quasi stellar radio sources", neatly dubbed quasars by one of the attendees. It seemed likely that general relativity might well be needed to understand this and other astronomical discoveries. Sure enough, the theory became a much more popular subject of study soon after, and several extremely strong research groups — notably in Moscow; Princeton, New Jersey; and Cambridge, UK — began to catch the eye of the physics community.

At the symposium was the British mathematical physicist Roger Penrose, who went on to work with Stephen Hawking to make

pioneering contributions to our understanding of the origins of the Universe, and of black holes, a term adopted by Wheeler after an audience member suggested it. Later, astronomers observed these exotic objects — regions of space-time where gravity is so strong that it appears nothing can escape. Einstein would surely have been delighted to see this and other demonstrations of the surprising consequences of his theory. His equations were smarter than he was, to paraphrase the physicist Paul Dirac.

Yet general relativity is not quite perfect. It takes no account of quantum theory and is extremely difficult to combine with the well-tested account of nature's other fundamental interactions — weak, electromagnetic and strong — in the standard model of particle physics. Ferreira lucidly sketches several attempts to generalize Einstein's theory, including string theory, which both describes gravity and offers an explanation of why it exists. Although enormously promising and mathematically rich, string theory is unpopular among some physicists in part because of the extreme difficulty of putting it to test, at least in the foreseeable future. Meanwhile, good old general relativity — once regarded as too recondite to be worth studying — is now the framework for planning and interpreting many astronomical experiments, as Ferreira describes in a moving coda.

"General relativity is now the framework for planning and interpreting many astronomical experiments."

When the sculptor Henry Moore visited Chicago, Illinois, in the late 1960s, the brilliant theoretical astrophysicist Subrahmanyan Chandrasekhar asked him how best to view a work of sculpture. Moore replied that the greatest of these works should be viewed from all distances, as new aspects of their beauty are revealed on every scale. Likewise, 50 years later, the mathematical aesthetic of relativity has been enhanced by the beautiful demonstrations of its veracity that Ferreira describes. These would probably have made Born ponder why he and his peers did not spend more time developing a deeper appreciation of the theory soon after Einstein first presented it. Maybe there's a lesson here for some of today's string-theory sceptics? ■

Graham Farmelo *is a by-fellow at Churchill College, Cambridge, and author of* Churchill's Bomb.
e-mail: graham@grahamfarmelo.com

SCIEPRO/SPL/CORBIS

NEUROSCIENCE

Joined-up thinking

Chris Frith explores a masterful model of how consciousness plays out in the theatre of the brain.

In 1874, Thomas Henry Huxley gave a prescient lecture on mind and brain. The biologist argued that subjective experience depends on the brain's "anterior divisions", and that consciousness has as little effect on behaviour as a steam whistle has on a locomotive's progress — rendering humans little more than "conscious automata". He raised two questions that remain key in contemporary studies of the neural basis of consciousness: what is special about the neural processes that underlie consciousness, and what, if anything, is consciousness for?

The 1870s seemed a likely time for a concerted research effort to answer those questions. Herman von Helmholtz had made the distinction between conscious and unconscious brain processes, and Gustav Theodor Fechner's 'psychophysics' had begun to allow the experimental study of the relationship between subjective experience and physical stimulation. But it was not until the 1970s that three-dimensional imaging of the living human brain became possible through physicist Peter Mansfield's work in magnetic resonance imaging. Among the first to realize the importance of this breakthrough for the study of mind and brain was cognitive neuroscientist Stanislas Dehaene. In his brilliant *Consciousness and the Brain*, Dehaene conveys the excitement of developing paradigms that such technologies have made possible.

Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts
STANISLAS DEHAENE
Viking Books: 2014.

For Dehaene, consciousness is simply this: we are conscious of whatever we choose to focus our attention on. He details many experiments, and presents the best attempt yet to answer the two questions raised by Huxley.

Regarding the first, on neural processes, the brain does a lot of work before we become conscious of a stimulus, as Helmholtz pointed out. When you read these words, you are rarely aware of the individual letters — yet you must have analysed them to have understood the meaning. How much unconscious analysis happens before what we are looking at emerges into consciousness? Dehaene relates how clever techniques have been developed to answer this question.

In backward masking, for example, a word (such as 'five') is presented, followed by a mask (a meaningless series of letters, for example). It has been found that the brain begins to analyse the word as soon as it appears, but that this analysis ceases when the mask appears. If the switch from word to mask is very rapid, there is no consciousness that the word was presented. Yet, as Dehaene has shown, the unconscious neural processing that goes ▶

on before the mask appears is enough for meaning to be extracted.

Combined with brain imaging, such studies show that activity in the region concerned with word recognition is not sufficient for consciousness. Instead, Dehaene reveals, conscious experience depends on interactions between sensory regions and the parietal and frontal areas of the brain. This is one of four neural signatures of consciousness that he lists.

These findings could be key in diagnosing locked-in syndrome, a state resembling coma in which a person is fully conscious, but unable to demonstrate it. Using brain-imaging techniques, it should soon be possible to detect consciousness in suspected cases: if a person with the syndrome imagines making a movement, for example, changes in brain activity linked to that could be detected.

Dehaene's special contribution is his global-workspace theory, the first step in a complete account of why some neural processes lead to conscious experience. The brain contains a number of discrete modules specialized for specific tasks, such as visual perception and motor output. Dehaene shows that for advanced cognitive processes — such as seeing things from the viewpoint of others — information generated by these modules must be maintained, manipulated and understood by several or all of them. The 'global workspace' is the virtual arena, created by long-range, synchronized neural connections, in which this happens. Only information that can be shared between modules enters consciousness. Effectively, without such conscious access, higher cognitive abilities would not be possible: consciousness is, Dehaene argues, no steam whistle.

I am not completely convinced that a global workspace is sufficient for consciousness. I believe that the ability to tell people about our experiences, as when tasting wine for example, is a crucial feature. However, our reportage is often erroneous, and that does not seem compatible with the precision needed for the information broadcast in Dehaene's global workspace. Nevertheless, Dehaene's account is the most sophisticated story about the neural basis of consciousness so far. It is essential reading for those who want to experience the excitement of the search for the mind in the brain. ■

Chris Frith *is emeritus professor of neuropsychology at the Wellcome Trust Centre for Neuroimaging at University College London and visiting professor at the Interacting Minds Centre at Aarhus University in Denmark. His books include* Making Up the Mind. *e-mail: c.frith@ucl.ac.uk*

MEDICINE

Bad medicine

Alison Abbott reviews an exhibition that reveals a lag in applying academic knowledge to medical practice.

DEUTSCHES MEDIZINHISTORISCHES MUSEUM INGOLSTADT

A nineteenth-century amulet used to guard against tooth ache and other ills.

Johannes Magirus enjoyed special status in Zerbst, southwest of Berlin, in the mid-seventeenth century. As the town's only academically trained physician, he treated rich and poor alike — and loved to impress the social elite with the breadth of his learning, from physics to astrology.

Magirus is one of eight physicians practising between the seventeenth and nineteenth centuries whose working lives are featured in *Praxiswelten* ('Practice worlds'), an unusual exhibition at the Berlin Medical History Museum. Anatomical knowledge increased dramatically over this time, and understanding of physiology and infection biology began their catch up in the late nineteenth century. As medicine became more scientific, barber-surgeons gradually gave way to university-trained physicians. But as this exhibition shows, the transition to scientific medicine was slow, perhaps because patients clung to the magical beliefs of other healers.

Praxiswelten showcases ongoing research by a consortium of medical historians who scoured libraries and the countryside for unusual source material: the original notebooks of doctors in German-speaking regions of Europe. It comes as a jolt to see that the notebooks are written in Latin. Also surprising is the enormous detail with which physicians recorded symptoms and the circumstances of patient visits. The notebooks reveal the very individual personal styles of the doctors, who, although exposed to modern knowledge at university, rarely applied it in daily practice. They tended to refer instead to imbalances of the four 'humours' of antiquity — black bile, yellow bile, blood and phlegm — or more recent theories not based on science.

For example, Friedrich von Bönninghausen, who opened his practice in 1864 in Münster, relied exclusively on homeopathy — despite having trained in Bonn and Berlin, the most prestigious German-speaking centres of medicine at the time. His notebook shows that he treated 11,500 people up to 1889, but he lost patients in droves thereafter. The germ theory of infectious diseases had emerged in Europe by then, thanks to the work of Louis Pasteur and Robert Koch, and public-hygiene measures such as using clean sources of water had proven so effective that scientific medicine gained in popularity.

In remote regions, neither physicians nor patients had it easy. The ill often had to send urine samples and descriptions of their symptoms using messengers, who needed to be fit. Franz von Ottenthal opened his practice in 1847 in the Alpine Ahrn Valley. His notebook records that he prescribed extract of meadow saffron as a painkiller for one Josef Brugger. But the treatment caused burning sensations in the stomach, as Brugger's messenger informed von Ottenthal. Von Ottenthal sent her back with the advice that Brugger supplement his treatment with sodium bicarbonate and powdered rhubarb. Whether that helped remains unrecorded, but the messenger had to trek a total of 26 rugged kilometres.

Back in 1653, Magirus claimed success in treating a toddler suffering from fever cramps with a range of strange medicines and ointments. The child's father was rich enough to pay for as much as Magirus's renowned knowledge could deliver. The physician consulted specialist literature, and used his mathematical skills to calculate the positions of stars and planets, applying his remedies when the celestial bodies were most propitiously aligned. The exhibition makes one wonder anew that 'alternative therapies' remain so popular today. ■

Praxiswelten
Berlin Medical Historical Museum
Until 21 September 2014.

Alison Abbott *is* Nature*'s senior European correspondent.*

BILL TRUSLOW

An etched glasswork by artist Peter Houk, part of his Big Dig series.

MATERIALS

Vitreous visions

Daniel Cressey celebrates the pending refit of the Glass Lab — an innovative crossroads of science and art at MIT.

The most oversubscribed programme at the Massachusetts Institute of Technology in Cambridge takes just 16 students a term and offers no credit nor classes in physics, chemistry or engineering. Instead, it teaches the art and science of glass-blowing — the creation of objects ranging from ornamental pumpkins to functional musical instruments, such as the flask-shaped 'vitreous membranophone'. Blowing through its neck creates audible oscillations in the thin glass base.

The Glass Lab is now attracting big-name glass artists. Dale Chihuly is one — famous for his exuberant, brilliantly hued sculptures resembling fantastical marine organisms or jungle flowers. The Venetian artist Lino Tagliapietra is also on board: his visually stunning vessels and geometric panels have earned him the title of the world's greatest glass-blower. Staff include the likes of mathematician–artist Martin Demaine, known as the father of Canadian glass.

NATURE.COM
For more lab art by Martin and Erik Demaine, see:
go.nature.com/3hprzz

The lab's sales of artworks created in its basement room, such as sculpted bowls and ornaments, are hugely popular.

This year, after almost three decades of operation, a lab refit costing around US$2.5 million will enable even more students to train in this modern alchemy — transforming sand into a frozen spray of colour. The number of workstation benches is to double. "The main reason driving the expansion and renovation is demand," says Peter Houk, Glass Lab director since 1997 and an established artist in the medium, whose work ranges from vases etched with exquisitely detailed cityscapes to huge coloured panels. "The joke is it's harder to get into the Glass Lab than MIT," he says. The odds are actually about the same, he explains, "but one is merit and the other is luck".

Engineering researcher Michael Cima, the lab's faculty director, was in on the idea from its unplanned beginnings. A junior faculty member in 1986, he was offered the lab to pursue his work, which involves engineering technologies in health and medicine. Two students and the artist Page Hazlegrove visited him before he had even glimpsed the lab, reporting that it contained a glass furnace "and would I mind if they used it". Cima had previous experience in lampworking — a process often used to make laboratory glassware — but none in the wilder realm of glass-blowing.

Today, training at the lab both channels artistic creativity and feeds directly into science, while providing valuable lessons on improvisation and other skills to future engineers and researchers. "The reason why the engineering school supports it is this learning how to improvise," says Cima. Glasswork is largely collaborative — the efforts of a team enduring scorching heat and the shards of failed attempts. "Glass-blowing teams have to adapt quickly while they work, changing their plans or methods in response to changes in the material they are manipulating," he adds.

The history of glass-blowing can be traced back to the fourth millennium BC, when it was realized that silicon dioxide, sodium oxide and calcium oxide, subjected to extremely high temperatures, would fuse into glass. Glass in the MIT lab, however, is ordered in as clear chunks called cullet. These are dropped into a furnace that can keep about 50 kilograms of glass molten at temperatures of 1,100–1,200 °C. Purified in one chamber of the furnace — which runs all day, every day — the refined glass flows into a second chamber, where it is retrieved by the glass-blowing team while it is still malleable.

The shape of every piece is determined by both the glass-blower and the person manipulating the glass on the bench. As its name implies, glass-blowing involves puffing into the blowpipe to which the glass blob is affixed, forming a bubble. Shaping is done with moulds or scissor-like tools called jacks, or simply by squeezing the mass by hand while protected by a wad of wet newspaper. Bubbles are also rolled on a steel table called a marver to shape them and to remove heat from certain parts, changing the way the bubbles grow when they are blown. The blobs can be repeatedly reheated to restore malleability.

Colour can be added with coloured glass rods; Cima likens them to a paint palette. These can be ground up and used to coat the bubbles of glass in a colour-saturated layer, or heated and pulled into strings to add lines and patterns.

Cima never ended up using the Glass Lab for his own research. But the adaptiveness and can-do inventiveness fostered by the lab, he says, is "a perfect example of why MIT is different". ■

Daniel Cressey *is a reporter for* Nature *in London.*

Correspondence

Gender: resolve bias, don't excuse it

It is difficult to make the claim that the disproportionate number of male reviewers and authors is not indicative of some level of gender bias (L. Koube *Nature* **505,** 291; 2014). As with many other challenges that female scientists face, the answer lies not in explaining why discrepancies exist, but in taking steps to resolve them.

The proportion of female referees (13% for *Nature* in 2013; *Nature* **504,** 188; 2013) remains considerably lower than the proportion of female researchers (roughly 30% in the United States, according to a 2013 report by the US National Science Foundation on Women, Minorities, and Persons with Disabilities in Science and Engineering). Not challenging this situation is tantamount to declaring that the quality of the pool of female referees is lower than that of their male counterparts, which is both short-sighted and wrong.

Arguments about personal or family responsibilities only serve to cloud the bigger issue, which is about finding a way to work towards a body of scientific literature that represents true gender balance among those contributing to it.

Morgan V. Fedorchak
University of Pittsburgh, Pennsylvania, USA.
mod8@pitt.edu

Gender: why publish an offensive letter?

I want an answer to this question. If the answer was to engender controversy, then it worked; but if it was to reinforce *Nature*'s "own positive views and engagement in the issues concerning women in science" (*Nature* **505,** 483; 2014), then it failed. Here is the context: two weeks ago, *Nature* published a Correspondence from Lukas Koube (*Nature* **505,** 291; 2014), which in my view implies that journals' pursuit of scientific quality will logically and inevitably result in women's invisibility. On the day that I read it, I was scheduled to do an interview about my research for the Careers section of *Nature*. I declined the interview.

Declining this interview was a strategic decision. Every young scientist is told that publication in *Nature* is a valuable prize, a harbinger of 'glory, laud and honour' and of job security. Thus, the assignment of a *Nature* DOI (digital object identifier) is a powerful force of reification, one that endures far beyond any squabbling that may precede or follow it.

Nature states that the correspondence it publishes does not necessarily reflect the opinions of the journal or its editors (*Nature* **505,** 483; 2014). However, people have a deep-seated tendency to associate the *Nature* brand with a stringent selection process for publication. Out of the many letters it receives, why did *Nature* want its readers to read Koube's? It is unclear why you should publish his Correspondence at all in an age when people's comments already have multiple outlets for mass distribution. My interview cancellation was meant to provide concrete evidence that at least one reader wants an answer.

Nature is a powerful institution in which its editors, reviewers, authors and readers invest a monumental amount of effort and care. For this very reason, it is also an institution at which each editorial choice merits exceptional scrutiny.

A. Hope Jahren *University of Hawaii, USA.*
jahren@hawaii.edu

Plume hypothesis challenged

The hundreds of Earth scientists who challenge the existence of plumes of hot rock rising from Earth's core–mantle boundary are not "a small but vocal subset" (*Nature* **504,** 206–207; 2013). Rather than simply promoting the conventional wisdom, you should be encouraging the development of multiple working hypotheses.

Many scientists have valid concerns that the originally postulated behavioural, geometric, chemical and thermal characteristics of mantle plumes have been widely discredited (W. J. Morgan and J. P. Morgan in *Plates, Plumes, and Planetary Processes* 65–78; Geological Society of America, 2007). The plume model has survived only by diversifying its supposed characteristics, which include a variety of compositions and feats such as tunnelling thousands of kilometres horizontally to emerge anywhere at any time, splitting, merging and pulsing (E. R. Lundin in *52 Things You Should Know About Geology* 66–67; Agile Libre, 2013).

There are no chemical or isotopic data that require deep-plume origins or anomalously high temperatures, and no reliable seismic-tomography results have ever revealed a plume. Plumes cannot account for the eruption rates of the largest flood basalts, which can best be explained by rapidly draining reservoirs of molten rock that have accumulated over long periods.

There has been significant progress in developing an alternative model for anomalous volcanism (see, for example, G. R. Foulger *Plates vs Plumes: A Geological Controversy*, Wiley-Blackwell; 2010). This is better explained as a passive response to the stretching of lithospheric plates — for example, at rift valleys — which permits melt to rise from shallow depths in the mantle.

Gillian R. Foulger *Durham University, UK.*
g.r.foulger@durham.ac.uk
Warren B. Hamilton *Colorado School of Mines, USA.*

Cut costs with open-source hardware

Sally Tinkle and others (see *Nature* **503,** 463–464; 2013) highlight the importance of open-source software and data sharing in materials science. But researchers should also be developing free and open-source hardware to radically reduce the costs of their experimental work.

Harnessing open-source methodology will ensure that funding used to develop scientific equipment is spent only once. A return on investment is achieved through digital replication of devices for just the cost of the materials required. This scaled replication saves 90–99% on conventional costs, making more scientific equipment available for research and education (see J. M. Pearce *Open-Source Lab*, Elsevier; 2013).

Dozens of free open-source designs for lab equipment already exist. For example, the University of Washington in Seattle has produced a magnetic rack for molecular and cell-separation applications that can be fabricated with a three-dimensional printer for less than it can be bought commercially. Even if the device is made only once, it justifies the price of the printer. A hand-held open-source colorimeter built in my department for US$50 matches the performance of similar tools that cost more than $2,000. And the University of Cambridge, UK, has developed a microscope for about $800 from open-source plans, to use instead of conventional equivalents costing up to 100 times as much.

Federal funding agencies could join forces to fund open-source scientific hardware to accelerate its development. A free online database of tested and validated tools should be set up, and governments should give preference to funding such hardware purchases.

Joshua M. Pearce *Michigan Technological University, USA.*
pearce@mtu.edu

FORUM: Crystallography

Sources of inspiration

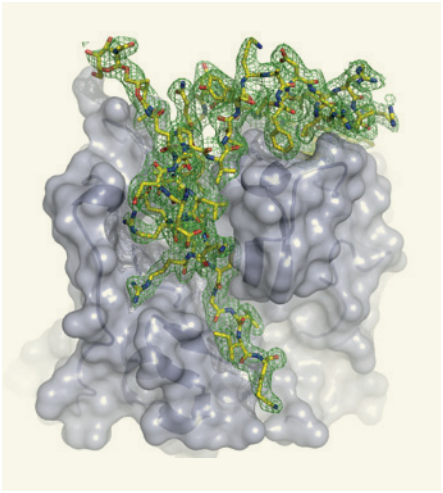
Synchrotrons have long been the preferred X-ray sources for crystallography, but competition has arrived with the advent of X-ray free-electron lasers. A synchrotron expert and an advocate of free-electron lasers discuss the prospects of the respective source types for applications in structural biology.

Sophisticated synchrotrons

SEAN MCSWEENEY

During the 20 years that biological structures have been solved using modern synchrotron sources, the hundreds of thousands of experiments performed have revolutionized the process of determining macromolecular structures. These high-intensity, well-collimated X-ray beams continually drive biologists to try new approaches, pushing our capabilities to reveal ever-larger molecular complexes at atomic resolution. The usefulness of these X-ray beams has also driven a steady rise in the number of crystallographic instruments at synchrotron facilities. Structural biology has thus increasingly been used as a major tool for generating fundamental biological knowledge — much of which has benefited society by aiding the discovery of new drugs.

When third-generation synchrotrons (also known as undulator-based storage rings) began operating in 1994, only two or three crystal structures were being deposited each week in the Protein Data Bank, an international repository for protein structures. Since then, the number of beamlines — the specialist instrumentation that enables light from synchrotrons to be used in experiments — has risen considerably, mostly at third-generation sources. The number of structural biologists has increased in parallel with the ease with which X-ray diffraction revealed natural structures. At present, around 8,000 distinct structures are deposited each year, approximately one per hour. This demonstrates how continual innovation by synchrotron-facility scientists and users has made the existing sources incredibly productive. The list of their achievements is encyclopaedic, and includes the development of: automated sample handling, advanced detectors, improved software, new crystallographic methods and stable X-ray optics that produce microscopic X-ray beams.



KAROL NASS

Figure 1 | Protein structures from micrometre-sized crystals. The cathepsin B protein of the parasitic microbe *Trypanosoma brucei* is a potential target for drugs to combat sleeping sickness. The crystal structure of the protein (grey) in complex with an inactivating peptide (multicoloured) was first determined[7] from micrometre-sized crystals grown *in vivo*, using X-rays from a free-electron laser (FEL). The structure has since been validated[4] on a synchrotron using a method inspired by techniques developed for FELs.

These micro-focused X-ray beams may prove to be crucial to structural biologists in the future[1,2]. Five Nobel prizes have been awarded for work that depended on synchrotron X-ray studies. The most recent of these — the Nobel Prize in Chemistry 2012, which was awarded, in part, for the determination of the structures of G-protein-coupled receptors[3] — required the sort of micro-beam that became available only recently. Such beams allowed the delivery of a high flux of X-rays in the tiny volume that was needed to collect crystallographic data from the fragile protein crystals involved.

Further beamline developments will continue until it is possible to truly tune experiments, controlling beam size, shape, flux and wavelength, thereby enabling optimal extraction of information from crystal samples. Storage-ring developments will also continue: the fourth generation of synchrotrons is currently under construction, and will eventually produce flux densities a thousand to a million times higher than those of current state-of-the-art instruments, allowing new experimental approaches and scientific discoveries.

Impressive results from free-electron lasers (FELs) have made some people wonder whether conventional storage-ring sources will continue to have a major role in driving structural biology. I contend that both tools are developing synergistically, and that we are still far from being able to realize the full potential of storage-ring sources in particular. In the next decade, scientists will benefit from synchrotrons even more than they do now, as a result of innovations that are spurred, in part, by FELs. For example, a recent study[4] reports how intense, micro-focused X-ray beams from a synchrotron, combined with data-analysis techniques previously developed for FEL experiments, have enabled structures to be determined from micrometre-scale crystals (Fig. 1). It is fair to say that the future is bright for synchrotrons in structural biology.

Sean McSweeney *is in the Department of Photon Sciences, Brookhaven National Laboratory, Upton, New York 11973-5000, USA.*
e-mail: smcsweeney@bnl.gov

Leading-edge lasers

PETRA FROMME

Free-electron lasers[5] have opened up a new era in structural biology[6], for several reasons. For starters, FELs allow structures to be determined from nanometre-scale crystals that contain only a few hundred molecules. These nanocrystals are easier to grow and have fewer defects than the macroscopic crystals used for conventional crystallography.

This is especially helpful for proteins that

are difficult to crystallize, such as large complexes and proteins embedded in membranes. Recently, a structure was determined with a FEL using nanocrystals prepared by overexpressing a protein in insect cells[7] (Fig. 1). This method of preparation seems to be applicable to many proteins, and could save years that would otherwise be spent crystallizing proteins using conventional methods.

FELs also overcome one of the main obstacles in crystallography: that proteins are often damaged by conventional X-ray sources. X-ray pulses from FELs are extremely intense and so completely destroy molecules and crystals. But because the pulses have only femtosecond duration (1 femtosecond is 10^{-15} seconds), diffraction patterns can be detected before the molecules are destroyed[8]. This overcomes the size limit for crystals, as noted earlier. It also allows damage-free structures to be determined from radiation-sensitive crystals. This is especially important for proteins that contain metal centres, which tend to undergo X-ray-induced chemical reduction.

Biomolecules are dynamic, but most crystal structures provide only a static picture of such molecules in one state. By contrast, time-resolved femtosecond crystallography using FELs allows researchers to make 'molecular movies' — a series of snapshots — of biomolecules in action. For proteins whose reactions can be triggered by light, X-ray pulses fired at different times after a light trigger enable the structures of different reaction intermediates to be obtained[9].

Not all protein reactions are light driven, however. Methods are therefore being developed in which rapid mixing of protein nanocrystals with a solution of the protein's substrate triggers a reaction; X-ray pulses are then fired at the sample at different time intervals after mixing. This should enable all the steps of drug transport through a receptor to be visualized, for example.

The current main limitation of structural biology research with FELs is access to beam time at the two sources in the United States and Japan. But, with the opening of the European FEL and the Swiss FEL in 2015 or 2016, available beam time will increase significantly. Furthermore, the European FEL will allow up to 10,000 images to be collected per second, so that a full data set can be acquired in 5 minutes, rather than the 3 hours required at present.

It is the dream of structural biologists to determine atomic structures from the X-ray diffraction of single molecules, but this is not yet within our grasp. To reach this goal major challenges have to be met: the flux of X-ray photons from FELs must be increased by at least 1,000-fold to detect the weak diffraction of individual biomolecules at atomic resolution. In addition, the duration of pulses may have to be shortened to less than a femtosecond, to allow for diffraction before destruction of single molecules. ■

Petra Fromme *is in the Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA.*
e-mail: pfromme@asu.edu

1. Riekel, C., Burghammer, M. & Schertler, G. *Curr. Opin. Struct. Biol.* **15,** 556–562 (2005).
2. Smith, J. L., Fischetti, R. F. & Yamamoto, M. *Curr. Opin. Struct. Biol.* **22,** 602–612 (2012).
3. www.nobelprize.org/nobel_prizes/chemistry/laureates/2012/
4. Gati, C. *et al. IUCrJ* http://dx.doi.org/10.1107/S2052252513033939 (2014).
5. Emma, P. *et al. Nature Photon.* **4,** 641–647 (2010).
6. Chapman, H. N. *et al. Nature* **470,** 73–77 (2011).
7. Redecke, L. *et al. Science* **339,** 227–230 (2013).
8. Barty, A. *et al. Nature Photon.* **6,** 35–40 (2012).
9. Aquila, A. *et al. Opt. Exp.* **20,** 2706–2716 (2012).

MOLECULAR BIOLOGY

A second layer of information in RNA

Three studies have characterized the full complement of RNA folding in cells. They find large numbers of secondary structures in RNA, some of which may have functional consequences for the cell. SEE LETTERS P.696, P.701 & P.706

SILVIA B. V. RAMOS & ALAIN LAEDERACH

The RNA molecule is generally understood as a messenger of genetic information in the cell: it is transcribed from DNA and then translated into proteins[1]. Stretches of RNA that are complementary in sequence have a propensity to pair, forming elements of secondary structure, such as hairpin loops, within RNA molecules. But the prevalence of secondary structure in messenger RNAs, and its role in RNA regulation, is not fully understood. In this issue, three reports[2–4] describe analyses of all the mRNA molecules present in different populations of cells — transcriptome-wide analyses — using structure-probing techniques. These studies begin to reveal the extent of secondary structure in the transcriptomes of plants, humans and yeast.

The chemical structure of RNA is analogous to that of DNA. It is comprised of a sugar–phosphate backbone and four distinct nucleotide bases: adenine (A), cytosine (C), guanine (G) and uracil (U). As with DNA, these bases interact by forming hydrogen bonds, resulting in aptly named Watson–Crick pairs (G–C and A–U). However, unlike DNA, complementary bases from two RNA molecules do not pair up to form a double helix, a formation that in DNA prevents secondary structures from arising. Instead, the nucleotides of RNA are free to interact with one another within each molecule, resulting in folding of the RNA chain into secondary structures (Fig. 1).

The functional consequences of secondary structural elements in RNA depend on their molecular context. Some specific structural elements have well-known regulatory roles after gene transcription, but these are restricted to small subsets of mRNAs[5,6]. In some cases, such as in ribosomal RNA (part of the cellular

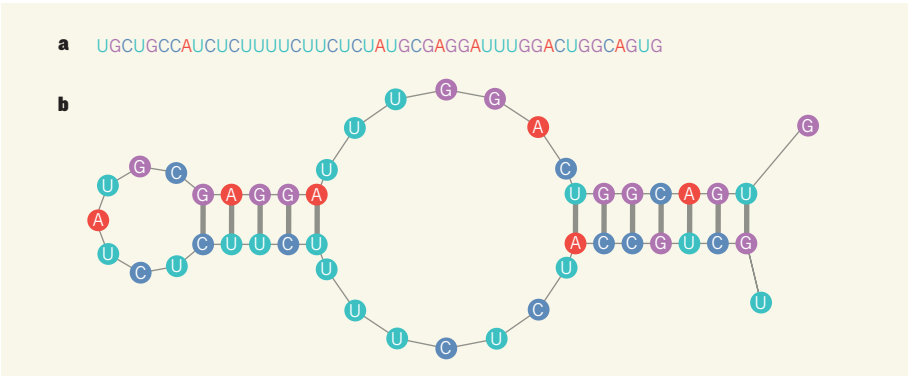


Figure 1 | Principles of RNA primary sequence and secondary structure. **a**, RNA is a single-stranded polymer, with nucleotide bases adenine (A), cytosine (C), guanine (G) and uracil (U). **b**, Unlike DNA, RNA molecules do not pair up to form helices. The bases of an individual molecule can therefore pair with one another (G–C and A–U), causing the RNA to fold into secondary structures. G bases can also pair with U bases, forming a G–U wobble pair. Three reports[2–4] find that such folding is commonplace in humans, plants and yeast. (Figure adapted from Fig. 3 of ref. 4.)

machinery that synthesizes proteins), secondary structural elements fold further into compact three-dimensional conformations that can catalyse reactions[7].

The three new studies, each analysing different cell populations, use a combination of well-established chemical and enzymatic structure-probing techniques for determining RNA secondary structure together with next-generation sequencing, a method that allows simultaneous sequencing of millions of stretches of nucleotides. Ding and colleagues[2] (page 696) examined seedlings from the plant *Arabidopsis thaliana*, Rouskin and co-workers[3] (page 701) investigated yeast, and both Rouskin *et al.* and Wan and colleagues[4] (page 706) report analyses of secondary structures in humans. All three papers report unprecedented coverage of the transcriptome[8]. In doing so, they demonstrate unequivocally that most mRNAs have a propensity to form secondary structures *in vitro*, in the absence of any other cellular components.

Each group reports that some of the RNA structures they observed *in vitro* were altered *in vivo*. In fact, Rouskin and colleagues found evidence in yeast that RNA structures in the cell are actively unfolded by proteins. Nonetheless, the papers show that structural patterns are evolutionarily conserved at several functional sites within RNA molecules. These results provide the first *in vivo* data to suggest that, if given the opportunity, RNA will fold. This is consistent with many previous *in vitro* studies[9] of RNA structure and folding. Because mRNA must be unfolded to successfully act as a messenger, the cell must therefore find ways to get around the folding problem.

In addition to their structural characterization of the human transcriptome, Wan and co-workers performed comparative structure probing in cell lines derived from a family trio (mother, father and child). In so doing, they were able to assess the structural consequences of natural human inter-generational genetic variation on the transcriptome, and discovered more than 1,900 single-nucleotide mutations that alter RNA structure. These experiments therefore yielded thousands of new putative 'ribosnitches'[7,10] — broadly defined as RNA sequences in which a specific single-nucleotide mutation alters structure[7]. Ribosnitches are analogous to bacterial riboswitches, which change structure on binding of a small molecule and regulate transcription or translation[11].

Because RNA structure has the potential to influence post-transcriptional processes in the cell, a subset of the putative ribosnitches could be functional. Indeed, mutations that disrupt certain RNA secondary structural elements can cause human disease[10]. Although the structural changes identified in Wan and colleagues' work are not by themselves indicators of malfunction — the three individuals studied are presumably healthy — the newly identified putative ribosnitches have the potential to help to identify mechanisms by which structural changes can give rise to disease, an exciting step forward.

The application of next-generation sequencing to the transcriptome has previously revealed the complexity of post-transcriptional regulatory networks[5]. The structural dimension of this complexity is now accessible with the publication of these three papers. Although the three studies reveal similar general structural features of transcripts, there are key differences in the specific features found by each approach. Such discrepancies may come from differences in experimental design, which can cause changes to the inherently dynamic structure of RNA. In this case, each study used different protocols for RNA extraction, library preparation and, in particular, determining levels of background noise. These experimental details must be taken into account when comparing structures discovered using the different approaches.

The trio of reports provides our first insight into the secondary structure of an entire transcriptome in eukaryotes — the class of organisms comprising plants, animals and fungi. However, a full characterization of transcriptome structure will require a concerted community effort, with an emphasis on standardization to allow quantitative comparisons of these data sets. Only then will it be possible to fully integrate these findings to determine the structural elements that are consequential in the transcriptome[12]. ■

Silvia B. V. Ramos *and* **Alain Laederach** *are in the Obstetrics and Gynecology Department and the Biology Department, University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA.*
e-mails: alain@unc.edu; silvia_ramos@med.unc.edu

1. Crick, F. *Nature* **227,** 561–563 (1970).
2. Ding, Y. *et al. Nature* **505,** 696–700 (2014).
3. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissmann, J. S. *Nature* **505,** 701–705 (2014).
4. Wan, Y. *et al. Nature* **505,** 706–709 (2014).
5. Woodson, S. A. *Curr. Opin. Chem. Biol.* **12,** 667–673 (2008).
6. Dominski, Z. & Marzluff, W. F. *Gene* **396,** 373–390 (2007).
7. Halvorsen, M., Martin, J. S., Broadaway, S. & Laederach, A. *PLoS Genet.* **6,** e1001074 (2010).
8. Tijerina, P., Mohr, S. & Russell, R. *Nature Protocols* **2,** 2608–2623 (2007).
9. Shcherbakova, I., Mitra, S., Laederach, A. & Brenowitz, M. *Curr. Opin. Chem. Biol.* **12,** 655–666 (2008).
10. Martin, J. S. *et al. RNA* **18,** 77–87 (2012).
11. Tucker, B. J. & Breaker, R. R. *Curr. Opin. Struct. Biol.* **15,** 342–348 (2005).
12. Ulitsky, I. & Bartel, D. P. *Cell* **154,** 26–46 (2013).

CELL BIOLOGY

Potency unchained

Differentiated cells have been reprogrammed to an embryonic-like state using a physical stimulus. This treatment generates a new cell population that contributes to both the embryo and the placenta. SEE ARTICLE P.641 & LETTER P.676

AUSTIN SMITH

Cell specialization in mammals is essential for diverse functions, such as muscle contraction and nerve conduction. These specializations become fixed during development, and conversion between differentiated cell types seems to be extremely rare. However, in this issue, two studies by Obokata *et al.*[1,2] show that cells isolated from newborn mice lose their identity on exposure to mildly acidic conditions. Remarkably, instead of triggering cell death or tumour growth, as might be expected, a new cell state emerges that exhibits an unprecedented potential for differentiation into every possible cell type.

Studies on tissue regeneration in amphibians, reptiles and birds indicate that differentiated cells have some ability to dedifferentiate or to switch identity. Mammalian cells are more resistant, but fate conversion is observed in certain cancers. It was only with the cloning of Dolly the sheep[3], in which nuclear material from the mammary cell of an adult sheep was transferred into an enucleated egg cell to produce a cloned animal, that the capacity for complete reprogramming of the mammalian genome was confirmed. However, cloning does not convert whole cells.

Whole cells can be induced to switch identity by genetic manipulation. The introduction of certain transcription factors can in specific contexts rewire gene circuitry, leading to changes in cell specialization[4]. In 2006, the cell-identity and cloning research paths were unified through the discovery of a phenomenon known as induced pluripotency[5]: when mouse fibroblast cells were treated with a quartet of embryonic regulatory factors, a small percentage adopted the molecular and functional attributes of embryonic stem cells. The resulting induced pluripotent stem cells (iPSCs) had the dual abilities to self-renew indefinitely and to differentiate into all somatic cell types (Fig. 1a). Subsequently, iPSCs have

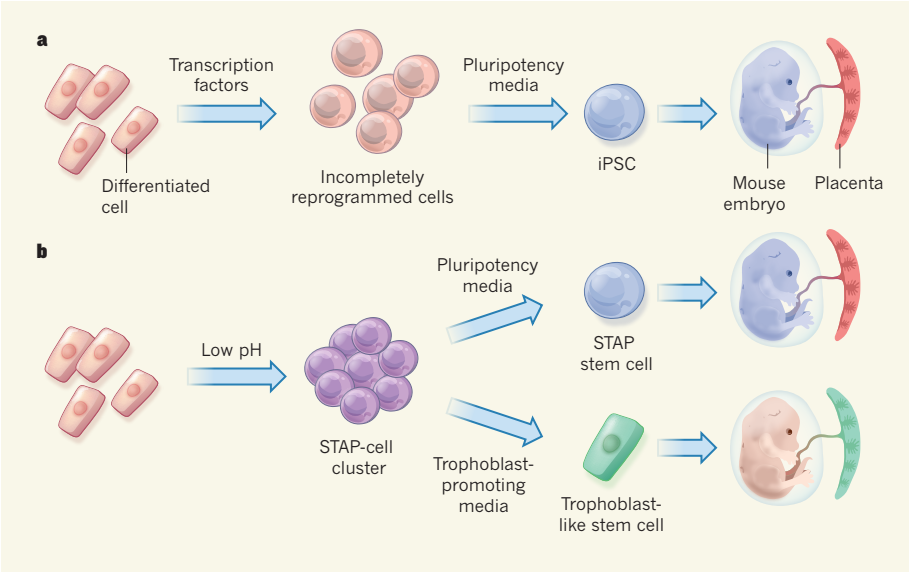


Figure 1 | Alternative methods for dedifferentiating specialized cells. **a**, Differentiated cells are typically reprogrammed to an embryonic-like (pluripotent) state using transcription factors and a cell-culture medium that promotes pluripotency[5]. This creates induced pluripotent stem cells (iPSCs), which can self-renew and contribute to all the cell types in a developing embryo, but not the placenta. iPSC generation occurs through a proliferative intermediate stage. **b**, Obokata and colleagues report[1,2] that dedifferentiation can also be achieved by short-term exposure of differentiated cells to a solution of low pH, a process they call stimulus-triggered acquisition of pluripotency (STAP). STAP cells do not proliferate, but subsequent treatment with pluripotency-promoting media produces STAP stem cells, which have the same properties as iPSCs. When cultured in a medium that promotes the growth of trophoblast stem cells (a placenta-generating cell type), STAP cells acquire trophoblast-like characteristics. Unlike iPSCs, the cells can contribute to the placenta.

been produced from a range of adult cell types, fostering enthusiasm worldwide for developing customized disease-modelling and cell-therapy applications.

What, then, is the significance of the two reports by Obokata and colleagues? The authors were inspired by the notion that physical stimuli might be sufficient to change a cell's identity. An example from nature is temperature-dependent sex determination in crocodile embryos[6]. And in the laboratory, frog cells fated to form skin will develop into brain tissue[7] if exposed to a solution of low pH. These changes in fate occur in embryonic progenitor cells. In their first report (page 641), Obokata and co-workers[1] investigated the effect of physical stimuli on cells from newborn mice.

In a similar manner to that used in iPSC reprogramming studies[5], the researchers monitored cells using a 'reporter' protein that fluoresces when a gene associated with pluripotency is turned on. They applied various stresses to white blood cells and found that, after a short exposure to a solution of low pH, the cells lost markers of blood identity, and a proportion activated the pluripotency reporter. The authors collected cells marked with the reporter and found that the cells had gene markers typical of early embryos. They describe this phenomenon as stimulus-triggered acquisition of pluripotency (STAP).

When injected into embryos, cells generated by STAP ('STAP cells') produced chimaeras — mice composed of cells originating from both the host embryo and the STAP cells. The ability to produce chimaeras is a property that was previously thought to be exclusive to embryonic stem cells and iPSCs. STAP cells differ from both of these cell types, however, in that they have little or no capacity for self-renewal and can be maintained for only a few days. The authors investigated this discrepancy and discovered that, if STAP cells are transferred into the culture conditions used to grow pluripotent stem cells, they begin to proliferate, and acquire structural features and gene markers diagnostic of embryonic stem cells. The researchers termed these self-renewing cells STAP stem cells (Fig. 1b).

The STAP-cell state might therefore be similar to the incompletely reprogrammed intermediate cells observed during iPSC formation[8]. However, as documented in Obokata and colleagues' second report[2] (page 676), a further surprise was in store. Examination of chimaeras produced from STAP cells revealed that the cells colonized extraembryonic layers such as the trophoblast, a structure that gives rise to the placenta, in addition to the embryo body. This colonization is rarely seen in chimaeras produced from embryonic stem cells or iPSCs, and it implies that cells generated by STAP have an unusually broad developmental potency. The authors then tested culture conditions normally used to obtain trophoblast stem cells. STAP cells again proliferated, but now acquired a trophoblast-like identity, confirming their broad potency (Fig. 1b).

The STAP state may represent a developmental stage that precedes segregation of extraembryonic and embryonic cell lineages. However, it is not evident that embryos ever contain single cells that have the complement of markers and behaviours exhibited by STAP cells. An alternative explanation could be that STAP cells constitute a mixture of cells predisposed to extraembryonic or embryonic differentiation. More provocatively, STAP cells might be an indeterminate, synthetic cell type — a blank slate — from which extraembryonic or embryonic gene circuitries emerge in appropriate environments.

The unexpected finding that a physical stimulus can trigger dedifferentiation of cells to a state of unrestricted potency opens up the possibility of obtaining patient-specific stem cells by a simple procedure, without genetic manipulation. STAP cells have yet to be produced from humans, however. Obokata and colleagues[1] provided evidence of reproducible STAP-cell generation from different mouse tissues, but they did not test other species. Furthermore, they used immature cells as a starting material, and it remains to be seen whether adult cells will respond similarly. Nonetheless, they have established a new principle: that a physical stimulus can be sufficient to dismember gene-control circuitry and create a 'plastic' state from which a previously unattainable level of potency can rapidly develop.

How pluripotent circuitry self-organizes, and how the body suppresses this, are fascinating questions. Notably, provision of leukaemia inhibitory factor (LIF), a cell-signalling molecule, promotes the emergence of STAP cells. This is tantalizing because LIF is the self-renewal factor for embryonic stem cells, has a crucial role in the formation of iPSCs[9,10] and drives the conversion of germ cells to pluripotent cells[11]. ■

Austin Smith *is at the Wellcome Trust–Medical Research Council Stem Cell Institute, and in the Department of Biochemistry, University of Cambridge, Cambridge CB2 1QR, UK.*
e-mail: austin.smith@cscr.cam.ac.uk

1. Obokata, H. *et al. Nature* **505,** 641–647 (2014).
2. Obokata, H. *et al. Nature* **505,** 676–680 (2014).
3. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385,** 810–813 (1997).
4. Graf, T. *Cell Stem Cell* **9,** 504–516 (2011).
5. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
6. Gilbert, S. F. *Developmental Biology* 6th edn (Sinauer, 2009).
7. Holtfreter, J. *J. Exp. Zool.* **106,** 197–222 (1947).
8. Silva, J. *et al. PLoS Biol.* **6,** e253 (2008).
9. Yang, J. *et al. Cell Stem Cell* **7,** 319–328 (2010).
10. Stuart, H. T. *et al. Curr. Biol.* http://dx.doi.org/10.1016/j.cub.2013.12.040 (2014).
11. Leitch, H. G. *et al. Stem Cell Rep.* **1,** 66–78 (2013).

ASTROPHYSICS

Portrait of a dynamic neighbour

Brown dwarfs are celestial objects that lack the mass to become fully fledged stars. High-resolution maps of one such object add to the evidence that these exotic worlds have highly dynamic weather and climate. SEE LETTER P.654

ADAM P. SHOWMAN

Humanity's study of stars and planets stretches back centuries, but our understanding of intermediate objects — brown dwarfs — is relatively primitive. Brown dwarfs are fluid, hydrogen-dominated objects that are generally presumed to form like stars, but that contain insufficient mass to fuse hydrogen into helium. As links between planets and stars, these dwarfs provide clues about the processes of star and planet formation, the physics of interior structure and the behaviour of atmospheres under exotic conditions. But because they are so far away, brown dwarfs are seen as only unresolved points of light in telescope images. So far, observations of these objects have all been measurements of the combined light from their Earth-facing hemisphere that preclude any detailed view of what they look like. That has now changed: on page 654 of this issue, Crossfield *et al.*[1] present the first spatially resolved maps of the visible surface of a nearby brown dwarf.

Crossfield and colleagues' maps of the brown dwarf, dubbed Luhman 16B, show large-scale bright and dark regions suggestive of patchy clouds. As such, the maps provide constraints on the dominant length scales of the meteorological motions and the overall nature of the atmospheric circulation on these exotic worlds. Luhman 16B was discovered in 2013 and lies a mere 2 parsecs away[2], making it and its companion brown dwarf (Luhman 16A) the third-closest stellar or sub-stellar system to Earth, after α-Centauri and Barnard's star. Still glowing from the heat of its formation billions of years ago, the brown dwarf's atmospheric temperatures reach a baking 1,200 kelvin.

Given that brown dwarfs are unresolved points of light in the sky, how did Crossfield *et al.* construct these maps? The answer lies in the technique of Doppler imaging[3] (Fig. 1). Brown dwarfs rotate rapidly — Luhman 16B rotates once every 4.9 hours. This fast rotation leads to movement of the atmospheric gas towards Earth on one side of the object and away from Earth on the other side. These rotational motions result in a change in frequency (Doppler shift) of emitted light, which, in turn, causes significant broadening of the emission lines observed in infrared spectra. If the visible surface of the brown dwarf were featureless, the spectral lines would be approximately symmetrical. But discrete, bright or dark atmospheric features that move across the dwarf's Earth-facing hemisphere during its rotation cause time-dependent asymmetries in the shape of the spectral lines that can be inverted to create a map, in longitude and latitude, of this surface patchiness. The technique has long been applied to stars[3], but Crossfield and colleagues' study is the first to apply it to brown dwarfs.

The observations add to a growing body of evidence demonstrating that brown dwarfs exhibit highly dynamic weather and climate. Atmospheric motions have long been hinted at from the presence of clouds and disequilibrium chemistry — the result of vertical mixing of atmospheric gas — that are inferred from infrared spectra of brown dwarfs. The first spectacular evidence for weather on brown dwarfs emerged[4] in 2009, when it became clear that the infrared emission of many brown dwarfs shows strong variability in integrated brightness on timescales of hours to days. Several lines of evidence indicate that this variability results from relatively cloudy and cloud-free patches coming into or out of view as the brown dwarf rotates. Luhman 16B is no exception, and recent observations[5,6] indicate that it exhibits peak-to-peak brightness variations of about 5–20%, fluctuating in time as the weather evolves. Although tantalizing, such variability provides only loose constraints on the size, shape and configuration of atmospheric features, rendering any direct assessment of atmospheric circulation for these objects difficult. In this context, Crossfield and colleagues' maps are potentially game changing.

Brown dwarfs generally, and Luhman 16B specifically, occupy a key position in our grand effort to understand the mechanisms and behaviour of atmospheric circulation over a wide range of conditions. As on planets such as Earth and Jupiter, the rapid rotation of brown dwarfs ensures that their atmospheric dynamics are rotationally dominated at large scales[7]. Unlike most known planets, however, brown dwarfs receive negligible external irradiation. Earth's global-scale weather is driven primarily by the contrast in solar heating between the Equator and the poles, a type of climate forcing that is ruled out for brown dwarfs such as Luhman 16B. Theories suggest that the vigorous convection that takes place in a brown dwarf's interior, which is necessary to

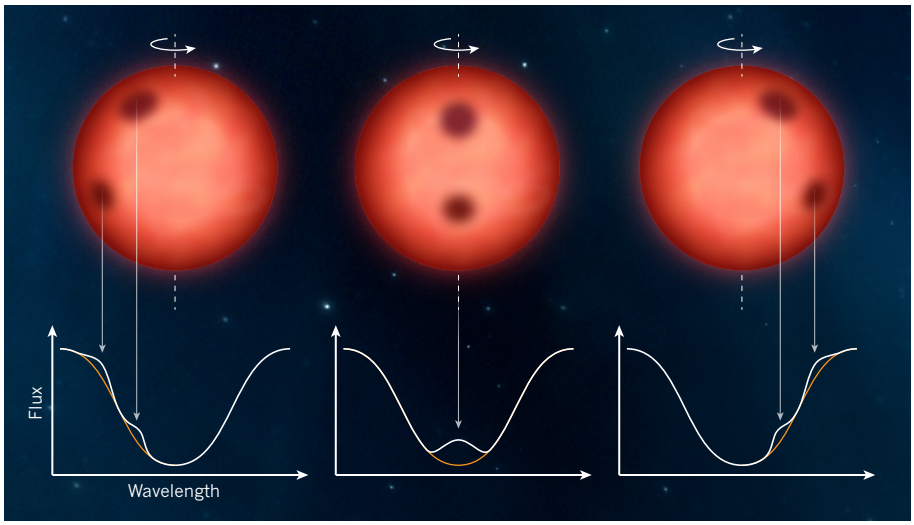


Figure 1 | The Doppler imaging technique. The rotation of a rapidly spinning star or brown dwarf causes a significant broadening of spectral lines through the Doppler shift in the frequency of emitted light. For a featureless brown dwarf, the broadened line is mirror symmetrical in wavelength with respect to the line centre (orange curves). However, the presence of discrete spots causes perturbations in the line shape (white curves). Such perturbations will move from the left to the right wing of the line as the spots move across the Earth-facing hemisphere owing to the rotation of the brown dwarf (left to right panels). Because the rotational velocities, and hence Doppler shifts, are greatest in low-latitude regions of the brown dwarf, a spot near the equator produces a perturbation that migrates from the extreme left wing to the extreme right wing. Spots at higher latitude exhibit smaller Doppler shifts and thus produce perturbations that begin and end closer to the line centre. In this way, time-resolved spectra, such as those obtained by Crossfield *et al.*[1], can be used to construct a map of the spot distributions on the surface of the brown dwarf. (Figure adapted from ref. 3.)

transport the enormous heat flux that they radiate into space, will trigger waves and turbulence in the atmosphere[7,8] that could potentially organize into coherent, large-scale weather features such as those seen in Crossfield and co-workers' maps. Jupiter's Great Red Spot — a vast, centuries-old vortex — and Saturn's recent massive convective storm[9] provide useful analogies.

That said, it is currently unclear how far the analogy with Jupiter extends. Although brown dwarfs are Jupiter-like in many ways, they radiate heat fluxes that are orders of magnitude greater. Recent work[10] suggests that, under these radiative conditions, the atmospheric circulation may comprise turbulence and vortices with no preferred directionality, rather than a banded pattern with multiple east–west jet streams like that of Jupiter and Saturn. Unfortunately, Crossfield and colleagues' analysis does not resolve this crucial issue; a well-known bias makes it a particular challenge to confidently infer banded patterns with the Doppler-imaging technique. Still, future attempts will be welcome, and, if successful, they could have implications for the interpretation of brown-dwarf variability as well as theories of atmospheric dynamics generally, including the multi-decade effort to build a theory for Jupiter's and Saturn's jet streams.

There are other caveats. The signal-to-noise ratio in the authors' maps is modest, and only a few of the largest atmospheric structures in the maps are statistically robust. The observations — which are based on carbon monoxide spectral lines at a wavelength near 2 micrometres — do not establish whether the patchiness results from spatial variations of clouds, temperature or chemistry, although the first is most likely, and observations at other wavelengths can break this degeneracy. Moreover, because Luhman 16B and its companion are the brightest brown dwarfs in the sky, they are the only ones to which the Doppler-imaging technique can currently be applied. Despite the caveats, these are exciting times for brown-dwarf science. The next few years should see the workings of these fascinating worlds gradually come into focus. ■

Adam P. Showman *is in the Department of Planetary Sciences, Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.*
e-mail: showman@lpl.arizona.edu

1. Crossfield, I. J. M. *et al. Nature* **505,** 654–656 (2014).
2. Luhman, K. L. *Astrophys. J.* **767,** L1 (2013).
3. Rice, J. B. *Astron. Nachr.* **323,** 220–235 (2002).
4. Artigau, E., Bouchard, S., Doyon, R. & Lafreniere, D. *Astrophys. J.* **701,** 1534–1539 (2009).
5. Gillon, M. *et al. Astron. Astrophys.* **555,** L5 (2013).
6. Biller, B. A. *et al. Astrophys. J.* **778,** L10 (2013).
7. Showman, A. P. & Kaspi, Y. *Astrophys. J.* **776,** 85 (2013).
8. Freytag, B., Allard, F., Ludwig, H.-G., Homeier, D. & Steffen, M. *Astron. Astrophys.* **513,** A19 (2010).
9. Sayanagi, K. M. *et al. Icarus* **223,** 460–478 (2013).
10. Zhang, X. & Showman, A. P. *223rd Am. Astron. Soc. Meet.*, Washington DC, abstr. 424.03 (2014).

EVOLUTIONARY BIOLOGY

Brotherly love benefits females

Mating competition between males often has harmful consequences for females. But it seems that fruit flies alter their behaviour among kin, with brothers being less aggressive and females reproducing for longer as a result. SEE LETTER P.672

SCOTT PITNICK & DAVID W. PFENNIG

The romantic notion of sexual reproduction as a cooperative endeavour has been trampled on by a growing number of cases in which sexual competition between males results in harm to females[1]. Examples include spiny-beetle penises that punch holes in the female reproductive tract, female frogs drowning as several males struggle to mount them, and toxic ejaculate proteins that reduce a female fruit fly's desire to re-mate and can cause her early death. Such costs incurred by females represent the collateral damage of male–male competition for access to successful reproduction[2]. But the picture is complicated when the competing males are related, because of the evolutionary benefit to an individual if a relative reproduces. Theory suggests that male relatedness should reduce sexual harm to females. In this issue, Carazo *et al.*[3] (page 672) show experimentally that this is indeed the case in the fruit fly *Drosophila melanogaster*.

Sexual harm to females is a 'reproductive tragedy of the commons' that may reduce a population's productivity and even lead to local extinctions[4]. But conflict and cooperation in social interactions lie along a continuum, and resolving the evolutionary pressures that move populations along this continuum is a major challenge. One such pressure is genetic relatedness among males. Natural selection favours individuals that are most successful at propagating their distinctive genes; these individuals are said to have the highest 'fitness'. However, an individual's overall ('inclusive') fitness is the sum of its direct fitness, which is the number of offspring it produces, and its indirect fitness, which includes the number of offspring produced by the individual's genetic relatives as a result of its behaviour. Essentially, by helping its genetic relatives to reproduce, an individual indirectly

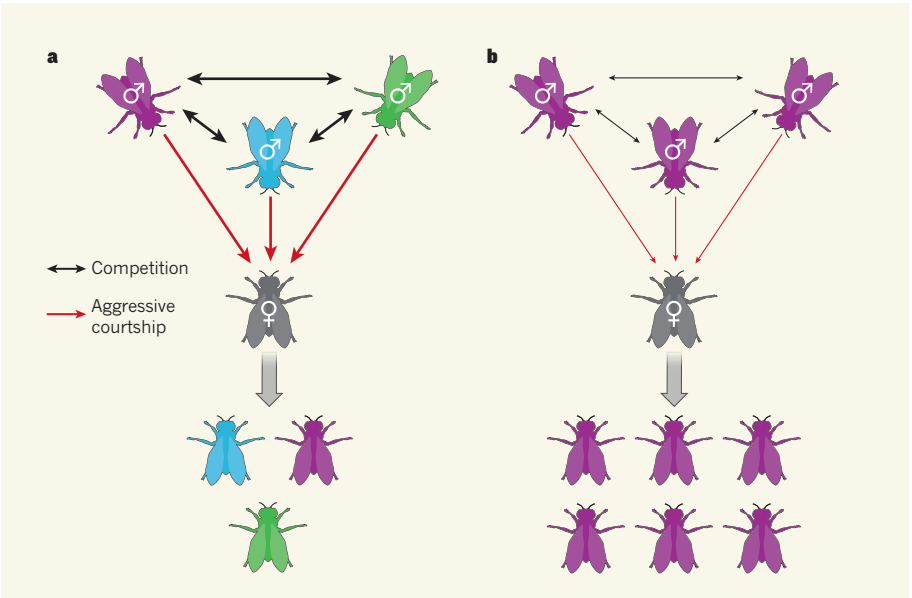


Figure 1 | Kindness to kin reduces harm to females. **a**, Unrelated male fruit flies compete with each other and court females aggressively. Carazo *et al.*[3] find that this behaviour harms females by causing them to age rapidly (in reproductive terms) and ultimately to produce fewer offspring. **b**, By contrast, the authors observe that brothers compete and court less aggressively; consequently, the females are reproductively successful for longer and produce more offspring. This reduced aggression between brothers also benefits the males: by helping his brothers to reproduce, a male indirectly propagates copies of some of his own genes.

propagates copies of some of its own genes[5].

It has been proposed that kin selection — natural selection that increases indirect fitness — can explain why males sometimes reduce the harm incurred by their mates[4,6]. Specifically, when kin compete, any harm imposed on a female should detrimentally affect the males' inclusive fitness by reducing the reproductive output of their male relatives. So, by favouring reduced competition between related males, kin selection should limit collateral harm to females. Although sexual cooperation between related males has been extensively studied in vertebrates[7,8], the fitness consequences for females have received little attention.

In a series of experiments, Carazo *et al.* paired one female with three males that were unrelated to the female, but that varied in relatedness to one another. The authors found that females paired with male triplets that were full siblings (AAA) had greater lifetime reproductive success than females paired with three males that were unrelated to each other (ABC). This difference was not a result of AAA-treatment females having higher fecundity or a longer lifespan, but rather because they exhibited reduced reproductive senescence — that is, their rate of offspring production declined with age more slowly than did that of females exposed to unrelated males. The researchers show that this pattern was attributable, at least in part, to a significantly slower decline in the survival of offspring as AAA- compared with ABC-treated females aged (Fig. 1).

The authors next sought to uncover the mechanisms underlying the reduced reproductive senescence of females when paired with brothers, by quantifying how males interact with the female and with one another. Again, females were randomly assigned to AAA or ABC trios of males, with the addition of a third, intermediary treatment of two full siblings and one unrelated male (AAB). As predicted by kin-selection theory, fighting between males was more common in ABC triplets than in either of the other conditions (Fig. 1). ABC males also courted females more intensely than AAA males. However, there were no treatment-related differences in mating rates. These observations suggest that harm to females is mediated by the aggressive behaviour of unrelated males towards each other and to females, reinforcing earlier findings[9].

One might propose that ABC males harm their mates by adjusting the contents of their ejaculate. For example, the seminal-fluid hormone Acp70A can reduce female lifespan, and *D. melanogaster* males are adept at facultatively adjusting both the sperm and seminal-fluid content of their ejaculates[10,11]. But Carazo *et al.* ruled out this explanation. They quantified female post-mating behaviours that are influenced by ejaculate content (latency to re-mating, and egg-laying rate) and found no differences between females inseminated by AAA compared with ABC males. Thus, the beneficial consequences of kin selection seem to involve pre-mating sexual selection. Nevertheless, another experiment revealed dramatic post-copulatory consequences of male competitive behaviour. By combining two brothers with one unrelated male (AAB), the authors found that the unrelated male did not court or mate more frequently than either of the brothers, yet sired on average twice as many offspring! Although the mechanism underlying this dramatic pattern remains a mystery, the evolutionary implications are clear: the gentler behaviour among brothers that reduces premature ageing of females is evolutionarily unstable. Such kindness will not be rewarded whenever selfish, unrelated males join the group.

Drosophila melanogaster has been an important model system for studying myriad topics in evolutionary biology, including sexual selection and sexual conflict, but not kin selection. Natural fruit-fly populations are typically large, and individuals are thought to disperse widely within their environment, so there would presumably be little opportunity for interaction among relatives. Yet Carazo and colleagues' findings suggest that *D. melanogaster* populations might occasionally be (or have been) structured such that they could be influenced by kin selection. We hope that this surprising and compelling study will tempt more *Drosophila* biologists to leave the laboratory to explore the ecology of this model system. ■

Scott Pitnick *is in the Department of Biology, Syracuse University, Syracuse, New York 13244, USA.* **David W. Pfennig** *is in the Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA.*
e-mails: sspitnic@syr.edu; dpfennig@unc.edu

1. Arnqvist, G. & Rowe, L. *Sexual Conflict* (Princeton Univ. Press, 2005).
2. Morrow, E. H., Arnqvist, G. & Pitnick, S. *Behav. Ecol.* **14,** 802–806 (2003).
3. Carazo, P., Tan, C. K. W., Allen, F., Wigby, S. & Pizzari, T. *Nature* **505,** 672–675 (2014).
4. Rankin, D. J., Dieckmann, U. & Kokko, H. *Am. Nat.* **177,** 780–791 (2011).
5. Hamilton, W. D. *J. Theor. Biol.* **7,** 1–16, 17–52 (1964).
6. Pizzari, T. & Gardner, A. *Phil. Trans. R. Soc. B* **367,** 2314–2323 (2012).
7. Solomon, N. G. & French, J. A. (eds) *Cooperative Breeding in Mammals* (Cambridge Univ. Press, 2007).
8. Concannon, M. R., Stein, A. C. & Uy, J. A. C. *Mol. Ecol.* **21,** 1477–1486 (2012).
9. Partridge, L. & Fowler, K. *J. Insect Physiol.* **36,** 419–425 (1990).
10. Lüpold, S., Manier, M. K., Ala-Honkola, O., Belote, J. M. & Pitnick, S. *Behav. Ecol.* **22,** 184–191 (2011).
11. Sirot, L. K., Wolfner, M. F. & Wigby, S. *Proc. Natl Acad. Sci. USA* **108,** 9922–9926 (2011).

This article was published online on 22 January 2014.

ATOMIC PHYSICS

Polar exploration

Magnetic monopoles — particles carrying a single magnetic charge — have never been seen. Analogues of these entities have now been produced in an ultracold cloud of rubidium atoms. SEE LETTER P.657

LINDSAY J. LEBLANC

If you have ever broken a magnet in two, you will know that each of the new pieces has a 'north' and a 'south' pole — just like the original. Despite being allowed in theory, a north pole separated from its south to create an isolated magnetic monopole has not been found. On page 657 of this issue, Ray *et al.*[1] report how they have created a 'Dirac monopole' by engineering an environment that mimics a monopole's magnetic field in a cloud of rubidium atoms. Using direct imaging, the authors observe a distinct signature of the Dirac monopole in this quantum system: a line of zero atomic density that pierces the cloud and terminates at the monopole. This 'Dirac string' is a defect that allows the system's quantum-mechanical phase to satisfy constraints imposed by the monopole's characteristic geometry and the wave-like nature of matter.

The duality of electric and magnetic fields in classical electromagnetism makes it especially surprising that no magnetic monopole has been found to complement the electric charge. In his 1931 paper[2], Paul Dirac showed that the theory of quantum mechanics, like its classical counterpart, allows the existence of monopoles. Furthermore, he demonstrated that if even a single monopole exists, electrical charge must come in discrete packets, which provides a possible explanation for the well-established observation that electrical charge is quantized. Although experiments have failed to find definitive evidence for the magnetic monopole[3], researchers continue to seek this elusive particle with ever more powerful tools (see, for example, refs 4–6).

To explore the quantum properties of matter near a monopole, Ray and colleagues used a Bose–Einstein condensate (BEC) of ultracold rubidium atoms. A BEC is a collection

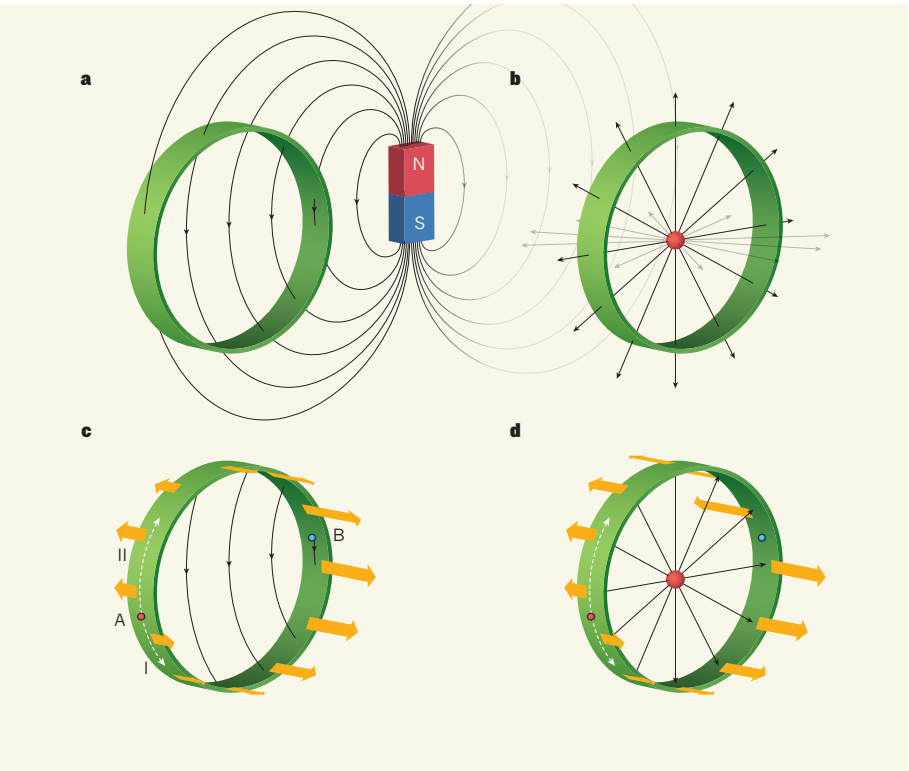


Figure 1 | Magnetic differences. Magnetic fields (black lines) produced by a conventional bar magnet (**a**) and a magnetic monopole (red sphere; **b**). Two semi-circular paths, I and II, connecting points A and B in the conventional (**c**) and monopole (**d**) magnetic fields. Magnetic-field lines in the plane of these paths are indicated. A charged particle with constant speed moving along either path experiences a Lorentz force that imparts a perpendicular velocity (orange arrows, the lengths of which are only approximate in this graphic). In the bar-magnet case, the final velocities at point B are identical, whereas in the case of the monopole they are not. The differing velocities are associated with a vortex that circulates an infinitely thin filament (not shown) extending from the monopole, with the quantum-mechanical phase along the filament being undefined. Ray *et al.*[1] have observed this filament in a Bose–Einstein condensate of rubidium atoms.

of quantum particles in which the wave-like nature of matter dominates and the ensemble behaves as a single wave. Although the wave's phase — the quantity that determines the local amplitude of the wave as it oscillates between its minimum and maximum values — is always evolving, phase relationships between different spatial points in a BEC are rigidly maintained and give the condensate well-defined long-range quantum correlations.

Magnetic fields exert a force, called the Lorentz force, on charged particles in a direction that is perpendicular to both that of the particle's velocity and the magnetic field. In quantum mechanics, the velocity at a specific point is proportional to the spatial variation in phase, and magnetic fields modify this variation to ensure that the Lorentz force is realized. Because a monopole's magnetic-field lines emerge radially from the source, its field geometry is fundamentally different from that of a conventional magnet, whose field lines have no end points (Fig. 1a, b). This geometric difference is reflected in distinct spatial phase relationships associated with each magnetic-field source.

The phase difference between two spatial points can be visualized by relating it to the change in velocity, owing to the Lorentz force, of a classical particle travelling along a trajectory between the points. For conventional magnetic fields, the change in velocity between start and finish is independent of the particle's path (Fig. 1c). By contrast, the monopole's geometry results in a path-dependent final velocity (Fig. 1d), which suggests that, because of the relationship between velocity and phase variation, the final phase of a quantum particle depends on its trajectory. In quantum mechanics, however, all paths are sampled in a journey from one point to another. Because the phase can have only a single value at each point in space, the system must account for all possibilities. One solution to this ambiguity is the emergence of an infinitely thin filament extending from the monopole, along which the phase is singular (undefined) and there is zero probability that any atom resides there. Rapid phase variations wrap around this Dirac string (see Fig. 1 of the paper[1]) and result in large, swirling velocities, which are physically manifested in a BEC as a vortex. This motion corresponds to the classical particles' acquisition of large velocities circling around the final point in the visualization described above.

In their study, Ray *et al.* produce a synthetic magnetic field in a BEC whose phase variations accompany spatial variations in intrinsic angular momentum (spin) of the BEC's atoms. Previous methods implemented synthetic fields using rotating BECs[7] or light-assisted atomic transitions[8]. To create a Dirac monopole, which is the magnetic monopole's generalization in a quantum system, the authors engineered an environment in which the preferred spin varies in space, and tailored these variations using an impressively stable and precise apparatus. With these techniques, they identified a zero-density Dirac string that terminated within the BEC at the Dirac monopole. They also observed that the phase variation around the Dirac string is consistent with predictions, and showed that the spatial distribution of the atoms' spin matches numerical calculations.

This creation of a Dirac monopole in a BEC is a beautiful demonstration of quantum simulation[9], a growing research field that uses real quantum systems to model others that are difficult to make, calculate or observe. Ray *et al.* have shown that experimental atomic-physics techniques can provide tangible systems in which to explore phenomena across disciplines. Although this technique is limited in its geometry, the authors' synthetic-magnetic-field method is free from the atomic-number losses caused by light-assisted heating that plague other techniques[8]. Their experiments will lead to further exploration of the dynamics and excitations of a Dirac monopole, and provide the promise of producing large effective magnetic fields by means of 'vortex pumping'[10], which may in turn yield analogues of quantum Hall states[11] and other exotic quantum configurations.

Although these results offer only an analogy to a magnetic monopole, their compatibility with theory reinforces the expectation that this particle will be detected experimentally. As Dirac said[2] in 1931, referring to the magnetic monopole: "under these circumstances one would be surprised if Nature had made no use of it." ■

Lindsay J. LeBlanc *is in the Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada.*
e-mail: lindsay.leblanc@ualberta.ca

1. Ray, M. W., Ruokokoski, E., Kandel, S., Möttönen, M. & Hall, D. S. *Nature* **505,** 657–660 (2014).
2. Dirac, P. A. M. *Proc. R. Soc. Lond. A* **133,** 60–72 (1931).
3. Milton, K. A. *Rep. Prog. Phys.* **69,** 1637–1711 (2006).
4. Pinfold, J. L. *Radiat. Measure.* **44,** 834–839 (2009).
5. Adrián-Martínez, S. *et al. Astropart. Phys.* **35,** 634–640 (2012).
6. Abbasi, R. *et al. Phys. Rev. D* **87,** 022001 (2013).
7. Fetter, A. L. *Rev. Mod. Phys.* **81,** 647–691 (2009).
8. Dalibard, J., Gerbier, F., Juzeliūnas, G. & Öhberg, P. *Rev. Mod. Phys.* **83,** 1523–1543 (2011).
9. Bloch, I., Dalibard, J. & Nascimbène, S. *Nature Phys.* **8,** 267–276 (2012).
10. Möttönen, M., Pietilä, V. & Virtanen, S. M. M. *Phys. Rev. Lett.* **99,** 250406 (2007).
11. Roncaglia, M., Rizzi, M. & Dalibard, J. *Sci. Rep.* **1,** 43 (2011).

REVIEW

doi:10.1038/nature12908

Solar System evolution from compositional mapping of the asteroid belt

F. E. DeMeo[1,2] & B. Carry[3,4]

Advances in the discovery and characterization of asteroids over the past decade have revealed an unanticipated underlying structure that points to a dramatic early history of the inner Solar System. The asteroids in the main asteroid belt have been discovered to be more compositionally diverse with size and distance from the Sun than had previously been known. This implies substantial mixing through processes such as planetary migration and the subsequent dynamical processes.

Although studies of exoplanetary systems have the advantage of numbers[1] to answer the question of how planetary systems are built, our Solar System has the advantage of detail. For nearly two centuries since their first discovery, asteroids have been viewed as remnants of planetary formation. Located between Mars and Jupiter in the main asteroid belt (Fig. 1), they were thought to have formed essentially where they now are[2].

Early measurements showed asteroids in the inner part of the main asteroid belt were more reflective and appear subtly 'redder' than the outer, 'bluer' ones[3–6]. In the 1980s, distinct colour groupings of major asteroid compositional types were discovered as a function of distance from the Sun[2]. In the classic theory, this was interpreted as the remnant of a thermal gradient across the main belt at the time the Solar System formed[2,7–9]. An understanding of that gradient promised to hold clues to the initial conditions during planet formation.

Yet, over the course of the discovery of over half a million asteroids since the 1980s, the idea of a static Solar System history has dramatically shifted to one of great dynamic change and mixing. Driving this view was the effect on the main asteroid belt of planetary migration models that aimed to recreate the structure of the rest of the Solar System, such as the orbits of the giant planets, Pluto and the transneptunian objects, and the Jupiter Trojan asteroids (which reside in the L4 and L5 Lagrange points of Jupiter's orbit)[10–13].

As the planetary migration models evolved, so also new compositional characteristics of the main belt were uncovered through observation that were increasingly inconsistent with the classic theory. At first, just a few rogue asteroids were found to be contaminating the distinct groupings[14–16]. Now, with tens of thousands of asteroids to analyse for which we have compositional measurements[17,18], we can see that this mixing of asteroid types is more of the rule, rather than the exception, across the main belt[19].

Today, all the newly revealed aspects of the main asteroid belt, including its orbital and compositional structure and the dynamical processes that sculpt it, contribute to a more coherent story. In modern dynamical models, the giant planets are thought to have migrated over substantial distances, shaking up the asteroids—which formed throughout the Solar System—like flakes in a snow globe, and transporting some of them to their current locations in the asteroid belt (Fig. 2). The main asteroid belt thus samples the conditions across the entire Solar System. Yet, at the same time, the Hilda asteroids (located 4 AU from the Sun between the main belt and Jupiter (one astronomical unit is approximately the Earth–Sun distance); see Fig. 1) and the Jupiter Trojans appear distinctly homogeneous, challenging us to untangle the various events of the Solar System's evolution. Our Solar System's path to creating the arrangement of the planets today and the conditions that made life on Earth possible will set the context for understanding the myriad of exoplanetary systems.

Send in the rogues

Their generally redder-to-bluer colour and compositional trend implied that asteroids tend to preserve their initial formation environment: the temperature and compositional gradient in that part of the disk at the time of planetesimal formation[2,9]. From what astronomers understood at the time (the 1980s), guided by comparison with meteorites, the reddish (with a positive slope from ultraviolet-to-visible wavelengths) ones filling the inner main belt were melted igneous bodies[20], and the bluish (with a neutral slope from the ultraviolet-to-visible) ones in the outer main belt had undergone little thermal alteration[6]. The goal of the next decade (the 1990s) was to explain how the thermal gradient could be so steep, creating such wildly different outcomes, from melted to primitive over a distance of just 1 AU (ref. 21).

That original interpretation of the compositions of reddish and bluish asteroids was wrong. In fact, direct sampling (by spacecraft[22]) of the reddish asteroid (25143) Itokawa definitively showed that it did experience some heating but was relatively primitive, compared with the previous interpretation of a melted body[23–27]. Although it was still a challenge to explain the asteroids' compositional and thermal trend from warm to cold, it was not as drastic a gradient as had been supposed.

Such compositional measurements for the largest asteroids seemed to explain the gradient better, but the few measurements becoming available for smaller objects were beginning to reveal the misfits. First was (1459) Magnya, a basaltic fragment discovered among the cold, bluish bodies[14]. Then, a handful more of these rogue igneous asteroids were found dispersed across the main asteroid belt[16,28,29]. Iron asteroids present in the main belt should have formed much closer to the Sun[15]. Primitive asteroids were discovered in the inner belt[30], and furthermore, the reddish objects occurred throughout the outer belt[31–33]. Other asteroids that appeared to be dry asteroids were discovered to contain volatiles on or just below the surface, suggesting that they formed beyond the snowline (the distance from the Sun at which the temperature is low enough for water to be ice)[34–38]. At first, these observations seemed to represent 'contamination' by individual, unusual asteroids, but gradually it has become clear that even the core groups of reddish and bluish asteroids were more broadly distributed, further challenging the classic theory of a static Solar System.

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS-16, Cambridge, Massachusetts 02138, USA. [2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology (MIT), 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [3]Institut de Mécanique Céleste et de Calcul des Éphémérides, Observatoire de Paris, UMR8028 CNRS, 77 avenue Denfert-Rochereau, 75014 Paris, France. [4]European Space Astronomy (ESA) Centre, PO Box 78, Villanueva de la Cañada 28691, Madrid, Spain.

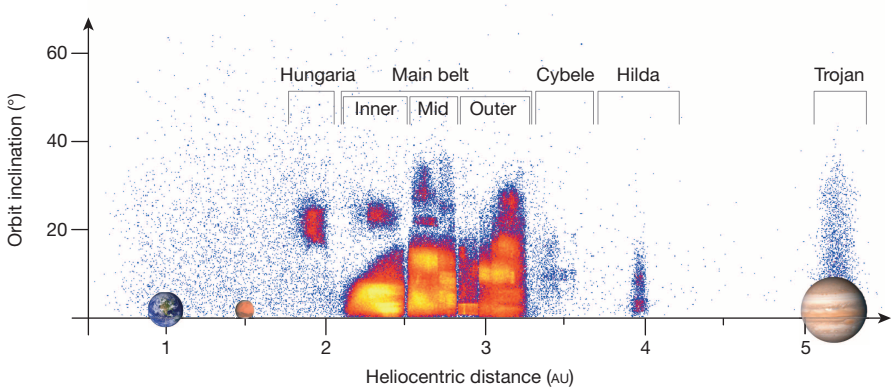


Figure 1 | The asteroid belt in context with the planets. This plot shows the location of the main belt with respect to the planets and the Sun as well as the orbital structure of asteroid inclinations and number density of objects (yellow represents the highest number density, blue the lowest). Asteroids have much higher orbital eccentricities and inclinations than do the planets. The structure of the main belt is divided by unstable regions, seen most prominently at 2.5 AU and 2.8 AU (locations where an asteroid's orbit is 'in resonance' with Jupiter's orbit), that separate the inner, middle and outer sections of the main belt. The Hungaria asteroids are located closer to the Sun than is the main belt and have orbital inclinations centred near 20 degrees. The Hildas are located near 4 AU and the Jupiter Trojans are in the L4 and L5 Lagrange points of Jupiter's orbit.

The compositional medley of asteroids

Equipped with an abundance of visible-wavelength colours and surface-brightness measurements from recent surveys[17,18] we can now reveal a new map of the distribution of asteroids down to diameters of 5 km (ref. 19) (Fig. 3). Traditionally, the distribution has been presented as the relative fraction of asteroid classes as a function of distance[2,9,31,32]. Now we compare bodies ranging from 5 km to 1,000 km in diameter, so an equal weighting would distort the view. By transforming the map of the asteroid belt to the distribution of mass[19,39], we are able to account for each asteroid type accurately, rather than the frequency or number of types (Fig. 3). Furthermore, we can now explore the change in distribution as a function of size (Fig. 4).

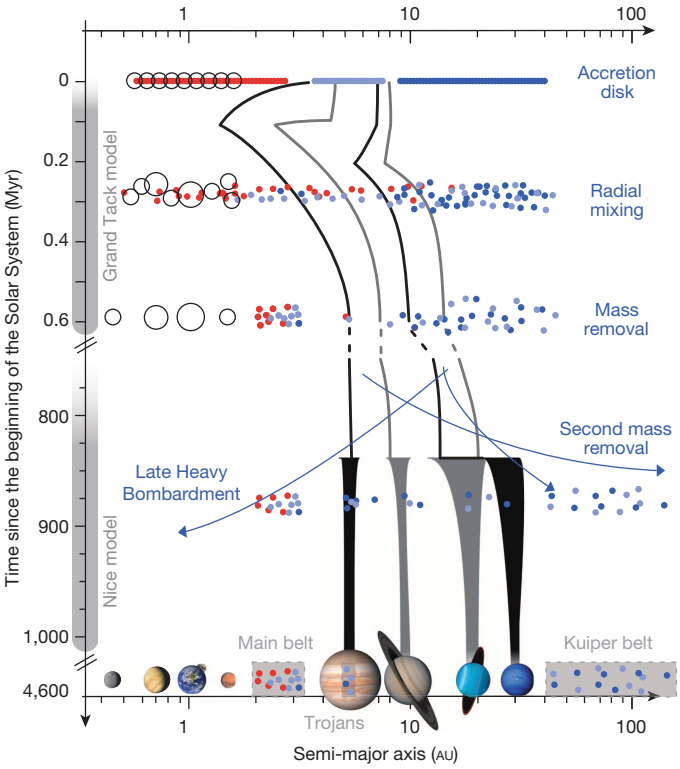


Figure 2 | Cartoon of the effects of planetary migration on the asteroid belt. This figure captures some major components of the dynamical history of small bodies in the Solar System based on models[11,12,51,54]. These models may not represent the actual history of the Solar System, but are possible histories. They contain periods of radial mixing, mass removal and planet migration—ultimately arriving at the current distribution of planets and small-body populations.

This is what we have found. The rarer asteroid types, such as the crust and mantle remnants of fully heated and melted bodies, are seen in all regions of the main belt[14,16]. We do not yet know whether this means that the locations of their respective parent bodies were ubiquitous in the inner Solar System or whether they were created close to the Sun and later injected into the main belt[15,40].

Asteroids that look compositionally Trojan-like (D-types; see Fig. 3) are detected in the inner belt, where they are not predicted to exist by dynamical models[19,41,42]. Their presence so close to the Sun demands an explanation for how they arrived there and whether they are really linked to the Trojan asteroids at all.

The Hungaria region is typically associated with its eponymous and brightest member, (434) Hungaria, and similarly super-reflective asteroids[2,8] (E-types; see Fig. 3). Despite this, most of the mass of this region is contained within a few reddish and bluish objects, which are also common elsewhere in the main belt[43–45].

The relative mass contribution of each asteroid class changes as a function of size in each region of the main belt. Most dramatic is the increase of bluish objects (C-types; see Fig. 3) as size decreases in the inner belt. Although these bluish objects are notoriously rare in the inner belt at large sizes[2,32], where they comprise only 6% of the total mass, half of the mass is bluish at the smallest sizes.

In the outer belt, reddish asteroids (S-types; see Fig. 3) make up a small fraction of the total there, yet their actual mass is still quite significant. In fact, we now find more than half of the mass of reddish objects outside the inner belt[19].

Just over a decade ago, astronomers still clung to the concept of an orderly compositional gradient across the main asteroid belt[46]. Since then the trickle of asteroids discovered in unexpected locations has turned into a river. We now see that all asteroid types exist in every region of the main belt (see Box 1 for a discussion of Hildas and Trojans). The smorgasbord of compositional types of small bodies throughout the main belt contrasts with the compositional groupings at large sizes. All these features demanded major changes in the interpretation of the history of the current asteroid belt and, in turn, of the Solar System.

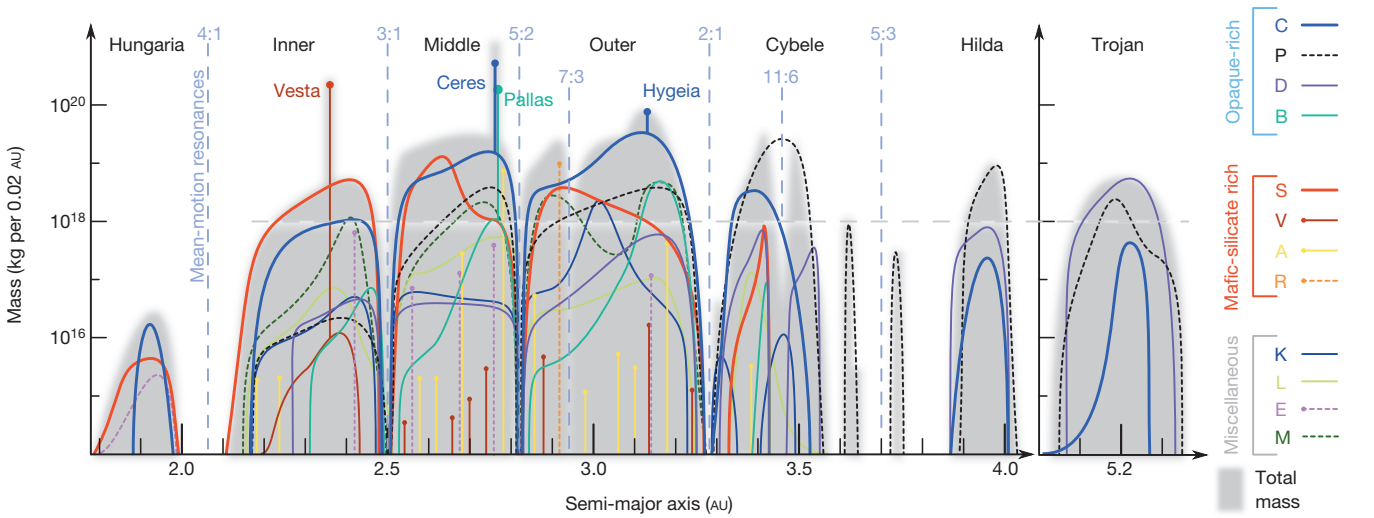


Figure 3 | The compositional mass distribution throughout the asteroid belt out to the Trojans. The grey background is the total mass within each 0.02-AU bin. Each colour represents a unique spectral class of asteroid, denoted by a letter in the key. The horizontal line at 10^{18} kg is the limit of the work from the 1980s[2,8,9]. The upper portion of the plot remains consistent with that work, but immense detail is now revealed at the lower mass range[19].

Cracking the 'compositional code' of the map

Earlier planetesimal-formation theories that explained the history of the asteroid belt invoked turbulence in the nebula, radial decay of material due to gas drag, sweeping resonances and scattered embryos[47,48]. Individually, each mechanism was, however, insufficient, and even together, although many of these mechanisms could deplete, excite and partially mix the main belt, they could not adequately reproduce the current asteroid belt[49].

The concept of planetary migration—whereby the planets change orbits over time owing to gravitational effects from the surrounding dust, gas or planetesimals—was not new, but its introduction as a major driver of the history of the asteroid belt came only recently. Migration models began by explaining the orbital structure and mass distribution of the outer Solar System, including the Kuiper belt past Neptune[50]. Individual models could successfully recreate specific parts, but we still sought to define a consistent set of events that would explain all aspects of the outer Solar System. Every action of the planets causes a reaction in the asteroid belt, so these models also needed to be consistent with the compositional framework within the main belt that we see today.

The Nice model was the first comprehensive solution that could simultaneously explain many unique structural properties of the Solar System[11–13,51,52],

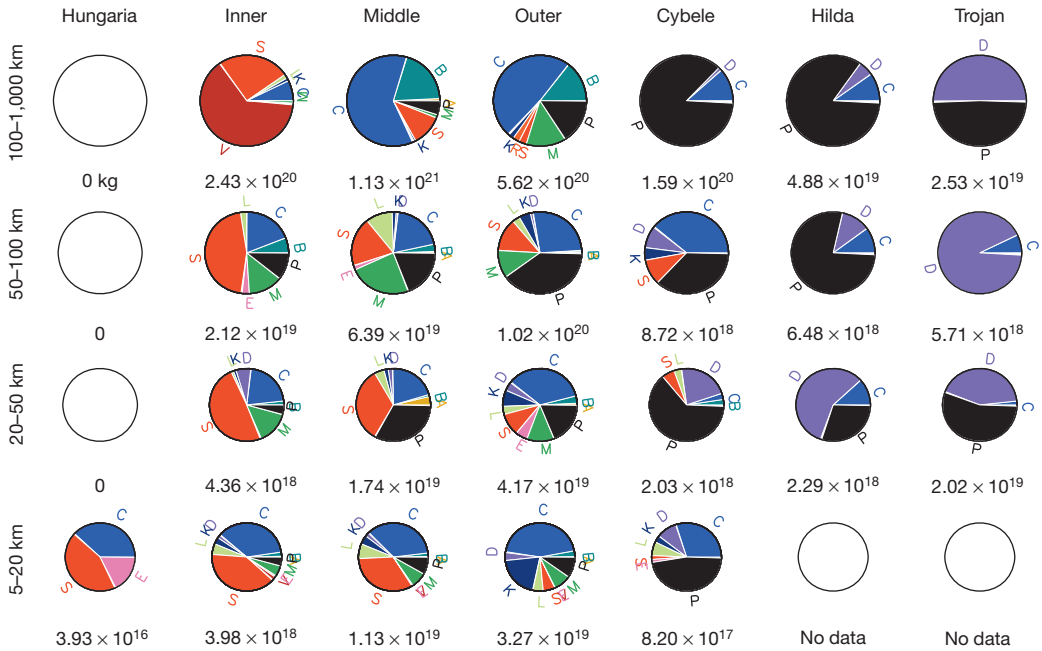


Figure 4 | The compositional mass distribution as a function of size throughout the main belt out to the Trojans. The mass is calculated for each individual object with a diameter of 50 km and greater, using its albedo to determine size and the average density[39] for that asteroid's taxonomic class. For the smaller sizes we determine the fractional contribution of each class at each size and semi-major axis, and then apply that fraction to the distribution of all known asteroids from the Minor Planet Center (http://minorplanetcenter.org/) including a correction for discovery incompleteness at the smallest sizes in the middle and outer belt[19]. Asteroid mass is grouped according to objects within four size ranges, with diameters of 100–1,000 km, 50–100 km, 20–50 km and 5–20 km. Seven zones are defined as in Fig. 1: Hungaria, inner belt, middle belt, outer belt, Cybele, Hilda and Trojan. The total mass of each zone at each size is labelled and the pie charts mark the fractional mass contribution of each unique spectral class of asteroid. The total mass of Hildas and Trojans are underestimated because of discovery incompleteness. The relative contribution of each class changes with both size and distance.

BOX 1

The Hilda and Trojan asteroids

The Hilda and Jupiter Trojan asteroids are located beyond the main asteroid belt at 4 AU and 5.2 AU, respectively (Fig. 1). The asteroid types in these regions are physically distinct from the main belt and from each other: the largest Hildas are dominated by spectral P-type asteroids and the largest Trojans are dominated by D-types[2] (Fig. 3). Despite interlopers of all types becoming more common throughout the main belt, the Hildas and Trojans remained curiously distinct and homogeneous.

Continued observations find that bright objects among Hildas and Trojans are scarce even at the small size scales[78–85]. It was recently and unexpectedly discovered, however, that the smallest bodies in these regions break rank (Fig. 4): most of the small Hildas have physical properties more similar to Trojans (D-type) and the makeup of the Trojans also changes with size with differing fractions of D-types and Hilda-like P-types[19,86,87]. Migration scenarios can now explain why the Hildas and Trojans look so different from the main belt, but they cannot yet explain the important details of why they look distinct from each other at the largest sizes (in relative fraction of the D-type and P-type asteroids) and are also different at the smaller sizes.

such as the locations of the giant planets and their orbital eccentricities[11], capture of the irregular satellites of Saturn[53], and the orbital properties of the Trojans[12] (Fig. 2). In the original model, Jupiter moves inward while the other giant planets migrate outward. As Jupiter and Saturn cross their 1:2 mean motion resonance, the system is destabilized[11]. In the most recent version of this model, the interaction between the giant planets and a massive, distant Kuiper disk causes the system to destabilize[13]. At that point, the primordial Jupiter Trojan region is emptied. Bodies that were scattered inward from beyond Neptune then repopulate this region. By reproducing the Trojans' orbital distributions and mass, the Nice model also naturally explains the why the Trojan region is compositionally distinct from the main belt: it would be populated solely by outer Solar System bodies and would not contain locally formed asteroids.

Missing from the Nice model, however, was an explanation of the large-scale mixing of reddish and bluish material in the asteroid belt that was becoming increasingly prominent. The Grand Tack model[54] showed that during the time of terrestrial planet formation (before the events of the Nice model would have taken place), Jupiter could have migrated as close to the Sun as Mars is today. Jupiter would have moved right through the primordial asteroid belt, emptying it and then repopulating it with scrambled material from both the inner and outer Solar System as Jupiter then reversed course and headed back towards the outer Solar System. Once the details of the resulting distribution in the Grand Tack model have been closely compared to the emerging observational picture, it will become clear whether this model can crack the asteroid belt's 'compositional code'.

Planetary migration ends well within the first billion years of our Solar System's 4.5-billion-year history. The asteroid belt, however, is still dynamic today. Collisions between asteroids are continuously grinding the bodies down to smaller and smaller sizes. The smaller ones ($<$40 km) are then subject to the Yarkovsky effect, according to which uneven diurnal heating and cooling of the body alters its orbit[55–59]. The Yarkovsky effect thoroughly mixes small bodies within each section of the main belt, but once they reach a major resonance—such as the 3:1 and 5:2 mean motion resonances at the locations where the orbital periods of an asteroid and of Jupiter are related by integers—they are swiftly ejected from the main belt[57–59]. Current observations[60,61] and models[28,62–64] indicate that the strong resonances with Jupiter inhibit the crossing of material from one region to another. These processes continue to mould the asteroid belt, erasing some of its past history and creating new structures in this complex system.

New observational evidence that reveals a greater mixing of bodies supports the idea of a Solar System that was and continues to be in a state of evolution and flux. Indeed, dynamical models have been leading us step-by-step to interpret the asteroid belt as a melting pot of bodies arriving from diverse backgrounds. Dynamical models have come a long way, but they have yet to explain the dichotomy between the orderly trend among the largest asteroids and the increased mixing of asteroid types at smaller sizes. Particularly noticeable are the scatter of igneous bodies and the existence of asteroids that look physically similar to Trojans in the inner belt. These details promise to teach us how our Solar System was built, providing context for other planetary systems.

The future

The ultimate goal of asteroid studies is to complete the picture of where these bodies formed and how they relate to the current chemistry and volatile abundance on Earth. No longer is our Solar System just an isolated example, and with only a minimal speculative extrapolation, asteroid-like building blocks seem likely to have influenced countless terrestrial-like planetary systems. The ongoing hunt for Earth-like planets has as its corollary the hunt for possible signatures of asteroid-like zones and an assessment of their uniqueness or commonality in all planetary systems.

Even though we now know the asteroid distributions in the Solar System down to 5 km, we are still literally only scratching the surface of what can be known about them. Asteroid interiors are the *terra incognita* for the next generation of asteroid researchers. At present we are frustrated by the inability of most physical measurements to provide any information on the interior of an asteroid. An asteroid's interior reveals its thermal history, which constrains the initial conditions of the protoplanetary disk during planetesimal formation. NASA's Dawn spacecraft mission recently provided a glimpse inside Vesta, determining the core mass fraction of this large asteroid from shape and gravity measurements[65]. When Dawn visits Ceres, we will learn to what extent this large asteroid differentiated into an ice mantle and rocky core[66–68]. Increased measurements of asteroid densities, provided mainly by the study of binary asteroids, will help us to infer their interior structure[39].

Each of our broad asteroid classes probably encompasses a wide variety of surface compositions[69,70]. Our meteorite collection has provided significant detail about the range of asteroid compositions, but to make firm links between the asteroids and meteorites, we need to observe an asteroid in space and then measure the same body in a laboratory. This will be achieved by asteroid sample return missions that are already under way[71–73], as well as 'free sample return' by meteorite falls such as the serendipitously discovered Almahata Sitta meteorite (formerly asteroid 2008 TC_3)[74,75].

Finally, the next step in distribution trends is to complement a refined understanding of asteroid compositions with physical measurements capable of detecting that detail on a large scale. The compositional trends discussed up to now cover broad taxonomic classes and combine objects into just a few major groups that do not accurately reflect the complexity of the asteroids' original and current compositions. Higher-spectral-resolution large-scale surveys at visible[76,77] and near- to mid-infrared wavelengths combined with the already available albedo information for hundreds of thousands of asteroids would be the most realistic data set to attain over the next decade or two.

Received 10 July; accepted 22 November 2013.

1. Wright, J. T. *et al.* The Exoplanet Orbit Database. *Publ. Astron. Soc. Pacif.* **123,** 412–422 (2011).
2. Gradie, J. & Tedesco, E. Compositional structure of the asteroid belt. *Science* **216,** 1405–1407 (1982).
This is a comprehensive view of the compositional trends of the asteroid belt that served for decades as the backbone for interpretations of the main belt.
3. Fischer, H. Farbmessungen an kleinen planeten. *Astron. Nachr.* **272,** 127–147 (1941).
4. Wood, X. H. J. & Kuiper, G. P. Photometric studies of asteroids. *Astrophys. J.* **137,** 1279–1285 (1963).
5. Chapman, C. R., Johnson, T. V. & McCord, T. B. in *Proc. IAU Colloq. 12* (ed. Gehrels, T.) SP 267, 1–47 (National Aeronautics and Space Administration, 1971).
6. Chapman, C. R., Morrison, D. & Zellner, B. Surface properties of asteroids—a synthesis of polarimetry, radiometry, and spectrophotometry. *Icarus* **25,** 104–130 (1975).

7. Zellner, B., Tholen, D. J. & Tedesco, E. F. The eight-color asteroid survey—results for 589 minor planets. *Icarus* **61,** 355–416 (1985).
8. Gradie, J. C., Chapman, C. R. & Tedesco, E. F. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 316–335 (Univ. Arizona Press, 1989).
9. Bell, J. F., Davis, D. R., Hartmann, W. K. & Gaffey, M. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 921–945 (Univ. Arizona Press, 1989).
10. Malhotra, R. The origin of Pluto's peculiar orbit. *Nature* **365,** 819–821 (1993).
11. Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435,** 459–461 (2005).
12. Morbidelli, A., Levison, H. F., Tsiganis, K. & Gomes, R. Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* **435,** 462–465 (2005).
13. Levison, H. F., Morbidelli, A., Tsiganis, K., Nesvorny, D. & Gomes, R. Late orbital instabilities in the outer planets induced by interaction with self-gravitating planetesimal disk. *Astron. J.* **142,** 152 (2011).
14. Lazzaro, D. *et al.* Discovery of a basaltic asteroid in the outer main belt. *Science* **288,** 2033–2035 (2000).
15. Bottke, W. F., Nesvorny, D., Grimm, R. E., Morbidelli, A. & O'Brien, D. P. Iron meteorites as remnants of planetesimals formed in the terrestrial planet region. *Nature* **439,** 821–824 (2006).
16. Moskovitz, N. A. *et al.* The distribution of basaltic asteroids in the main belt. *Icarus* **198,** 77–90 (2008).
17. Ivezić, Z. *et al.* Solar System objects observed in the Sloan Digital Sky Survey commissioning data. *Astron. J.* **122,** 2749–2784 (2001).
This is the original article on asteroid measurements from the Sloan Digital Sky Survey (SDSS), which provided multi-filter photometry in the visible spectrum for over 100,000 asteroids.
18. Mainzer, A. *et al.* Preliminary Results from NEOWISE: an enhancement to the Wide-field Infrared Survey Explorer for Solar System science. *Astrophys. J.* **731,** 53 (2011).
This is the original article on asteroid measurements from the Wide-field Infrared Survey Explorer (WISE), which provides diameters and albedos (surface brightness) for over 100,000 asteroids.
19. DeMeo, F. E. & Carry, B. The taxonomic distribution of asteroids from multi-filter all-sky photometric surveys. *Icarus* **226,** 723–741 (2013).
This work created a new framework to quantify the compositional makeup of the asteroid belt by looking at the distribution by mass rather than numbers.
20. McCord, T. B. & Gaffey, M. J. Asteroids—surface composition from reflection spectroscopy. *Science* **186,** 352–355 (1974).
21. Grimm, R. E. & McSween, H. Y. Heliocentric zoning of the asteroid belt by aluminum-26 heating. *Science* **259,** 653–655 (1993).
22. Nakamura, T. *et al.* Itokawa dust particles: a direct link between S-type asteroids and ordinary chondrites. *Science* **333,** 1113–1116 (2011).
23. Chapman, C. R. S-type asteroids, ordinary chondrites, and space weathering: the evidence from Galileo's fly-bys of Gaspra and Ida. *Meteorit. Planet. Sci.* **31,** 699–725 (1996).
24. Sasaki, S. *et al.* Production of iron nanoparticles by laser irradiation in a simulation of lunar-like space weathering. *Nature* **410,** 555–557 (2001).
25. Chapman, C. R. Space weathering of asteroid surfaces. *Annu. Rev. Earth Planet. Sci.* **32,** 539–567 (2004).
26. Brunetto, R. *et al.* Space weathering of silicates simulated by nanosecond pulse UV excimer laser. *Icarus* **180,** 546–554 (2006).
27. Binzel, R. P., Rivkin, A. S., Bus, S. J., Sunshine, J. M. & Burbine, T. H. MUSES-C target asteroid (25143) 1998 SF36: a reddened ordinary chondrite. *Meteorit. Planet. Sci.* **36,** 1167–1172 (2001).
28. Roig, F., Ribeiro, A. O. & Gil-Hutton, R. Taxonomy of asteroid families among the Jupiter Trojans: comparison between spectroscopic data and the Sloan Digital Sky Survey colors. *Astron. Astrophys.* **483,** 911–931 (2008).
29. Carvano, J. M., Hasselmann, H., Lazzaro, D. & Mothé-Diniz, T. SDSS-based taxonomic classification and orbital distribution of main belt asteroids. *Astron. Astrophys.* **510,** A43 (2010).
30. Clark, B. E., Bell, J. F., Fanale, F. P. & O'Connor, D. J. Results of the seven-color asteroid survey: infrared spectral observations of ~50-km size S-, K-, and M-type asteroids. *Icarus* **113,** 387–402 (1995).
31. Bus, S. J. & Binzel, R. P. Phase II of the Small Main-Belt Asteroid Spectroscopic Survey: a feature-based taxonomy. *Icarus* **158,** 146–177 (2002).
32. Mothé-Diniz, T., Carvano, J. & Lazzaro, D. Distribution of taxonomic classes in the main belt of asteroids. *Icarus* **162,** 10–21 (2003).
This was the first extensive look at asteroid compositional distributions at smaller sizes; it indicates that the distribution differs from that in ref. 2.
33. Lazzaro, D. *et al.* S3OS2: the visible spectroscopic survey of 820 asteroids. *Icarus* **172,** 179–220 (2004).
34. Hsieh, H. & Jewitt, D. A population of comets in the main asteroid belt. *Science* **312,** 561–563 (2006).
35. Campins, H. *et al.* Water ice and organics on the surface of the asteroid 24 Themis. *Nature* **464,** 1320–1321 (2010).
36. Rivkin, A. S. & Emery, J. P. Detection of ice and organics on an asteroidal surface. *Nature* **464,** 1322–1323 (2010).
37. Licandro, J. *et al.* (65) Cybele: detection of small silicate grains, water-ice, and organics. *Astron. Astrophys.* **525,** A34 (2011).
38. Jewitt, D. The active asteroids. *Astron. J.* **143,** 66 (2012).
39. Carry, B. Density of asteroids. *Planet. Space Sci.* **73,** 98–118 (2012).
This work conducted a detailed analysis of 994 mass estimates and 1,500 volume determinations of 300 asteroids, demonstrating density trends per asteroid taxonomic class.
40. Elkins-Tanton, L. T., Weiss, B. P. & Zuber, M. T. Chondrites as samples of differentiated planetesimals. *Earth Planet. Sci. Lett.* **305,** 1–10 (2011).
41. DeMeo, F. E., Binzel, R. P., Carry, B., Polishook, D. & Moskovitz, N. A. Unexpected D-type interlopers in the inner main belt. *Icarus* **229,** 392–399 (2014).
42. Levison, H. *et al.* Contamination of the asteroid belt by primordial trans-Neptunian objects. *Nature* **460,** 364–366 (2009).
43. Carvano, J. M., Lazzaro, D., Mothé-Diniz, T., Angeli, C. A. & Florczak, M. Spectroscopic survey of the Hungaria and Phocaea dynamical groups. *Icarus* **149,** 173–189 (2001).
44. Assandri, M. C. & Gil-Hutton, R. Surface composition of Hungaria asteroids from the analysis of the Sloan Digital Sky Survey colors. *Astron. Astrophys.* **488,** 339–343 (2008).
45. Warner, B., Harris, A. W., Vokrouhlicky, D., Nesvorny, D. & Bottke, W. F. Analysis of the Hungaria asteroid population. *Icarus* **204,** 172–182 (2009).
46. Meibom, A. & Clark, B. E. Evidence for the insignificance of ordinary chondritic material in the asteroid belt. *Meteorit. Planet. Sci.* **34,** 7–24 (1999).
47. Ruzmaikina, T. V., Safronov, V. S. & Weidenschilling, S. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 681–700 (Univ. Arizona Press, 1989).
48. Petit, J. M., Chambers, J., Franklin, F. & Nagasawa, M. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 711–723 (Univ. Arizona Press, 2002).
49. O'Brien, D. P., Morbidelli, A. & Bottke, W. F. The primordial excitation and clearing of the asteroid belt—revisited. *Icarus* **191,** 434–452 (2007).
50. Gomes, R. The origin of the Kuiper Belt high-inclination population. *Icarus* **161,** 404–418 (2003).
51. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435,** 466–469 (2005).
52. Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F. & Gomes, R. Dynamics of the giant planets of the solar system in the gaseous protoplanetary disk and their relationship to the current orbital architecture. *Astron. J.* **134,** 1790–1798 (2007).
53. Nesvorny, D., Vokrouhlický, D. & Morbidelli, A. Capture of irregular satellites during planetary encounters. *Astron. J.* **133,** 1962 (2007).
54. Walsh, K. J., Morbidelli, A., Raymond, S. N., O'Brien, D. P. & Mandell, A. M. A low mass for Mars from Jupiter's early gas-driven migration. *Nature* **475,** 206–209 (2011).
55. Chesley, S. R. *et al.* Direct detection of the Yarkovsky effect by radar ranging to asteroid 6489 Golevka. *Science* **302,** 1739–1742 (2003).
56. Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Nesvorný, D. The Yarkovsky and YORP effects: implications for asteroid dynamics. *Annu. Rev. Earth Planet. Sci.* **34,** 157–191 (2006).
57. Gladman, B. J. *et al.* Dynamical lifetimes of objects injected into asteroid belt resonances. *Science* **277,** 197–201 (1997).
58. Farinella, P., Vokrouhlicky, D. & Hartmann, W. K. Meteorite delivery via Yarkovsky orbital drift. *Icarus* **132,** 378–387 (1998).
59. Bottke, W. F., Vokrouhlicky, D., Broz, M., Nesvorny, D. & Morbidelli, A. Dynamical spreading of asteroid families by the Yarkovsky effect. *Science* **294,** 1693–1696 (2001).
60. Parker, A. *et al.* The size distributions of asteroid families in the SDSS Moving Object Catalog 4. *Icarus* **198,** 138–155 (2008).
61. Masiero, J. R. *et al.* Asteroid family identification using the hierarchical clustering method and WISE/NEOWISE physical properties. *Astrophys. J.* **770,** 7 (2013).
62. Wetherill, G. W. Steady state populations of Apollo-Amor objects. *Icarus* **37,** 96–112 (1979).
63. Wisdom, J. Chaotic behavior and the origin of the 3/1 Kirkwood gap. *Icarus* **56,** 51–74 (1983).
64. Nesvorný, D. *et al.* Fugitives from the Vesta family. *Icarus* **193,** 85–95 (2008).
65. Russell, C. T. *et al.* Dawn at Vesta: testing the protoplanetary paradigm. *Science* **336,** 684–686 (2012).
66. Thomas, P. C. *et al.* Differentiation of the asteroid Ceres as revealed by its shape. *Nature* **437,** 224–226 (2005).
67. Carry, B. *et al.* Near-infrared mapping and physical properties of the dwarf-planet Ceres. *Astron. Astrophys.* **478,** 235–244 (2008).
68. Castillo-Rogez, J. Ceres—neither a porous nor salty ball. *Icarus* **215,** 599–602 (2011).
69. Gaffey, M. J. *et al.* Mineralogical variations within the S-type asteroid class. *Icarus* **106,** 573–602 (1993).
70. Tholen, D. J. & Barucci, M. A. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 1139–1150 (Univ. Arizona Press, 1989).
71. Lauretta, D. S. *et al.* OSIRIS-REx—exploration of asteroid (101955) 1999 RQ36. *AGU Fall Meet. Abstr.* P21E–01 (2011).
72. Yano, H. *et al.* Hayabusa's follow-on mission for surface and sub-surface sample return from a C-type NEO. In *38th COSPAR Scientific Assembly* 635; http://adsabs.harvard.edu/abs/2010cosp...38..635Y (2010).
73. Barucci, M. A. *et al.* MarcoPolo-R near earth asteroid sample return mission. *Exp. Astron.* **33,** 645–684 (2012).
74. Jenniskens, P. *et al.* The impact and recovery of 2008 TC_3. *Nature* **458,** 485–488 (2009).
75. Brown, P. *et al.* The fall of the Grimsby meteorite—I: Fireball dynamics and orbit from radar, video, and infrasound records. *Meteorit. Planet. Sci.* **46,** 339–363 (2011).
76. Mignard, F. *et al.* The Gaia Mission: expected applications to asteroid science. *Earth Moon Planets* **101,** 97–125 (2007).
77. Jones, R. L. *et al.* Solar System science with LSST. *Earth Moon Planets* **105,** 101–105 (2009).
78. Emery, J. P. & Brown, R. H. Constraints on the surface composition of Trojan asteroids from near-infrared (0.8–4.0 μm) spectroscopy. *Icarus* **164,** 104–121 (2003).
79. Emery, J. P. & Brown, R. H. The surface composition of Trojan asteroids: constraints set by scattering theory. *Icarus* **170,** 131–152 (2004).

80. Emery, J. P., Burr, D. M. & Cruikshank, D. P. Near-infrared spectroscopy of Trojan asteroids: evidence for two compositional groups. *Astron. J.* **141,** 25 (2011).
81. Yang, B. & Jewitt, D. Spectroscopic search for water ice on Jovian Trojan asteroids. *Astron. J.* **134,** 223–228 (2007).
82. Yang, B. & Jewitt, D. A near-infrared search for silicates in Jovian Trojan asteroids. *Astron. J.* **141,** 95 (2011).
83. Fornasier, S. *et al.* Visible spectroscopic and photometric survey of L5 Trojans: investigation of dynamical families. *Icarus* **172,** 221–232 (2004).
84. Fornasier, S. *et al.* Visible spectroscopic and photometric survey of Jupiter Trojans: final results on dynamical families. *Icarus* **190,** 622–642 (2007).
85. Gil-Hutton, R. & Brunini, A. Surface composition of Hilda asteroids from the analysis of the Sloan Digital Sky Survey colors. *Icarus* **193,** 567–571 (2008).
86. Grav, T. *et al.* WISE/NEOWISE observations of the Hilda population: preliminary results. *Astrophys. J.* **744,** 197 (2012).
87. Grav, T., Mainzer, A. K., Bauer, J. M., Masiero, J. R. & Nugent, C. R. WISE/NEOWISE observations of the Jovian Trojan population: taxonomy. *Astrophys. J.* **759,** 49 (2012).

Acknowledgements We are grateful to R. Binzel for help in shaping this review, and to K. Walsh, W. Bottke, N. Moskovitz, D. Polishook, T. Burbine, J. Wisdom and A. Morales for discussions. We thank C. Chapman for a review. We acknowledge support from the ESAC faculty for F.E.D.'s visit. This material is based upon work supported by the National Science Foundation under grant number 0907766 and by the National Aeronautics and Space Administration (NASA) under grant number NNX12AL26G. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or NASA. Support for this work was provided by NASA through the Hubble Fellowship grant HST-HF-51319.01-A, awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. F.E.D. is a Hubble Fellow.

Author Contributions Both authors worked jointly on the scientific analysis that resulted in Figs 3 and 4. F.E.D. led the manuscript writing effort and B.C. created the figures.

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 F.E.D. (fdemeo@cfa.harvard.edu).

ARTICLE

doi:10.1038/nature12943

The evolution of lncRNA repertoires and expression patterns in tetrapods

Anamaria Necsulea[1,2]†, Magali Soumillon[1,2]†, Maria Warnefors[1,2], Angélica Liechti[1,2], Tasman Daish[3], Ulrich Zeller[4], Julie C. Baker[5], Frank Grützner[3] & Henrik Kaessmann[1,2]

Only a very small fraction of long noncoding RNAs (lncRNAs) are well characterized. The evolutionary history of lncRNAs can provide insights into their functionality, but the absence of lncRNA annotations in non-model organisms has precluded comparative analyses. Here we present a large-scale evolutionary study of lncRNA repertoires and expression patterns, in 11 tetrapod species. We identify approximately 11,000 primate-specific lncRNAs and 2,500 highly conserved lncRNAs, including approximately 400 genes that are likely to have originated more than 300 million years ago. We find that lncRNAs, in particular ancient ones, are in general actively regulated and may function predominantly in embryonic development. Most lncRNAs evolve rapidly in terms of sequence and expression levels, but tissue specificities are often conserved. We compared expression patterns of homologous lncRNA and protein-coding families across tetrapods to reconstruct an evolutionarily conserved co-expression network. This network suggests potential functions for lncRNAs in fundamental processes such as spermatogenesis and synaptic transmission, but also in more specific mechanisms such as placenta development through microRNA production.

Evolutionary analyses of protein-coding gene sequences[1] and expression patterns[2] have provided important insights into the genetic basis of lineage-specific phenotypes and into individual gene functions. For lncRNAs, such analyses remain scarce, despite growing interest in these genes. Recent studies have identified thousands of lncRNAs in human[3–5], mouse[6–9], fruitfly[10], nematode[11] and zebrafish[12]. Although most lncRNAs have unknown functions, some are involved in fundamental processes like X-chromosome dosage compensation[13], genomic imprinting[14], cellular pluripotency and differentiation[15]. As a class, lncRNAs seem to be versatile expression regulators that recruit chromatin-modifying complexes to specific locations[16], enhance transcription in *cis*[17] or provide decoy targets for microRNAs (miRNAs)[18]. Thus, lncRNA evolutionary studies can also be informative in the wider scope of regulatory networks evolution.

Although several highly conserved lncRNAs are known[19], lncRNAs generally have modest sequence conservation[6,20,21]. Furthermore, in mouse liver, lncRNA transcription undergoes rapid evolutionary turnover[22]. These observations suggest that many lncRNAs may have no biological relevance. Detailed evolutionary analyses can clarify lncRNA functionality, but such analyses have been hampered by lack of annotations in non-model organisms.

The evolutionary history of lncRNAs in 11 tetrapods

We used RNA sequencing (RNA-seq) to determine lncRNA repertoires of 11 tetrapod species. We analysed 185 samples and approximately 6 billion RNA-seq reads (Supplementary Table 1), representing the polyadenylated transcriptomes of 8 organs (cortex or whole brain, cerebellum, heart, kidney, liver, placenta, ovary and testes) and 11 species (human, chimpanzee, bonobo, gorilla, orangutan, macaque, mouse, opossum, platypus, chicken and frog), which diverged approximately 370 million years (Myr) ago[23]. We included 47 strand-specific samples (approximately 2 billion reads), which allowed us to confirm gene orientation and to predict antisense transcripts (Methods).

Using this data set, we recovered spliced transcripts for most known genes (Extended Data Table 1a and Supplementary Discussion). We evaluated the protein-coding potential of transcripts using genome-wide codon substitution frequency scores (CSF[24]) and the presence of sequence similarity with known proteins and protein domains (Methods), obtaining correct classifications for approximately 96% of protein-coding genes and 97% of known noncoding RNAs, on average (Extended Data Table 1b). We thus identified between approximately 3,000 and 15,000 multi-exonic lncRNAs in each species, including known lncRNAs for human[4,5] and mouse[6], as well as approximately 10,000 novel human and 9,000 novel murine lncRNAs (Fig. 1a and Extended Data Table 2). Although part of the variability in lncRNA repertoire size may be biologically meaningful, much is likely to be explained by unequal sequencing depth and by variable genome sequence and assembly quality (Supplementary Discussion).

We reconstructed homologous families based on sequence similarity and we inferred a stringent minimum evolutionary age of lncRNAs, requiring transcription evidence as an additional criterion (Methods). We also estimated a 'maximum' evolutionary age by explicitly accounting for between-species variations in RNA-seq coverage and annotation quality (Methods and Extended Data Table 3a). We thus identified 13,533 lncRNA families transcribed in at least 3 species. Most (81%) lncRNA families were primate-specific, but 2,508 (19%) families likely originated more than 90 Myr ago and 425 (3%) more than 300 Myr ago (Fig. 1a). Most homologous lncRNAs were found in conserved synteny, even for distantly related species (Extended Data Table 3b).

The large proportion of inferred young lncRNAs may be due to fast lncRNA evolution, which prevents detection of distant homologues. Furthermore, the phylogenetic distribution of the species in our data set may contribute to the skewed distribution of estimated ages. To investigate these possibilities, we evaluated DNA sequence conservation across placental mammals[25] and variation within populations[26]

[1]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. [2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [3]The Robinson Institute, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005, Australia. [4]Department of Systematic Zoology, Faculty of Agriculture and Horticulture, Humboldt University Berlin, 10099 Berlin, Germany. [5]Department of Genetics, Stanford University School of Medicine, Stanford University, Stanford, California 94305, USA. †Present addresses: Laboratory of Developmental Genomics, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (A.N.); Harvard Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA, and Broad Institute, Cambridge, Massachusetts 02142, USA (M.S.).

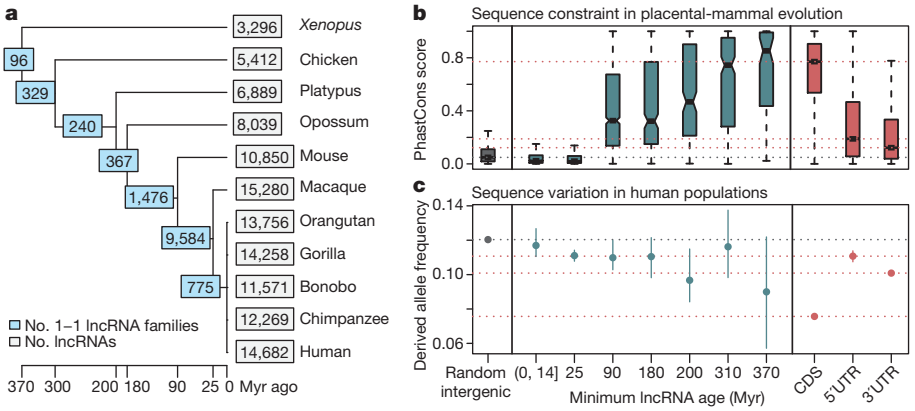


Figure 1 | Evolutionary age and genomic characteristics of lncRNA families. **a**, Simplified phylogenetic tree. Internal branches and root, numbers of 1–1 orthologous lncRNA families for each minimum evolutionary age. Tree tips, lncRNA numbers for each species. **b**, Exonic sequence conservation (placental PhastCons score), for random intergenic regions, lncRNA evolutionary age classes, coding (CDS) and untranslated exons of protein-coding genes. **c**, Mean derived allele frequency of autosomal non-CpG single nucleotide polymorphisms (SNPs) segregating in African populations (1000 Genomes Project[26]). Intergenic SNPs were randomly drawn in regions matching lncRNA recombination rates (Methods). Error bars, 95% confidence intervals based on 100 bootstrap resampling replicates. Round brackets indicate that the boundary is excluded from the interval; square brackets indicate that the boundary is included in the interval.

for human lncRNAs (Fig. 1b, c and Methods). We found that young lncRNAs (inferred minimum ages of 25 Myr or younger) have low levels of long-term exonic sequence conservation (median score ~0.02), significantly lower than random intergenic regions (median score ~0.05, Wilcoxon text, $P < 10^{-10}$). However, single nucleotide polymorphisms found in primate-specific (minimum evolutionary age 25 Myr) lncRNA exons have significantly lower derived allele frequencies (mean 0.11) than those found in intergenic regions (mean 0.12, randomization test, $P < 0.01$), consistent with recent purifying selection[27]. The same conclusions were reached using maximum evolutionary age estimates (Extended Data Fig. 1a, b), and when controlling for GC-biased gene conversion[28] (Extended Data Fig. 1c) and for linkage to protein-coding genes (Extended Data Fig. 1d). The presence of selective constraint in recent evolution, but not on a broader timescale, is compatible with a recent origination or acquisition of novel functions for a fraction of primate-specific lncRNAs.

Overall, the two measures of selective constraint correlate with evolutionary age estimates (Fig. 1c, d). Remarkably, older lncRNAs (minimum age 90 Myr) have higher levels of long-term exonic sequence conservation than untranslated regions (UTRs), and the oldest age classes are comparable with coding exons (Fig. 1c, Wilcoxon test, $P > 0.05$). Furthermore, lncRNA promoters are as conserved as protein-coding gene promoters even for younger classes (Extended Data Fig. 1e, f), suggesting stronger selective constraints at the transcriptional level, as previously observed[8].

Active regulation of ancient lncRNAs

We next asked whether lncRNA expression patterns vary with evolutionary age. We found that lncRNAs are lowly transcribed, highly organ-specific and preferentially expressed in testes (Fig. 2a–c and Extended Data Fig. 2), consistent with previous observations[4,5]. However, the testes specificity is stronger for young lncRNAs (55%) than for old lncRNAs (46%, Fig. 2a, chi-squared test, $P < 10^{-10}$), in agreement with the hypothesis that the permissive testes chromatin favours new gene origination[29]. After testes, neural tissues generally express the largest numbers of lncRNAs (Fig. 2a and Extended Data Fig. 2), consistent with a previously reported enrichment of lncRNAs in mouse brain[9]. Surprisingly, for platypus, ovary appears to be the second most favourable tissue for lncRNA expression (Extended Data Fig. 2).

The low expression levels and the testes specificity raise the question of whether lncRNAs are actively regulated, or whether they result from non-specific transcription in open chromatin regions. To test these hypotheses, we analysed the occurrence of transcription-factor-binding sites as an indicator of active regulation. Using a genome-wide set of evolutionarily conserved binding sites predicted *in silico*[30] and ChIP-seq transcription-factor-binding data[31] (Methods), we found that lncRNA promoters were more frequently associated with transcription factors than random intergenic regions (Fig. 2d and Extended Data Fig. 3a, c). Moreover, binding site sequence conservation was stronger in lncRNA promoters than in random intergenic regions and even protein-coding gene promoters, in particular for ancient lncRNAs (Fig. 2e, Wilcoxon test $P < 10^{-10}$). Consistently, the evolutionary turnover of CEBPA and HNF4A binding[32] between human and mouse is significantly slower for lncRNAs than expected by chance (Extended Data Fig. 3f, g, Fisher's exact test $P < 10^{-10}$). Taken together, these results suggest that lncRNA

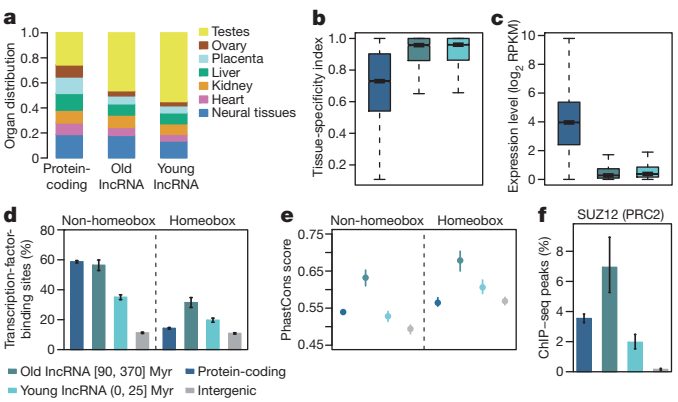


Figure 2 | lncRNA expression patterns and evidence for developmental regulation of old lncRNAs. **a**, Distribution of the organ in which maximum expression is observed, for human protein-coding genes, old lncRNAs (minimum age 90–370 Myr, 2,556 lncRNAs) and young lncRNAs (minimum age 0–25 Myr, 12,126 lncRNAs). **b**, Tissue-specificity index. Values close to 1 represent high tissue specificity. **c**, Distribution of the maximum expression level ($\log_2$-transformed RPKM). **d**, Frequency of *in silico*-predicted binding sites for homeobox and non-homeobox transcription factors, in human gene promoters (2 kb upstream) and in random intergenic regions. Error bars, 95% binomial proportion confidence intervals. **e**, Mean sequence conservation (PhastCons score) for transcription-factor-binding sites. Error bars, 95% confidence intervals based on 100 bootstrap replicates. **f**, Frequency of SUZ12 (part of the PRC2 complex) binding (ENCODE ChIP-seq). Error bars, 95% binomial proportion confidence intervals. We analysed 793 'old' lncRNAs, 3,418 'young' lncRNAs and 16,566 protein-coding genes for which the predicted transcription start site was within 100 bp of a cap analysis gene expression (CAGE) tag.

transcription is overall actively regulated, in particular for ancient lncRNAs.

Using *in silico* binding-site predictions, we also uncovered a remarkable difference between two transcription-factor classes: homeobox transcription factors, which function in embryonic development, bind preferentially in lncRNA promoters, whereas non-homeobox transcription factors bind more frequently in protein-coding promoters (Fig. 2d and Extended Data Fig. 3b). Notably, 31% of old lncRNA promoters have homeobox transcription-factor-binding sites, more than twice the frequency observed for protein-coding genes (14%, Fisher's exact test, $P < 10^{-10}$). The ChIP-seq data set consisted largely (95%) of non-homeobox transcription factors, 117 (98%) of which were associated significantly more often with protein-coding than with lncRNA promoters (Extended Data Fig. 3d). However, two factors bound more frequently in old lncRNA than in protein-coding promoters: SUZ12, a member of the polycomb repressive complex 2 (PRC2) that functions in pluripotency and differentiation[33] (Fig. 2f) and OCT4 (also known as POU5F1), a homeobox transcription factor that controls pluripotency[34] (Extended Data Fig. 3e). The association with homeobox transcription factors and PRC2 suggests that lncRNAs (especially ancient ones) may be important for embryonic development, pluripotency and differentiation[15].

Rapid evolution of lncRNA expression patterns

We next assessed the evolutionary conservation of lncRNA expression patterns. We first estimated the presence of shared transcription across species. To reduce the impact of weak lncRNA sequence conservation, we compared intergenic lncRNAs across closely related primate species (Fig. 3a) and we analysed lncRNAs transcribed in antisense of protein-coding exons (Extended Data Fig. 4a). We found that lncRNA transcription evolves rapidly: only approximately 92% of human intergenic lncRNAs were also detected as expressed in chimpanzee or bonobo and only approximately 72% were expressed in macaque, whereas more than 98% of conservation was observed for protein-coding genes, for all primates (Fig. 3a). Likewise, the evolutionary turnover of antisense lncRNAs is rapid compared to protein-coding genes (Extended Data Fig. 4a). The discrepancy between lncRNAs and protein-coding genes remained considerable when controlling for low lncRNA expression with a read resampling procedure (Fig. 3a and Extended Data Fig. 4a), indicating that rapid transcription evolution is a genuine feature of lncRNAs[22].

We also measured correlations of lncRNA expression levels between pairs of species (Fig. 3b). The difference between lncRNAs and protein-coding genes is marked (Fig. 3c): Spearman's correlation coefficient for lncRNA brain expression between human and chimpanzee (which diverged 6 Myr ago) is approximately 0.55, lower than the correlation (0.66) observed for protein-coding genes between human and *Xenopus* (which diverged ~370 Myr ago). However, low lncRNA expression levels explain much of this discrepancy, as differences between correlation coefficients for the two classes of genes were much lower after resampling controls (Fig. 3c). For both protein-coding genes and lncRNAs, the testes have the fastest rates of evolution (Extended Data Fig. 4b).

We also observed that lncRNA tissue specificity is well conserved among primates, but not beyond. Indeed, a hierarchical clustering of samples based on pairwise correlations for eutherian lncRNA families revealed preferential grouping among related organs for primates, though all mouse samples clustered together (Fig. 3c and Extended Data Fig. 4f, g). Moreover, 47% of human tissue-specific lncRNAs had

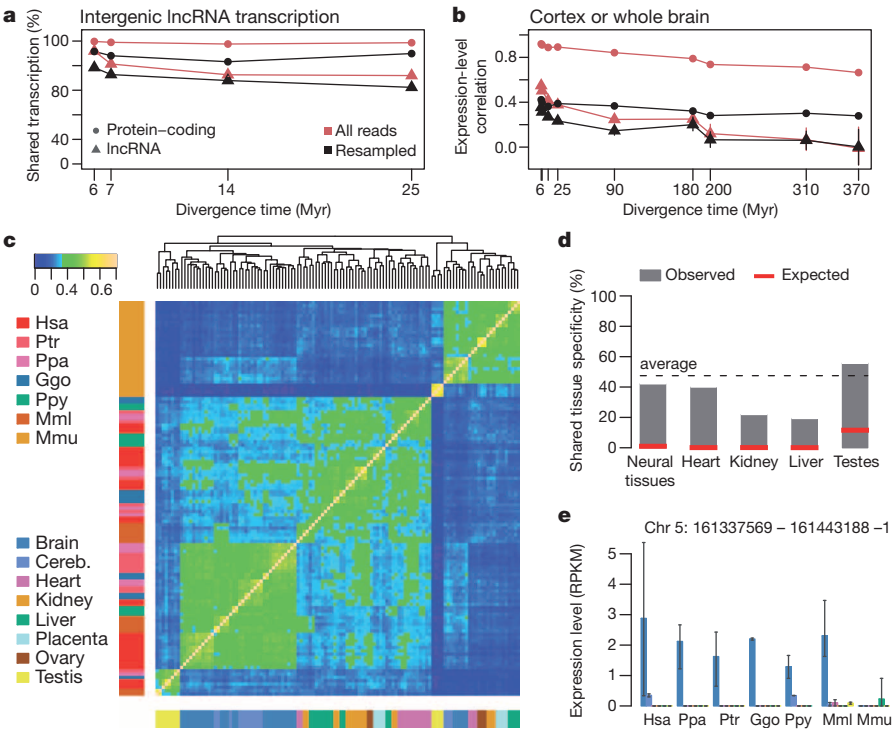


Figure 3 | Evolution of lncRNA expression patterns in tetrapods. **a**, Percentage of human lncRNAs (4,430 intergenic primate lncRNA families) transcribed in other primates, in a pool of 5 somatic tissues (Methods). **b**, Pairwise Spearman correlations between human and other species, for cortex or whole brain. In **a** and **b**, 'all reads' represents estimates obtained with all reads, and 're-sampled' represents estimates obtained after resampling identical numbers of mapped reads per species and tissue; error bars, 95% confidence intervals obtained with 100 bootstrap resampling replicates. **c**, Hierarchical clustering of pairwise Spearman correlations, for 1,716 lncRNA families with 1–1 orthologues in all eutherians. Samples are colour-coded according to the organ (horizontal) and species (vertical). Hsa, *Homo sapiens*; Ptr, *Pan troglodytes* (chimpanzee); Ppa, *Pan paniscus* (bonobo); Ggo, *Gorilla gorilla*; Ppy, *Pongo pygmaeus* (orangutan); Mml, *Macaca mulatta* (macaque); Mmu, *Mus musculus* (mouse). **d**, Proportion of human organ-specific lncRNAs (771 lncRNAs with minimum evolutionary age >90 Myr, tissue-specificity index >0.9, RPKM >0.1) for which organ specificity is shared across primates. Red lines, random expectation; dashed line, average conserved specificity across organs. **e**, A lncRNA with conserved neural tissue specificity across primates. Error bars, range observed in biological replicates. Chromosomal coordinates are given in the plot title.

conserved specificity in all primates, while only 28% had conservation across all eutherians (Fig. 3d and Extended Data Fig. 4c–e). These proportions are significantly lower than for protein-coding genes, for which 81% are conserved across all primates and 72% across all eutherians (Fisher's exact test, $P < 10^{-10}$), but higher than randomly expected (randomization test, $P < 0.01$). The extent of conservation varies among tissues (Fig. 3d and Extended Data Fig. 4c–e), but is always significantly higher than expected by chance (randomization test, $P < 0.01$). These observations are illustrated by a lncRNA identified within a cluster of GABA (γ-aminobutyric acid) receptors on human chromosome 5, expressed in neural tissues for primates, but detected only in liver in mouse (Fig. 3e).

Evolutionarily conserved co-expression network

Finally, we evaluated the co-expression of lncRNAs and protein-coding genes, which can indicate functional relatedness[35] or regulatory relationships[36]. As co-expression may also arise spuriously, we used evolutionary conservation as a criterion for significance[35]. We analysed a set of 16,076 protein-coding gene families and 1,770 lncRNA families expressed in at least 3 species (Methods). We evaluated expression correlations for all gene pairs and tested if the combination of correlation coefficients across species was significantly higher (for positive associations) or lower (for negative associations) than expected by chance[35] (Methods). The conserved co-expression relationships formed a network with 9,388 nodes (8,971 protein-coding and 417 lncRNAs) and 97,556 edges (Supplementary Table 2). The same criteria applied on randomized gene families identified only approximately 160 co-expression relationships, proving the reconstruction specificity (Supplementary Discussion).

The co-expression network can predict functional relatedness, as illustrated by the high frequency of connections within gene ontology (GO) categories: out of 115 GO categories with at least 100 members, 101 (88%) had within-category connections more often than randomly expected (Fig. 4a). To verify if the direction of network connections may also predict regulatory associations, we analysed 710 connections annotated as expression activation/inhibition relationships in the String[37] database. We found that approximately 70% of positive connections are annotated as activation relationships, significantly more than negative connections (30%, Fisher's exact test, $P = 0.01$; Extended Data Fig. 4a). Consistent with this, we found an overwhelming majority of negative connections for the *REST* and *HBP1* transcriptional repressors (Fig. 4b). Positive co-expression also often arises for genes that participate in complexes, such as the sodium channel subunit *SCNN1B* (Fig. 4b). Most (72%) network connections are positive co-expression cases. However, these occur more frequently between protein-coding genes, whereas lncRNAs have more negative connections (Fig. 4b). Interestingly, the imprinted lncRNA *H19*, which functions as a miRNA precursor[38], has a majority of negative connections (Fig. 4b).

The network connectivity depends on expression levels, as more connections were detected for highly expressed genes (Extended Data Fig. 5b, c). Expectedly, lncRNAs generally had lower connectivity (median degree 2) than protein-coding genes (median degree 5, Wilcoxon test $P < 10^{-10}$; Extended Data Fig. 5d), and transcription factors were less well connected (median degree 4) than non-transcription-factor protein-coding genes. However, when resampling genes with similar expression levels, lncRNAs had higher degrees (median 3) than protein-coding genes (median 2, randomization test $P < 0.01$), and transcription factors had higher connectivity than other protein-coding genes (median 3, randomization test $P = 0.02$; Extended Data Fig. 5d), consistent with their central roles in regulatory networks. The highly connected lncRNAs may represent interesting candidates for further studies of gene expression regulation. Notably, lncRNAs had connections in

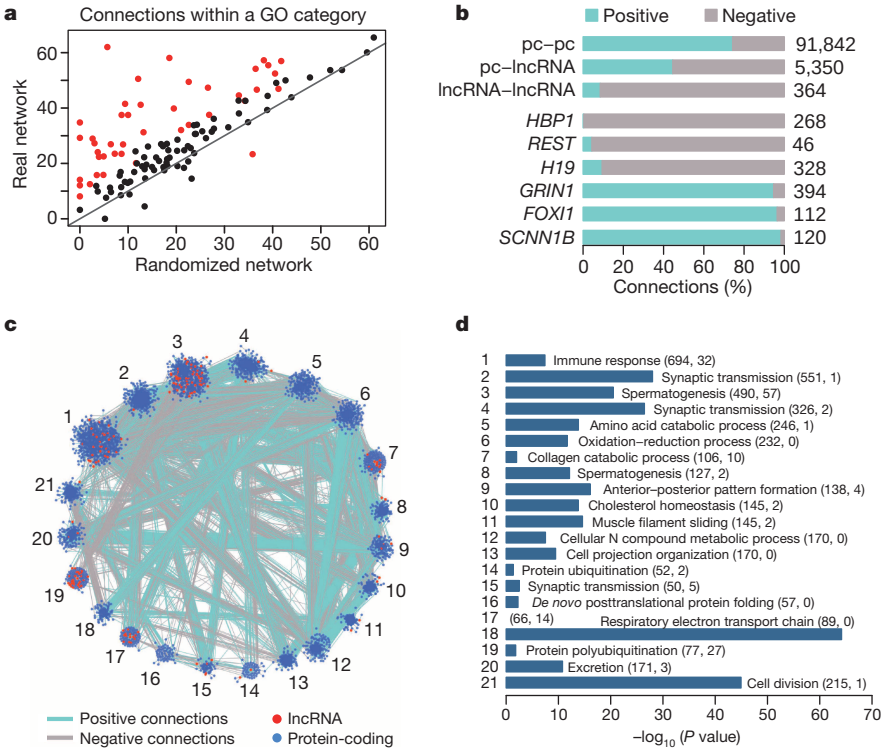


Figure 4 | Evolutionary conserved co-expression network of protein-coding genes and lncRNAs. **a**, Percentage of genes with connections within the same GO category, in real and randomized co-expression networks, for 115 biological process categories. Red, significant difference between real and randomized data ($P < 0.05$). **b**, Percentage of positive connections, for the entire network and for six genes with extreme positive:negative ratios. pc–pc, connections between two protein-coding genes; pc–lncRNA, connections between a protein-coding and a lncRNA gene; lncRNA–lncRNA, connections between two lncRNAs. Total numbers of connections per category are indicated on the right side of the plot. **c**, Cytoscape[47] representation of the 21 largest MCL clusters in the co-expression network. **d**, GO enrichment for the 21 largest MCL clusters; only the most significant GO category is displayed. Numbers of coding and lncRNA genes in each cluster are shown in brackets.

cis more often than protein-coding genes (Extended Data Fig. 5e). An excess of connections in *cis* was also found for protein-coding genes acting in body plan development, in particular for *HOX* genes (Extended Data Fig. 5e and Supplementary Table 3).

Finally, we used the co-expression network to infer potential functions for lncRNAs. Using the Markov clustering algorithm (MCL[39]), we identified 1,326 groups of highly inter-connected genes, including 21 clusters with at least 50 genes (Fig. 4c and Supplementary Table 4). The proportion of lncRNAs in these clusters varied between 0 and 26% (Fig. 4d). The clusters were enriched for organ-specific functions, such as spermatogenesis (testes), synaptic transmission (neural tissues), catabolic processes (liver), muscle functions (heart) (Methods, Fig. 4c and Supplementary Table 4). We also recovered specific processes, such as anterior–posterior pattern formation in a cluster that includes *HOX* genes (Fig. 4c). The clusters with highest lncRNA proportions were enriched in spermatogenesis functions (Fig. 4c), in agreement with the predominant lncRNA testes specificity. GO enrichment analyses for individual nodes suggested potential lncRNA involvement in, for example, nervous system development, cell adhesion, transcription (Supplementary Table 5).

miRNA precursors in the *H19* co-expression network

The only MCL cluster without significant GO enrichments (Fig. 4d) contains a high proportion (17.5%) of lncRNAs, including *H19*. As *H19* is a precursor for *miR-675*, which targets *IGF1R* and thus stalls placenta growth during late gestation[38], we scanned the network for other potential miRNA precursors (Methods). Unexpectedly, genes positively connected with *H19* had the highest average density of embedded miRNAs (Extended Data Fig. 6a). These include one exceptional case: a lncRNA that could potentially promote the transcription of between 2 and 7 miRNAs in different species (Fig. 5a, Supplementary Table 6 and Supplementary Discussion). This lncRNA (that we name *H19X*, for *H19* X-linked co-expressed lncRNA) is transcribed in all studied species and thus likely originated at least 370 Myr ago, in the tetrapod ancestor. Notably, its expression pattern appears to have dramatically shifted during evolution, from an ancestral testes-predominant pattern to preferential expression in the chorioallantoic placenta of eutherians (Fig. 5a).

The miRNAs associated with *H19X* comprise two conserved tetrapod families, four placental-mammal-specific families and one rodent-specific miRNA (Supplementary Discussion). Interestingly, the two oldest families (with representative members *miR-503*, *miR-16c*, and *miR-424*, *miR322*, *mir-15c*, respectively) seem to have undergone accelerated sequence evolution in the eutherian ancestor (Extended Data Fig. 6b). In human and mouse, these miRNAs are in general highly expressed in the placenta (Extended Data Fig. 6c, d).

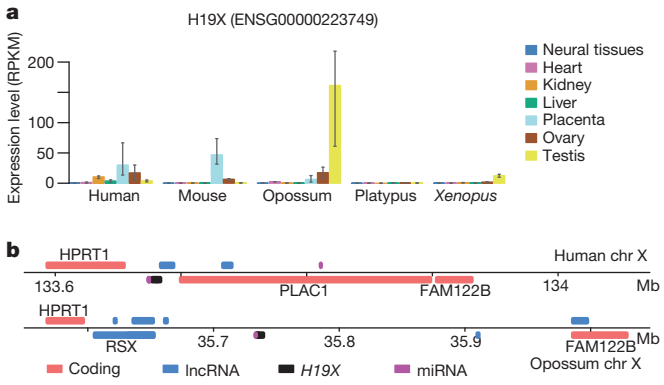


Figure 5 | *H19* co-expression network and miRNA precursors. a, Expression pattern for an X-linked *H19* co-expressed lncRNA (*H19X*, identified as ENSG00000223749 in the Ensembl database), in five tetrapod species. The error bars represent the range observed in biological replicates. **b**, Genomic neighbourhood of *H19X* in human and opossum.

Finally, *H19X* is a neighbour of *Rsx*, the lncRNA that drives imprinted X-inactivation in marsupials[40] (Fig. 5b), suggesting that *H19X* may itself be imprinted. These results suggest that *H19X* may function like *H19*, by promoting miRNA transcription, preferentially in the placenta and in an imprinted manner. Although validation is needed, this illustrates how the reconstruction of a conserved co-expression network, enabled by the broad evolutionary perspective of our study, can predict lncRNA functions and stimulate further investigations.

METHODS SUMMARY

We sequenced poly-adenylated transcriptomes of 11 species and 8 tissues with Illumina GAII and HiSeq2000 technologies. We detected multi-exonic transcripts based on transcribed island and splice junction coordinates, using TopHat[41] and Cufflinks[42]. Protein-coding potential was inferred using codon substitution frequency scores (CSF[24]) and sequence similarity with known proteins[43] and protein domains[44]. We included published lncRNA annotations for human and mouse[45] and projected annotations across species. We reconstructed homologous families based on DNA sequence similarity, with single-link clustering. We inferred lncRNA evolutionary ages based on the phylogenetic distribution of species with transcription evidence, or for which its absence was due to low coverage or incomplete annotation. We computed RPKM (reads per kilobase per million mapped reads) levels using non-overlapping exonic regions and unambiguously mapped reads, and we normalized them through median-scaling[2]. We computed tissue-specificity indexes as previously described[46]. To control for unequal coverage, we simulated read distributions by resampling identical numbers of reads per species and tissue, keeping proportions among genes unchanged. We reconstructed an evolutionarily conserved co-expression network by computing expression correlations between gene pairs and identifying cross-species combinations that are significantly higher or lower than randomly expected[35]. Network analysis was done with MCL[39] and Cytoscape[47]. For all statistics and graphics we used R[48].

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 31 December 2012; accepted 5 December 2013.
Published online 19 January 2014.

1. Kosiol, C. *et al.* Patterns of positive selection in six mammalian genomes. *PLoS Genet.* **4,** e1000144 (2008).
2. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478,** 343–348 (2011).
3. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106,** 11667–11672 (2009).
4. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).
5. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22,** 1775–1789 (2012).
6. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).
7. Guttman, M. *et al. Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnol.* **28,** 503–510 (2010).
8. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–1563 (2005).
9. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA* **105,** 716–721 (2008).
10. Young, R. S. *et al.* Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biol. Evol.* **4,** 427–442 (2012).
11. Nam, J. W. & Bartel, D. Long non-coding RNAs in *C. elegans*. *Genome Res.* (2012).
12. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147,** 1537–1550 (2011).
13. Chow, J. C., Yen, Z., Ziesche, S. M. & Brown, C. J. Silencing of the mammalian X chromosome. *Annu. Rev. Genomics Hum. Genet.* **6,** 69–92 (2005).
14. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415,** 810–813 (2002).
15. Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18,** 1433–1445 (2008).
16. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
17. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143,** 46–58 (2010).
18. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147,** 358–369 (2011).

19. Chodroff, R. A. *et al.* Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **11,** R72 (2010).
20. Marques, A. C. & Ponting, C. P. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol.* **10,** R124 (2009).
21. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).
22. Kutter, C. *et al.* Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **8,** e1002841 (2012).
23. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22,** 2971–2972 (2006).
24. Lin, M. F. *et al.* Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* **17,** 1823–1836 (2007).
25. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).
26. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **135,** 56–65 (2012).
27. Ward, L. D. & Kellis, M. Evidence of abundant purifying selection in humans for recently-acquired regulatory functions. *Science* **337,** 1675–1678 (2012).
28. Galtier, N. & Duret, L. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* **23,** 273–277 (2007).
29. Soumillon, M. *et al.* Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3,** 2179–2190 (2013).
30. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–482 (2011).
31. The Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
32. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328,** 1036–1040 (2010).
33. Walker, E., Manias, J. L., Chang, W. Y. & Stanford, W. L. PCL2 modulates gene regulatory networks controlling self-renewal and commitment in embryonic stem cells. *Cell Cycle* **10,** 45–51 (2011).
34. Chambers, I. & Tomlinson, S. R. The transcriptional foundation of pluripotency. *Development* **136,** 2311–2322 (2009).
35. Stuart, J. M., Segal, E., Koller, D., Kim, S. K. & A. Gene-coexpression network for global discovery of conserved genetic modules. *Science* **302,** 249–255 (2003).
36. Shkumatava, A., Stark, A., Sive, H. & Bartel, D. P. Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes Dev.* **23,** 466–481 (2009).
37. Franceschini, A. *et al.* STRING v9.1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41,** D808–D815 (2013).
38. Keniry, A. *et al.* The *H19* lincRNA is a developmental reservoir of miR-675 that suppresses growth and *Igf1r*. *Nature Cell Biol.* **14,** 659–665 (2012).
39. Van Dongen, S. & Abreu-Goodger, C. Using MCL to extract clusters from networks. *Methods Mol. Biol.* **804,** 281–295 (2012).
40. Grant, J. *et al. Rsx* is a metatherian RNA with *Xist*-like properties in X-chromosome inactivation. *Nature* **487,** 254–258 (2012).
41. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
42. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
43. UniProt. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40,** D71–D75 (2012).
44. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40,** D290–D301 (2012).
45. Flicek, P. *et al.* Ensembl 2012. *Nucleic Acids Res.* **40,** D84–D90 (2012).
46. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21,** 650–659 (2005).
47. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27,** 431–432 (2011).
48. R Development Core Team. *R: A language and environment for statistical computing* http://www.r-project.org (R Foundation for Statistical Computing, 2011).

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

Acknowledgements We thank L. Froidevaux and D. Cortéz for help with genome sequencing, J. Meunier for help with preliminary miRNA analyses, K. Harshman and the Lausanne Genomics Technology Facility for high-throughput sequencing support, I. Xenarios for computational support, S. Bergmann and Z. Kutalik for advice on co-expression analyses. Human embryonic and fetal material was provided by the Joint MRC/Wellcome Trust (grant 099175/Z/12/Z) Human Developmental Biology Resource (http://www.hdbr.org). The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. This research was supported by grants from the European Research Council (Starting Independent Researcher Grant 242597, SexGenTransEvolution) and the Swiss National Science Foundation (grant 31003A_130287) to H.K. A.N. was supported by a FEBS long-term postdoctoral fellowship.

Author Contributions A.N. conceived and performed all biological analyses and wrote the manuscript, with input from all authors. A.N. and M.W. processed RNA-seq data. M.S. and A.L. generated RNA-seq data. T.D. and F.G. collected platypus samples. U.Z. collected opossum samples. J.C.B. provided mouse placenta samples and contributed to *H19X* analyses. The project was supervised and originally designed by H.K.

Author Information The sequencing data have been deposited in the Gene Expression Omnibus (accession GSE43520) and SRA (PRJNA186438 and PRJNA202404). 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 A.N. (anamaria.necsulea@epfl.ch) or H.K. (henrik.kaessmann@unil.ch).

METHODS

RNA sequencing and initial analysis. Our main data set consists of 185 RNA-seq (135 previously published[2] and 50 new) samples, amounting to approximately 6 billion raw reads (Supplementary Table 1). The libraries were prepared using standard Illumina protocols and sequenced with Illumina GAII or HiSeq2000, as single-end reads, after poly(A) selection. After ensuring data comparability (Supplementary Discussion), we included 47 samples that we generated with a strand-specific RNA-seq protocol, for six species (human, mouse, opossum, platypus, chicken and *Xenopus*). To gain statistical power for co-expression network reconstruction, we incorporated 44 Illumina and 4 Applied Biosystems (ABI) Solid RNA-seq samples published by other groups (Supplementary Table 1 and Supplementary Discussion). We aligned the reads and detected splice junctions *de novo* using TopHat[41] v1.4.0 and Bowtie[49] v0.12.5. The genome sequences were retrieved from Ensembl[45] v62. Given the genetic similarity between chimpanzee and bonobo and the unavailability of the bonobo genome sequence when we started our project, we used the chimpanzee genome as a reference for all bonobo analyses.

lncRNA detection. To detect genes *de novo* with RNA-seq, we developed an algorithm that predicts multi-exonic transcribed loci based on transcribed island and splice junction coordinates and we used Cufflinks[42] to assemble transcripts from genomic read alignments (Supplementary Discussion). We combined multi-exonic transcripts detected with the two methods and Ensembl 62 annotations (including GENCODE lncRNAs[5]) into non-redundant data sets for each species. For human, we included approximately 8,000 lncRNAs predicted with RNA-seq[21]. To assess the evolution of sense-antisense transcripts, we repeated the detection procedure using only strand-specific samples. After the initial detection procedure, which used mainly in-house generated samples, we added to our analyses several previously published RNA-seq samples, mainly from the human ENCODE[5] and Illumina Human Body Map[4] projects, as well as several strand-specific samples that we generated at a later stage to increase coverage for the placenta, ovary and testes for several species (Supplementary Table 1). We did not repeat the entire detection procedure with these new samples, but we used the additional splice junction information to join fragmented lncRNA loci. We also discarded *de novo* detected loci which thus appeared to be unannotated UTRs, as they were joined with protein-coding genes. We determined the coding potential of genes based on the codon substitution frequency (CSF[24]) score and on the presence of sequence similarity with known proteins (SwissProt[43] database) or protein domains (Pfam-A[44] database). As *de novo* gene predictions can be incomplete or fragmented, we chose to assess the coding potential genome-wide rather than only for predicted exonic regions. We used the CSF score to define potential coding regions on a genome-wide scale, by scanning multiple species alignments (available through the UCSC Genome Browser[50]). Genes were said to be potentially noncoding if they were sufficiently distant (>2 kb away) from a CSF-predicted coding region. Several distance thresholds were tested (Supplementary Discussion). We evaluated two additional methods (reading frame conservation[51] and presence of open reading frames), but these performed less well and were not used in our final analyses (Supplementary Discussion). After estimating the coding potential independently for each species, we verified that the classifications of the members of homologous families agreed, thus further reducing the possibility of misclassifications.

Cross-species annotation projection. To reduce the inequalities in annotation depth among species, we projected the annotations across species and included the projected gene models in each species' data set. To do this, we searched for sequence similarity (blastn[52]) between the complementary DNAs of a reference species and the repeat-masked genomes of the target species. We accepted projections without rearrangements or internal repeats and with inferred intron sizes below 100 kb. To avoid redundancy, the projections were added recursively, and only if they did not overlap with already annotated genes (Supplementary Methods).

We reduced the occurrence of fragmented gene predictions (a single gene is annotated as multiple neighbour loci), using a homology-directed defragmentation procedure that takes advantage of the availability of multiple species. We searched for sequence similarity (blastn[52]) between the cDNA sequences of each species and classified as potentially 'fragmented' those neighbouring loci that could be reliably aligned with different regions of a single locus in another species (Supplementary Methods). For our final lncRNA data set, we excluded candidates that clustered with protein-coding sequences (thus reducing the possibility of misclassifying UTRs as lncRNAs) and we used 'de-fragmented' lncRNA annotations as controls for our analyses.

lncRNA filtering. We applied several filters to ensure reliability of the lncRNA data set. For species-specific lncRNAs we required: minimum exonic length 200 bp, at least 75% or 500 bp of non-overlapping exonic sequence, minimum 5 kb distance between lncRNA exons and Ensembl-annotated protein-coding gene exons, support by at least 5 non-strand-specific and 5 strand-specific reads (including splice junction reads), Ensembl gene biotypes (when available) 'lincRNA' or 'processed_transcript', no clustering (fragmentation) with protein-coding genes. For families of lncRNAs with n species, we required noncoding classification with both CSF and sequence similarity in at least $n-1$ species and with at least one of the two criteria in all species, minimum exonic length 200 bp (50 bp for projected genes) in all species, support by at least two reads in at least two species, minimum distance 5 kb to protein-coding gene exons for all species. For families that included Ensembl-annotated lncRNAs, we required the above criteria to be satisfied in at least $n-1$ out of n species. For genes that overlapped on the antisense strand with other genes, we required support with strand-specific reads. We note that the list of lncRNAs provided for each species includes projected genes for which transcription evidence could not be found in the corresponding species, if these genes belonged to homologous families in which at least two species had transcription evidence.

Reconstruction of homologous lncRNA families and lncRNA evolutionary age. We reconstructed homologous lncRNA families based on DNA sequence similarity. We searched for similarity between the cDNA sequences of each species, using blastn[52]. As in Ensembl Compara[53], we extracted reciprocal best hits for each pair of species and significant self-hits for each species and we clustered genes with single-linkage. As lncRNAs can overlap with protein-coding genes or with transposable elements, we repeated the procedure after masking these regions, with no significant change. For improved sensitivity, we searched for alignments of wider regions, including 5 kb of flanking sequences, in whole genome alignments generated with blastz and multiz[54] (available through the UCSC Genome Browser). Potential homologues were called for alignments that mapped to a single target species gene. This homology inference was used as a control for our analyses. We inferred the minimum lncRNA evolutionary age with parsimony, based on the phylogenetic distribution of the species with transcription evidence in the homologous gene families. We note that this estimate represents a strict lower boundary, since transcription may be undetectable for lowly expressed genes, in particular for the species with lower overall read coverage.

In addition, we tested whether the absence of transcription in some species can be simply attributed to differences in RNA-seq read coverage, and we provide an additional estimate of the potential evolutionary age of lncRNAs. We estimated the proportion of mapped reads assigned to a given lncRNA, separately for each species and tissue. For each lncRNA family and for each tissue, we then estimated the minimum such proportion (p_min), over all species in which the lncRNA was detected as transcribed. Given that for projected genes we often recover only a limited fraction of the original exonic length, the p_min probability was further adjusted to reflect the difference in exonic length between the species with no transcription evidence and the species in which p_min was observed (p_min was multiplied by the ratio of the two exonic lengths). We then assessed the probability of observing 0 reads out of the total n mapped reads, given a theoretical detection probability of p_min and assuming a binomial distribution, in the species for which transcription could not be detected in that tissue. If the tissue was not sampled for a given species (such as orangutan testes or non-human great ape placenta), the probability was set to 1. Finally, these probabilities were multiplied over all available tissues, to obtain a combined estimate of the likelihood that the absence of transcription in that species is simply due to differences in read coverage and/or annotated exonic length. We then re-estimated the evolutionary age of the lncRNA family, taking into account the phylogenetic distribution of the species in which transcription was either detected, or for which the absence of transcription could be attributed to read coverage and/or exonic length issues. This additional age estimate is termed the 'maximum' evolutionary age.

Selective constraint on DNA sequences. We computed average PhastCons[25] scores for exons and promoter regions, using genome-wide nucleotide resolution scores from the UCSC Genome Browser[50]. We downloaded SNP data from the 1000 Genomes Project[26], we filtered the SNPs to exclude potential CpG sites and we computed the average derived allele frequency (DAF) for the African population. For DAF comparisons, we derived 95% confidence intervals from 100 bootstrap resampling replicates (parametric statistics cannot be applied due to non-normal distributions). We analysed only autosomal SNPs, residing in regions of moderate recombination (<2 cM per Mb), as measured using the DECODE[55] sex-averaged recombination maps in 20 kb windows centred on the SNP. As a neutral control, we resampled intergenic SNPs (>5 kb away from coding or noncoding genes) found in regions of similar recombination rates as lncRNAs (Supplementary Discussion). For overlapping genes (for example, sense–antisense transcripts), both measures of selective constraint were estimated using non-overlapping exonic regions.

Expression-level estimation and normalization. We estimated RPKM values from unambiguous read alignments obtained with TopHat[41]. To ensure an unbiased measurement, we considered only exonic regions that could be unambiguously assigned to a single gene. We also measured expression levels with Cufflinks v2.0.0, using all mapped reads, with the embedded multi-read and fragment bias correction methods (Supplementary Discussion). For projected genes, for which exon annotations are often incomplete, we included 1-kb flanking sequences on each side in the expression computation, if this extended region did not overlap with

other transcribed loci. We normalized expression levels among samples with a median scaling, using the 1000 least-varying genes as a reference, as described previously[2].

Transcription-factor-binding analysis. We used a genome-wide set of human transcription-factor-binding sites (~2.7 million sites, for 375 transcription factors), predicted *in silico*[30], as well as ChIP-seq peaks for 127 transcription factors (excluding those directly associated with PolII or PolIII) from the human ENCODE project[56]. We analysed the occurrence of transcription-factor-binding sites or peaks in promoter regions, exclusively for genes for which the predicted transcription start site was found within 100 bp of a CAGE tag cluster (data from the FANTOM project[57]). Two promoter region sizes were tested (2 kb and 5 kb), reaching similar conclusions. We also used ChIP-seq data for HNF4A and CEBPA for human and mouse[32]. We aligned promoter regions for the two species and considered that transcription-factor binding was conserved if peaks were found in both species within 10 kb of the aligned transcription start site. As a control, we analysed transcription-factor binding and binding conservation for 20,000 randomly drawn intergenic regions.

Expression evolution analyses. For the qualitative assessment of transcription conservation, we analysed 4,430 intergenic lncRNAs (>5 kb away from protein-coding genes) that had 1–1 orthologues in all primate species and which had at least 2 mapped reads in human in a pool of brain, cerebellum, heart, kidney and liver samples, as well as 2,492 human lncRNAs that overlapped on the antisense strand with exons of protein-coding genes, which had orthologues in at least one of the other species with strand-specific data (mouse, opossum, platypus, chicken and *Xenopus*). These antisense lncRNAs were further filtered to extract genes that were expressed in human brain and testes. We evaluated Spearman's correlation coefficients between pairs of samples, on lncRNA or protein-coding gene RPKM values. All available 1–1 orthologues were used. As a control for our expression evolution analyses, we resampled the same average number of reads per gene for each species and tissue, keeping the proportions among genes identical to the original distribution.

Tissue-specific expression. We evaluated the tissue specificity of the expression pattern with a previously proposed index[46], which varies between 0 for housekeeping genes and 1 for tissue-restricted genes:

$$\frac{\sum_{i=1}^{n}\left(1-\frac{\exp_i}{\exp_{\max}}\right)}{n-1}$$

where n is the number of tissues, $\exp_i$ is the expression value in tissue i, and $\exp_{\max}$ the maximum expression level over all tissues. We used RPKM and $\log_2$-transformed RPKM for expression values, reaching the same conclusions. The randomly expected proportion of conserved specificity across species was computed as the product of the observed proportions of tissue-specific genes in each species, for each tissue.

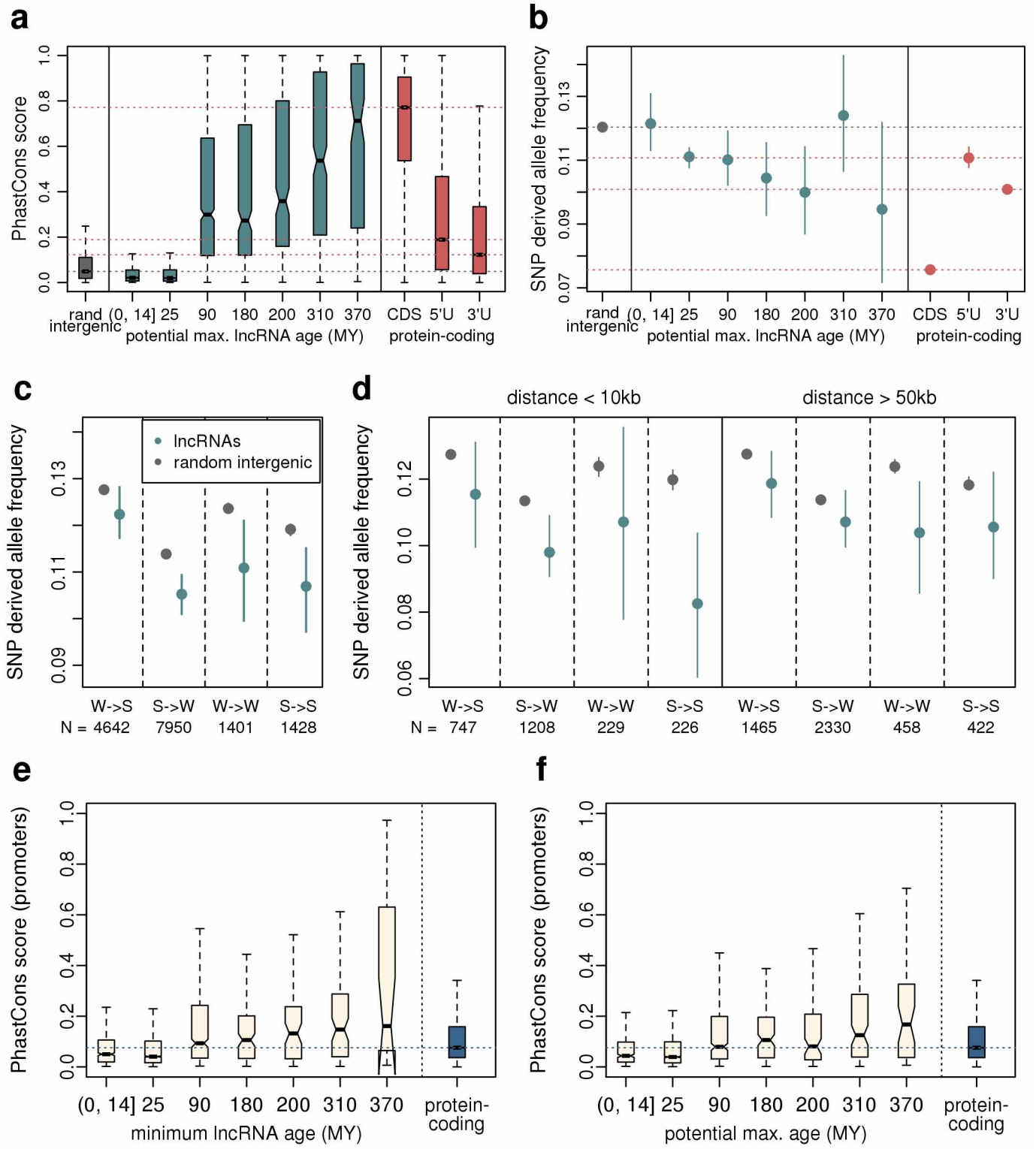
Reconstruction and analysis of the co-expression network. We reconstructed the evolutionarily conserved co-expression network for lncRNAs and protein-coding genes following a previously proposed method[35] (Supplementary Discussion). For each species and for each pair of genes (lncRNA or protein-coding), we computed the Pearson correlation coefficients of expression patterns. Given two homologous families, we examined whether the combination of correlation coefficients measured in each species was significantly higher or lower than expected by chance. The statistical tests were carried out by comparing the observed ranks of the correlation coefficients with a random n-dimensional order statistics[35]. We computed correlations only for genes expressed in at least three samples for each species, and we computed P values only if correlations were evaluated in at least three species. We allow negative connections, which have lower than expected rank combinations. We considered only lncRNAs estimated to have originated in the Eutherian ancestor or earlier, but without requiring representatives in all descendant species. As P-value computations are highly time-consuming with a large number of species, analyses were carried out using a representative subset of seven species: human, macaque, mouse, opossum, platypus, chicken and *Xenopus*. For greater accuracy of the reconstruction we extended our in-house generated data set to include previously published, comparable RNA-seq samples (Supplementary Table 1). We visualized the network with Cytoscape[47] and we detected clusters of highly inter-connected genes with the Markov Cluster (MCL) algorithm[39].

Defining potential miRNA precursors. To search for lncRNAs that may promote transcription of miRNAs or are potentially processed into miRNAs, we extracted all miRNA hairpin sequences from miRBase[58] 18 and searched for sequence similarity (blastn[52]) against all annotated gene regions, including 10 kb of flanking sequences. Genes with at least one miRNA hairpin alignment (95% identity, aligned on the entire length) on the same strand were considered potential miRNA precursors.

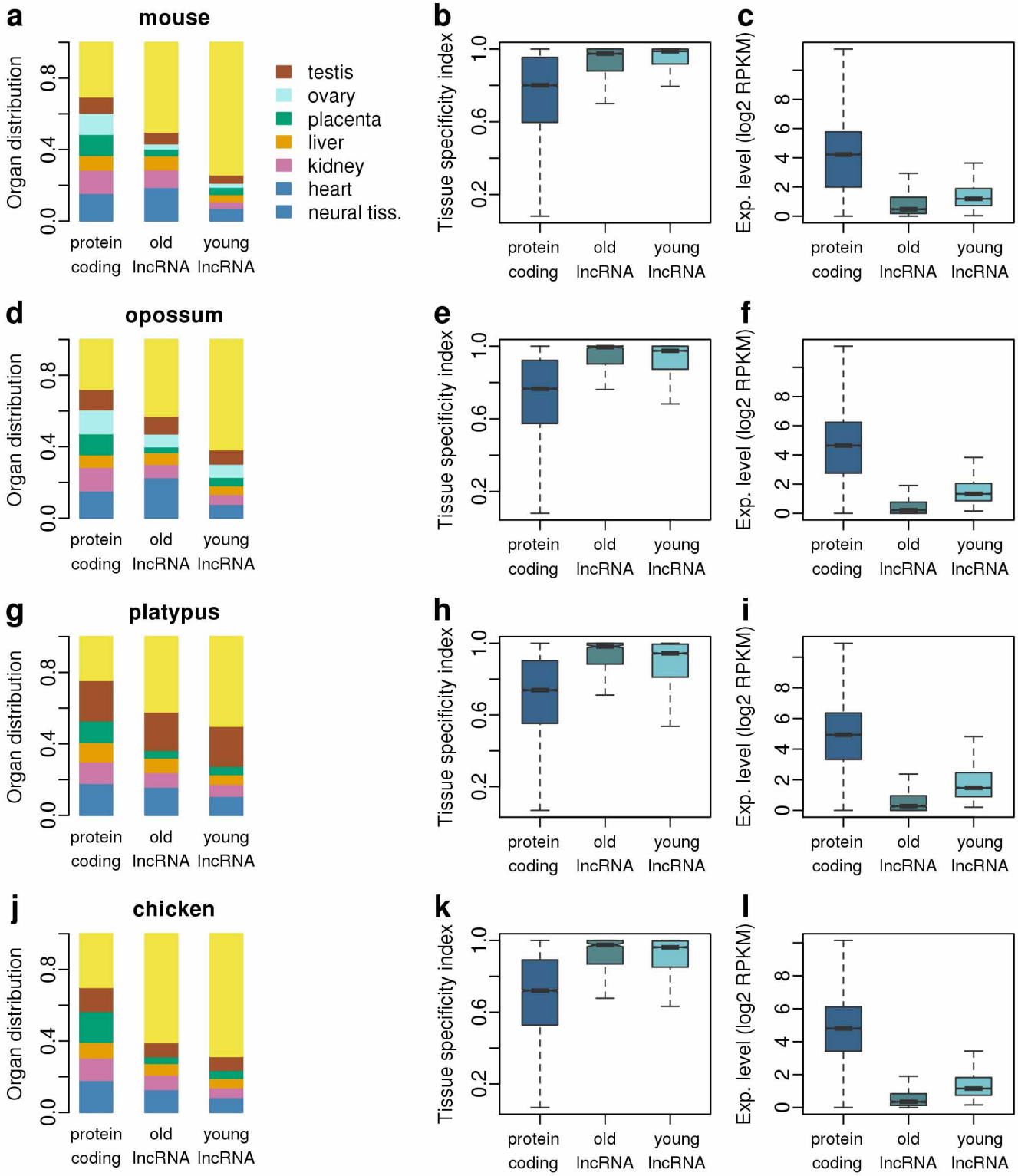
Statistical analyses. All statistical analyses and graphical representations (including gene expression clustering, principal component analysis, randomization tests for statistical significance) were done in R[48]. For statistical tests involving the co-expression network, we generated a set of 100 randomized networks by permuting the gene identifiers of the nodes for each edge. The randomized networks had the same distribution of edges types (positive, negative, coding–coding, coding–noncoding), and the node degree was preserved. To test the significance of the network properties (for example, *cis* connections), we derived a P value by comparing the values observed in real and randomized networks. To compare the degrees of connectivity among gene types by controlling for unequal expression levels, we extracted lncRNAs with maximum expression levels ($\log_2$ RPKM) between 3 and 6, and divided them into 6 discrete expression classes ([3, 3.5], (3.5, 4], … , (5.5,6] $\log_2$ RPKM) (round brackets represent open (excluded) boundaries of intervals, square brackets represent closed (included) boundaries). We then drew transcription-factor and non-transcription-factor protein-coding genes matching the relative proportions of lncRNAs in each expression class. The resampling was repeated 100 times.

Data availability. The sequencing data have been submitted to GEO (accession GSE43520) and SRA (PRJNA186438 and PRJNA202404). The lncRNA annotations and homologous families have been made available on the publisher's website (Supplementary Data 1 and 2), as well as gene expression levels for lncRNAs and protein-coding genes (Supplementary Data 3) and miRNAs (Supplementary Data 4).

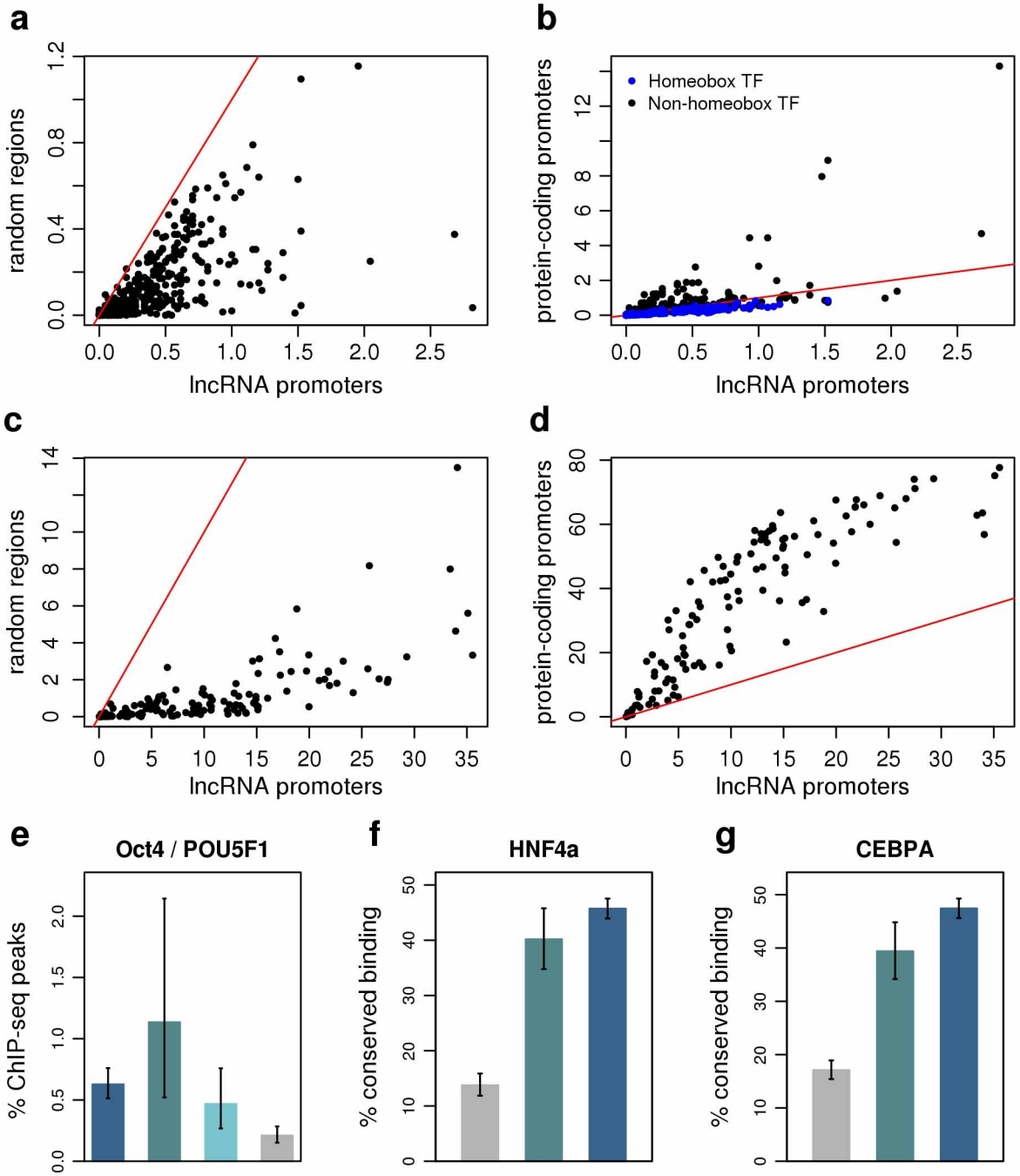
49. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
50. Rhead, B. *et al.* The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.* **38,** D613–D619 (2010).
51. Kellis, M., Patterson, N., Birren, B., Berger, B. & Lander, E. S. Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery. *J. Comput. Biol.* **11,** 319–355 (2004).
52. Altschul, S. F., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
53. Vilella, A. J. *et al.* EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19,** 327–335 (2009).
54. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14,** 708–715 (2004).
55. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467,** 1099–1103 (2010).
56. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489,** 91–100 (2012).
57. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet.* **38,** 626–635 (2006).
58. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39,** D152–D157 (2011).



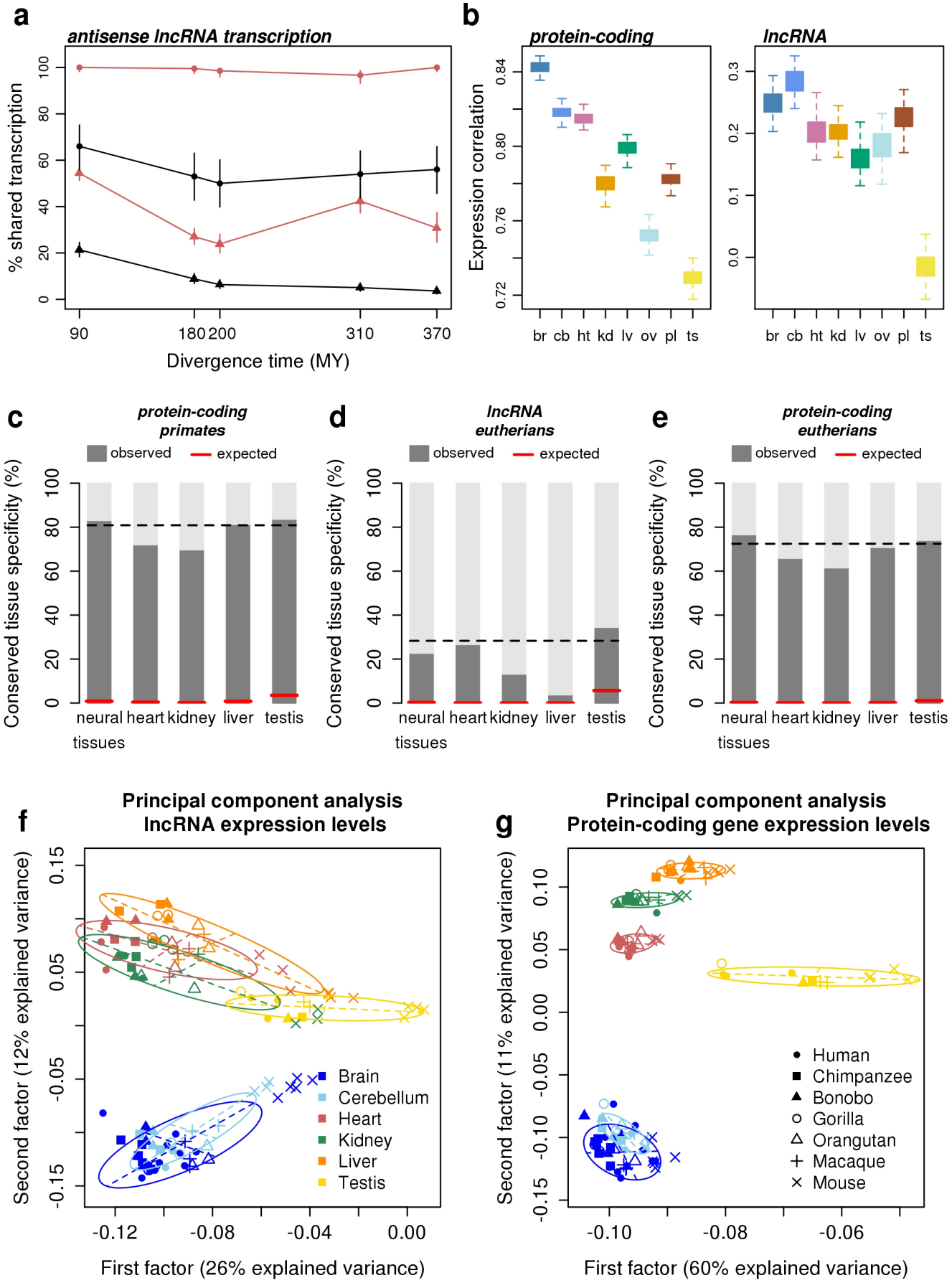
Extended Data Figure 1 | lncRNA evolutionary age and sequence conservation patterns. **a**, Exonic sequence conservation (mean placental PhastCons score), for random intergenic regions, lncRNA maximum evolutionary age classes, coding and untranslated exons of protein-coding genes. **b**, Mean DAF of autosomal non-CpG SNPs segregating in African populations (1000 Genomes project[26]). Intergenic SNPs were randomly drawn in regions matching lncRNA recombination rates (Methods). **c**, Mean DAF for the four classes of mutation orientation (W to S (W→S) or AT to GC; S to W (S→W) or GC to AT; W to W (W→W), or AT to AT; and S to S (S→S), or GC to GC) for autosomal non-CpG SNPs found in primate-specific (age 25 Myr) lncRNA exonic regions (blue) or in intergenic regions with matching recombination rates (grey). The W→S and S→W mutation classes are known to be affected by GC-biased gene conversion. **d**, Same as **c** but for lncRNAs that are found close to (left panel, maximum distance 10 kb) or far from (right panel, minimum distance 50 kb) Ensembl-annotated coding or noncoding genes. **e**, Mean placental PhastCons score for promoter regions (1 kb upstream) of lncRNA minimum evolutionary age classes (beige) and protein-coding genes (blue). **f**, Mean placental PhastCons score for promoter regions (1 kb upstream) of lncRNA maximum evolutionary age classes (beige) and protein-coding genes (blue). Error bars, 95% confidence intervals based on 100 bootstrap resampling replicates.



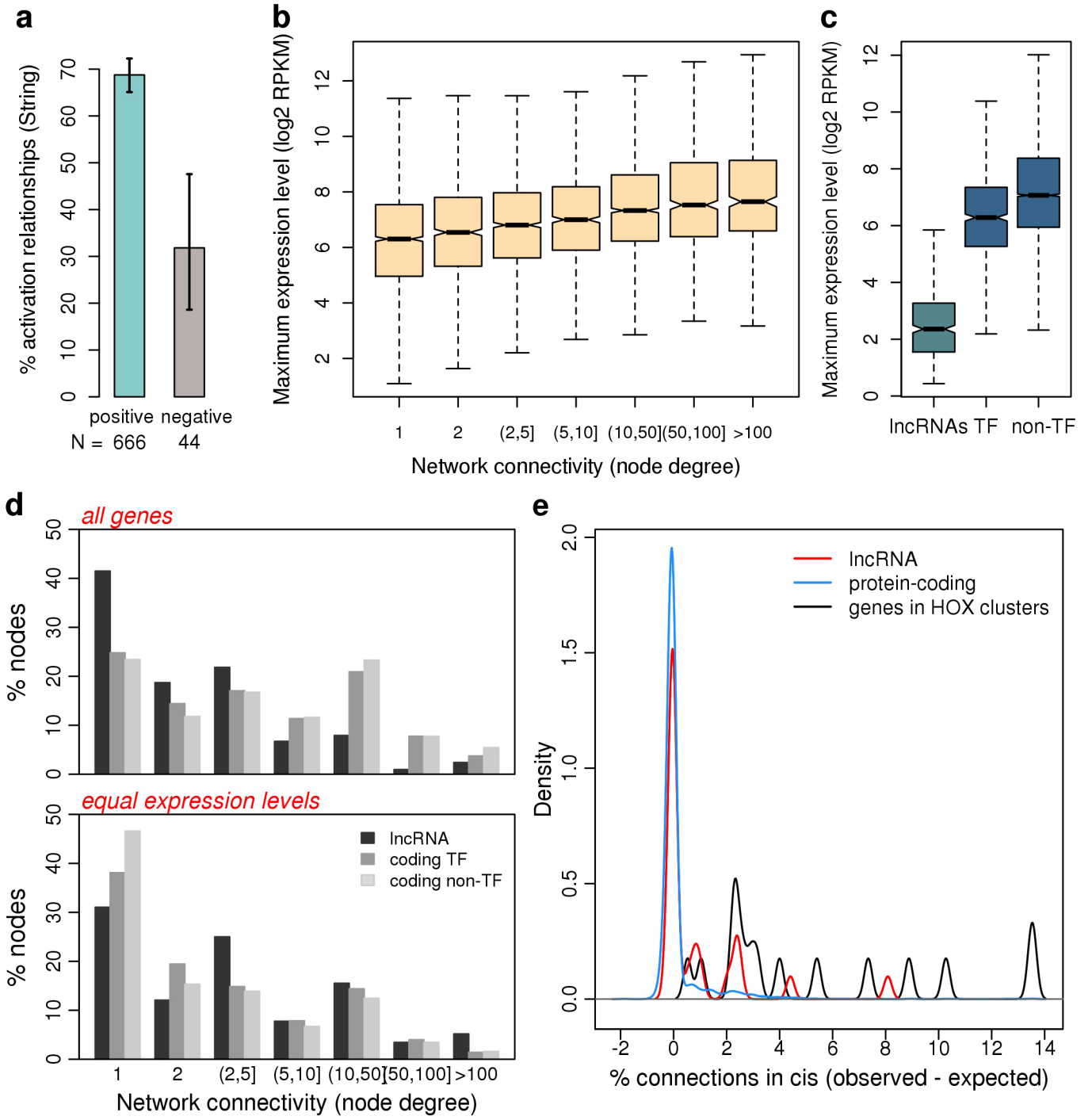
Extended Data Figure 2 | lncRNA expression patterns in four tetrapod species. **a**, Proportions of genes with observed maximum expression in different organs for mouse protein-coding genes, old lncRNAs (shared across at least two species) and young lncRNAs (species-specific). **b**, Tissue-specificity index, for the same classes of mouse genes. Values close to 1 represent high tissue specificity. **c**, Distribution of the maximum expression level ($\log_2$-transformed RPKM). **d**–**f**, Same as **a**–**c** but for the opossum. **g**–**i**, Same as **a**–**c** but for the platypus. **j**–**l**, Same as **a**–**c** but for the chicken.



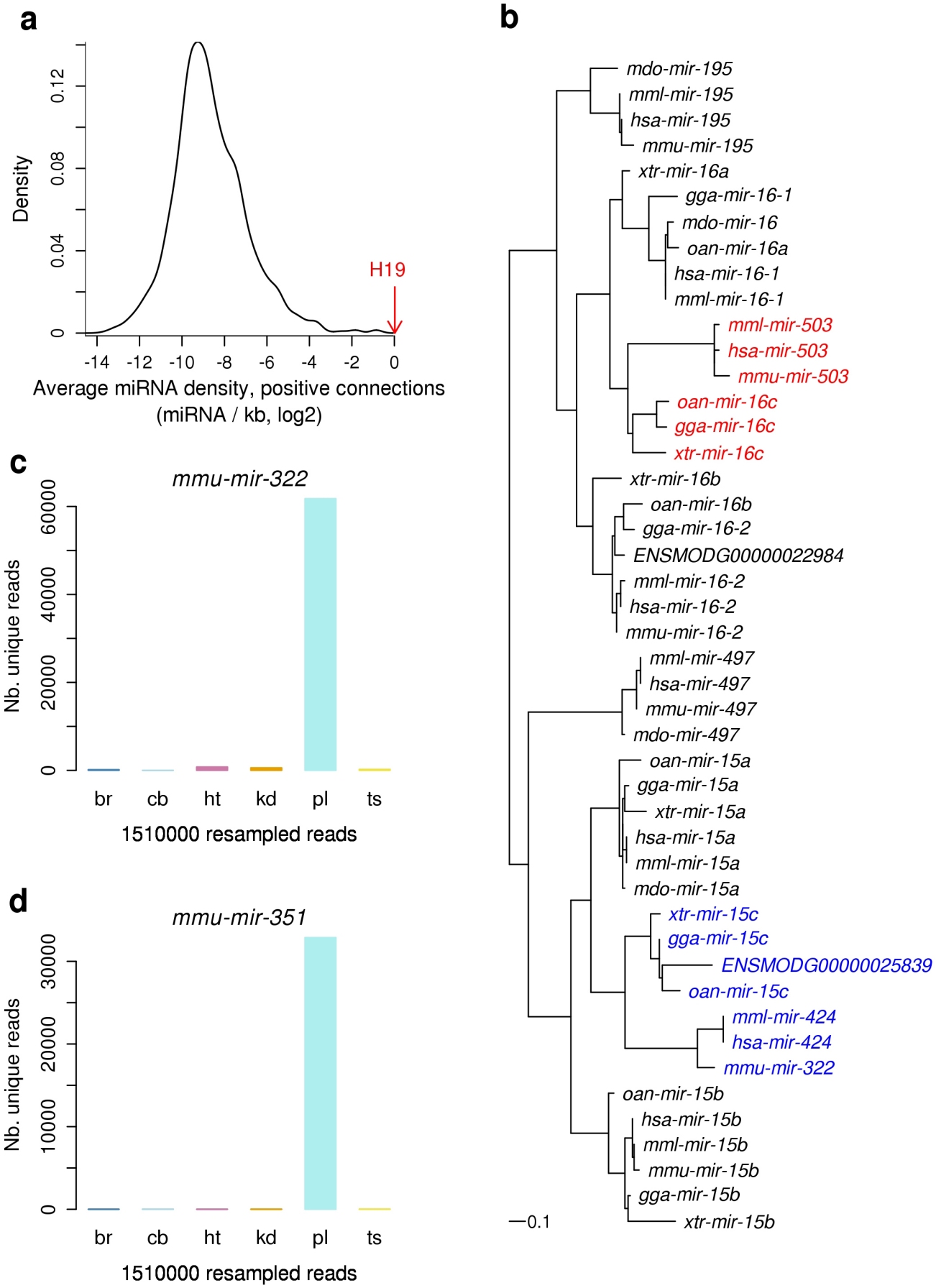
Extended Data Figure 3 | Transcription-factor binding at lncRNA promoters. **a**, Comparison between the frequencies of *in silico*-predicted transcription-factor (TF)-binding sites in lncRNA promoters (2 kb upstream) and in random intergenic regions. **b**, Comparison between the frequencies of *in silico*-predicted TF-binding sites in lncRNA and protein-coding gene promoters (2 kb upstream). Homeobox TFs are shown in blue. **c**, Comparison between the frequencies of experimentally determined (ChIP-seq ENCODE) TF-binding sites in lncRNA promoters (2 kb upstream) and in random intergenic regions. **d**, Comparison between the frequencies of experimentally determined (ChIP-seq ENCODE) predicted TF-binding sites in lncRNA and protein-coding gene promoters (2 kb upstream). **e**, Frequency of binding (Encode ChIP-seq data) for OCT4 (also known as POU5F1). **f**, **g**, Proportion of HNF4A- CEBPA-binding events shared between human and mouse, for random intergenic regions, lncRNA (321 lncRNAs with binding events and liver expression, supported by CAGE data) and protein-coding gene promoters (5 kb upstream).



Extended Data Figure 4 | Evolution of lncRNA expression patterns. **a**, Percentage of human lncRNAs (found in antisense of protein-coding genes) that have transcription evidence in other species, as a function of the divergence time. Transcription evidence was assessed in a pool of brain and testes strand-specific RNA-seq data, for 2,535 human antisense lncRNAs that had 1–1 orthologues in at least one other species and transcription evidence in human (Methods). **b**, Spearman correlation of human and mouse expression levels, in different tissues. The boxplots represent the variation observed in 100 bootstrap replicates. **c**, Proportion of human organ-specific protein-coding genes (tissue-specificity index >0.9, RPKM >0.1) for which the organ specificity is shared across primates. Red lines, random expectation of shared organ specificity; horizontal black line, average conserved specificity for all organs. **d**, Proportion of human organ-specific lncRNAs (minimum evolutionary age >90 Myr, tissue-specificity index >0.9, RPKM >0.1) for which the organ specificity is shared across eutherians. Red lines, random expectation of shared organ specificity; horizontal black line, average conserved specificity for all organs. **e**, Same as **c**, conservation across eutherian species. **f**, Principal component analysis of lncRNA expression levels for families of eutherian 1–1 orthologues. **g**, Principal component analysis of protein-coding gene expression levels for families of eutherian 1–1 orthologues.



Extended Data Figure 5 | Characteristics of the evolutionarily conserved co-expression network. **a**, Proportion of activation/inhibition relationships annotated in the String database, for positive and negative co-expression network connections. **b**, Gene expression levels (maximum over all available sample and species for each co-expression network node) for different network connectivity classes. **c**, Gene expression levels (maximum over all available sample and species for each co-expression network node) for connected lncRNAs, transcription factors (TFs) and non-TF protein-coding genes. **d**, Network connectivity (node degree) for lncRNAs (black), transcription factors (medium grey) and for non-transcription factors protein-coding genes (light grey). Top, raw data; bottom, after correcting for expression level differences. **e**, Difference between observed and expected proportions of connections in *cis*, for lncRNAs (red), protein-coding genes (blue) and for genes found in *HOX* clusters (black). The expected proportions were computed through randomizations (Methods).



Extended Data Figure 6 | Expression patterns and sequence evolution of *H19X*-associated miRNAs. **a**, Distribution of the average embedded miRNA density (miRNA hairpins per kb, in the gene body or 10 kb downstream), for genes that are positively connected with each network node. Red arrow, average miRNA density for genes that are positively connected with *H19*. **b**, Maximum likelihood reconstruction of the phylogeny of the ancient *H19X*-associated miRNA family (representative members *miR-503*, *miR-322*, *miR-424*, *miR-15c*, *miR-16c*). miRNAs associated with *H19X* are displayed in red (subfamily containing *miR-503* and *miR-16c*) and blue (subfamily containing *miR-424*, *miR-322* and *miR-15c*). miRNA names are derived from miRBase where available, including three-letter species abbreviations. Hsa, *Homo sapiens*; Mdo, *Monodelphis domestica* (opossum); Mml, *Macaca mulatta* (macaque); Mmu, *Mus musculus* (mouse); Oan, *Ornithorhynchus anatinus* (platypus); Gga, *Gallus gallus* (chicken), Xtr, *Xenopus tropicalis*. Ensembl identifiers are given for two opossum miRNAs. **c**, Expression pattern of the mouse miRNA *mmu-miR-322*, associated with *H19X*. The expression level was computed as the number of uniquely mapping reads per miRNA, after resampling the same number of reads per tissue. **d**, Same as **c** but for the mouse miRNA *mmu-miR-351*.

Extended Data Table 1 | Validation of the *de novo* detection and classification methods

(a)

species	protein-coding		lincRNA		processed transcript	
	partial	complete	partial	complete	partial	complete
human	17006 (88%)	9510 (49%)	942 (77%)	442 (36%)	4708 (54%)	2331 (27%)
chimp / bonobo	15457 (93%)	10292 (62%)	-	-	NA	NA
gorilla	14623 (92%)	9693 (61%)	-	-	NA	NA
orangutan	13222 (87%)	8676 (57%)	-	-	NA	NA
macaque	14617 (93%)	9235 (59%)	-	-	NA	NA
mouse	16824 (90%)	12072 (64%)	1000 (78%)	647 (51%)	1568 (64%)	1021 (41%)
opossum	12204 (95%)	7936 (62%)	-	-	NA	NA
platypus	9221 (90%)	3782 (37%)	-	-	NA	NA
chicken	13611 (89%)	9598 (63%)	-	-	NA	NA
Xenopus	13373 (89%)	7052 (47%)	-	-	NA	NA
average	90 %	57 %	78%	44%	56%	30%

(b)

species	protein-coding	lincRNA	processed transcript	tRNA, rRNA
human	19247 (92%)	721 (58%)	6710 (78%)	511 (96%)
chimp / bonobo	17537 (99%)	-	-	505 (97%)
gorilla	16519 (99%)	-	-	508 (97%)
orangutan	16514 (99%)	-	-	516 (97%)
macaque	16807 (100%)	-	-	696 (97%)
mouse	20651 (95%)	1043 (81%)	2062 (84%)	306 (96%)
opossum	13607 (99%)	-	-	170 (97%)
platypus	10521 (99%)	-	-	203 (99%)
chicken	14105 (86%)	-	-	201 (96%)
Xenopus	15492 (96%)	-	-	268 (99%)
average	96 %	70 %	79 %	97 %

a, Proportion of Ensembl-annotated (release 62) multi-exonic protein-coding genes, lncRNAs and processed transcripts recovered with our *de novo* detection methods. Partial overlap: number (percentage) of Ensembl-annotated multi-exonic genes for which at least half of the exons were recovered *de novo*. Complete: number (percentage) of multi-exonic genes for which all exons were recovered *de novo*. Protein-coding genes were filtered to retain those with 'known' or 'known by projection' gene status. **b**, Proportion of Ensembl-annotated protein-coding genes, lncRNAs, processed transcripts and other noncoding RNA genes (transfer RNA (tRNA), ribosomal RNA (rRNA)) that were correctly classified as coding or noncoding with our approach.

Extended Data Table 2 | LncRNA repertoires in 11 tetrapod species

(a)

species	total	orphan	in 1-1 fam.	intergenic	intragenic	de novo	known	projected
Hsa	14682	481	14201 (92%)	12286	2396	2030	4619 (3263)	8032
Ptr/Ppa	14654	347	14307 (90%)	12695	1959	3450	0 (0)	11203
Ggo	14258	501	13757 (91%)	12546	1712	4530	0 (0)	9726
Ppy	13756	229	13527 (61%)	12189	1566	1099	0 (0)	12655
Mml	15280	1060	14220 (88%)	13463	1817	5931	0 (0)	9348
Mmu	10850	7895	2955 (79%)	9045	1805	7485	1580 (1580)	1784
Mdo	8039	6579	1460 (56%)	7171	868	6815	0 (0)	1223
Oan	6889	5890	999 (59%)	6576	313	6097	0 (0)	790
Gga	5412	4730	682 (71%)	4951	461	4857	0 (0)	554
Xtr	3296	3059	237 (49%)	3133	163	3091	0 (0)	204

(b)

species	total	orphan	in 1-1 fam.	intergenic	intragenic	de novo	known	projected
Hsa	12677	8080	4597 (61%)	6823	5854	4161	5646 (4485)	2869
Mmu	15934	11604	4330 (70%)	9138	6796	11448	2482 (2482)	2003
Mdo	9635	6898	2737 (42%)	5847	3788	7416	0 (0)	2218
Oan	8037	5941	2096 (42%)	5704	2333	6346	0 (0)	1689
Gga	8358	6733	1625 (47%)	5095	3263	7099	0 (0)	1258
Xtr	5314	4557	757 (31%)	3748	1566	4683	0 (0)	630

a, lncRNA repertoires determined using all RNA-seq samples available for each species, including both strand-specific and non-strand-specific data. Gga, *Gallus gallus* (chicken); Ggo, *Gorilla gorilla*; Hsa, *Homo sapiens*; Mdo, *Monodelphis domestica* (opossum); Mml, *Macaca mulatta* (macaque); Mmu, *Mus musculus* (mouse); Oan: *Ornithorhynchus anatinus* (platypus); Ppa, *Pan paniscus* (bonobo); Ppy, *Pongo pygmaeus* (orangutan); Ptr, *Pan troglodytes* (chimpanzee); Xtr: *Xenopus tropicalis*. Orphan, lncRNAs for which no orthologues could be detected; 1–1 fam, lncRNAs found in 1–1 orthologous families; Intergenic, lncRNAs found >5 kb away from Ensembl-annotated protein-coding genes; Intragenic, lncRNAs that overlap with Ensembl-annotated protein-coding genes on the opposite strand, but are found at least 5 kb away from their exons; *De novo*, previously unknown lncRNAs detected with RNA-seq; Known, lncRNAs that confirm previously known loci (including GENCODE/Ensembl human and mouse annotations (numbers in parentheses) and a set of 8,264 human lncRNAs previously detected with RNA-Seq[4]). Projected, lncRNAs derived from cross-species annotation projections. **b**, lncRNA repertoires determined with strand-specific data.

Extended Data Table 3 | LncRNA evolutionary age estimates and synteny conservation

(a)

	hominins	african apes	great apes	primates	eutherians	therians	mammals	amniotes	tetrapods
hominins	47.1%	1%	30.8%	20.7%	0.5%	0%	0%	0%	0%
african apes	0%	32.5%	40.4%	25.4%	0.4%	0.8%	0%	0.2%	0.4%
great apes	0%	0%	80.8%	16.6%	1.2%	0.8%	0.2%	0.2%	0.2%
primates	0%	0%	0%	97.1%	1.3%	0.7%	0.4%	0.3%	0.1%
eutherians	0%	0%	0%	0%	88.6%	5.7%	2.9%	2%	0.9%
therians	0%	0%	0%	0%	0%	81.2%	12.2%	3.4%	3.2%
mammals	0%	0%	0%	0%	0%	0%	92.5%	4.4%	3.1%
amniotes	0%	0%	0%	0%	0%	0%	0%	91.3%	8.7%
tetrapods	0%	0%	0%	0%	0%	0%	0%	0%	100%

(b)

	Hsa	Ptr / Ppa	Ggo	Ppy	Mml	Mmu	Mdo	Oan	Gga	Xtr
Hsa		94.7%	93.6%	89.4%	90.8%	90.9%	67.2%	51.1%	90.1%	79.4%
Ptr / Ppa	97.5%		94.2%	90.5%	92.1%	91.6%	68.1%	55.6%	92.2%	81.5%
Ggo	95.9%	94.9%		89.8%	90.2%	91.2%	68.6%	51.3%	89.4%	82.8%
Ppy	95.4%	94.2%	92.7%		91%	92.3%	69.2%	53.5%	89.5%	87%
Mml	94.4%	93.4%	91.7%	88.2%		89.6%	67.4%	48.8%	90.9%	78.1%
Mmu	86%	83.6%	81.3%	77.9%	79.8%		60.2%	50.4%	87.2%	82.6%
Mdo	88.5%	87.9%	87%	83.6%	84.2%	89.8%		55.2%	90.1%	82%
Oan	59.6%	58.8%	55.9%	54%	55.8%	60.6%	44.7%		84%	75.5%
Gga	54.8%	51.2%	47.6%	47.5%	50%	58.1%	40.9%	48.2%		72.7%
Xtr	61.3%	59.5%	55.7%	51.8%	56.8%	61.1%	46.2%	46.6%	79.4%	

a, Comparison between the minimum evolutionary age of lncRNA families (requiring transcription evidence in all species), and the maximum potential evolutionary age (Methods). The numbers represent the percentage of cases in which a given 'minimum age' estimate (rows) is associated with a given 'maximum age' estimate (columns). **b**, Synteny conservation for pairs of neighbouring genes that contain at least one lncRNA. The neighbouring gene pairs in the reference species (see Extended Data Table 2 legend) were genes with 1–1 orthologues in the target species, separated by 5–100 kb in the reference genome. The numbers represent the percentage of neighbouring gene pairs in the reference species (rows) for which the 1–1 orthologues in the target species (columns) were found on the same chromosome, separated by at most 100 kb.

ARTICLE

doi:10.1038/nature12968

Stimulus-triggered fate conversion of somatic cells into pluripotency

Haruko Obokata[1,2,3], Teruhiko Wakayama[3]†, Yoshiki Sasai[4], Koji Kojima[1], Martin P. Vacanti[1,5], Hitoshi Niwa[6], Masayuki Yamato[7] & Charles A. Vacanti[1]

Here we report a unique cellular reprogramming phenomenon, called stimulus-triggered acquisition of pluripotency (STAP), which requires neither nuclear transfer nor the introduction of transcription factors. In STAP, strong external stimuli such as a transient low-pH stressor reprogrammed mammalian somatic cells, resulting in the generation of pluripotent cells. Through real-time imaging of STAP cells derived from purified lymphocytes, as well as gene rearrangement analysis, we found that committed somatic cells give rise to STAP cells by reprogramming rather than selection. STAP cells showed a substantial decrease in DNA methylation in the regulatory regions of pluripotency marker genes. Blastocyst injection showed that STAP cells efficiently contribute to chimaeric embryos and to offspring via germline transmission. We also demonstrate the derivation of robustly expandable pluripotent cell lines from STAP cells. Thus, our findings indicate that epigenetic fate determination of mammalian cells can be markedly converted in a context-dependent manner by strong environmental cues.

In the canalization view of Waddington's epigenetic landscape, fates of somatic cells are progressively determined as cellular differentiation proceeds, like going downhill. It is generally believed that reversal of differentiated status requires artificial physical or genetic manipulation of nuclear function such as nuclear transfer[1,2] or the introduction of multiple transcription factors[3]. Here we investigated the question of whether somatic cells can undergo nuclear reprogramming simply in response to external triggers without direct nuclear manipulation. This type of situation is known to occur in plants—drastic environmental changes can convert mature somatic cells (for example, dissociated carrot cells) into immature blastema cells, from which a whole plant structure, including stalks and roots, develops in the presence of auxins[4]. A challenging question is whether animal somatic cells have a similar potential that emerges under special conditions. Over the past decade, the presence of pluripotent cells (or closely relevant cell types) in adult tissues has been a matter of debate, for which conflicting conclusions have been reported by various groups[5–11]. However, no study so far has proven that such pluripotent cells can arise from differentiated somatic cells.

Haematopoietic cells positive for CD45 (leukocyte common antigen) are typical lineage-committed somatic cells that never express pluripotency-related markers such as Oct4 unless they are reprogrammed[12,13]. We therefore addressed the question of whether splenic CD45$^+$ cells could acquire pluripotency by drastic changes in their external environment such as those caused by simple chemical perturbations.

Low pH triggers fate conversion in somatic cells

CD45$^+$ cells were sorted by fluorescence-activated cell sorting (FACS) from the lymphocyte fraction of postnatal spleens (1-week old) of C57BL/6 mice carrying an *Oct4-gfp* transgene[14], and were exposed to various types of strong, transient, physical and chemical stimuli (described below). We examined these cells for activation of the *Oct4* promoter after culture for several days in suspension using DMEM/F12 medium supplemented with leukaemia inhibitory factor (LIF) and B27 (hereafter called LIF+B27 medium). Among the various perturbations, we were particularly interested in low-pH perturbations for two reasons. First, as shown below, low-pH treatment turned out to be most effective for the induction of *Oct4*. Second, classical experimental embryology has shown that a transient low-pH treatment under 'sublethal' conditions can alter the differentiation status of tissues. Spontaneous neural conversion from salamander animal caps by soaking the tissues in citrate-based acidic medium below pH 6.0 has been demonstrated previously[15–17].

Without exposure to the stimuli, none of the cells sorted with CD45 expressed *Oct4*-GFP regardless of the culture period in LIF+B27 medium. In contrast, a 30-min treatment with low-pH medium (25-min incubation followed by 5-min centrifugation; Fig. 1a; the most effective range was pH 5.4–5.8; Extended Data Fig. 1a) caused the emergence of substantial numbers of spherical clusters that expressed *Oct4*-GFP in day-7 culture (Fig. 1b). Substantial numbers of GFP$^+$ cells appeared in all cases performed with neonatal splenic cells ($n = 30$ experiments). The emergence of *Oct4*-GFP$^+$ cells at the expense of CD45$^+$ cells was also observed by flow cytometry (Fig. 1c, top, and Extended Data Fig. 1b, c). We next fractionated CD45$^+$ cells into populations positive and negative for CD90 (T cells), CD19 (B cells) and CD34 (haematopoietic progenitors[18]), and subjected them to low-pH treatment. Cells of these fractions, including T and B cells, generated *Oct4*-GFP$^+$ cells at an efficacy comparable to unfractionated CD45$^+$ cells (25–50% of surviving cells on day 7), except for CD34$^+$ haematopoietic progenitors[19], which rarely produced *Oct4*-GFP$^+$ cells (<2%; Extended Data Fig. 1d).

Among maintenance media for pluripotent cells[20], the appearance of *Oct4*-GFP$^+$ cells was most efficient in LIF+B27 medium, and did not occur in mouse epiblast-derived stem-cell (EpiSC) medium[21,22] (Extended Data Fig. 1e). The presence or absence of LIF during days 0–2 did not substantially affect the frequency of *Oct4*-GFP$^+$ cell generation on day 7 (Extended Data Fig. 1f), whereas the addition of LIF during days 4–7 was not sufficient, indicating that LIF dependency started during days 2–4.

[1]Laboratory for Tissue Engineering and Regenerative Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Laboratory for Cellular Reprogramming, RIKEN Center for Developmental biology, Kobe 650-0047, Japan. [3]Laboratory for Genomic Reprogramming, RIKEN Center for Developmental biology, Kobe 650-0047, Japan. [4]Laboratory for Organogenesis and Neurogenesis, RIKEN Center for Developmental biology, Kobe 650-0047, Japan. [5]Department of Pathology, Irwin Army Community Hospital, Fort Riley, Kansas 66442, USA. [6]Laboratory for Pluripotent Stem Cell Studies, RIKEN Center for Developmental biology, Kobe 650-0047, Japan. [7]Institute of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo 162-8666, Japan. †Present address: Faculty of Life and Environmental Sciences, University of Yamanashi, Yamanashi 400-8510, Japan.

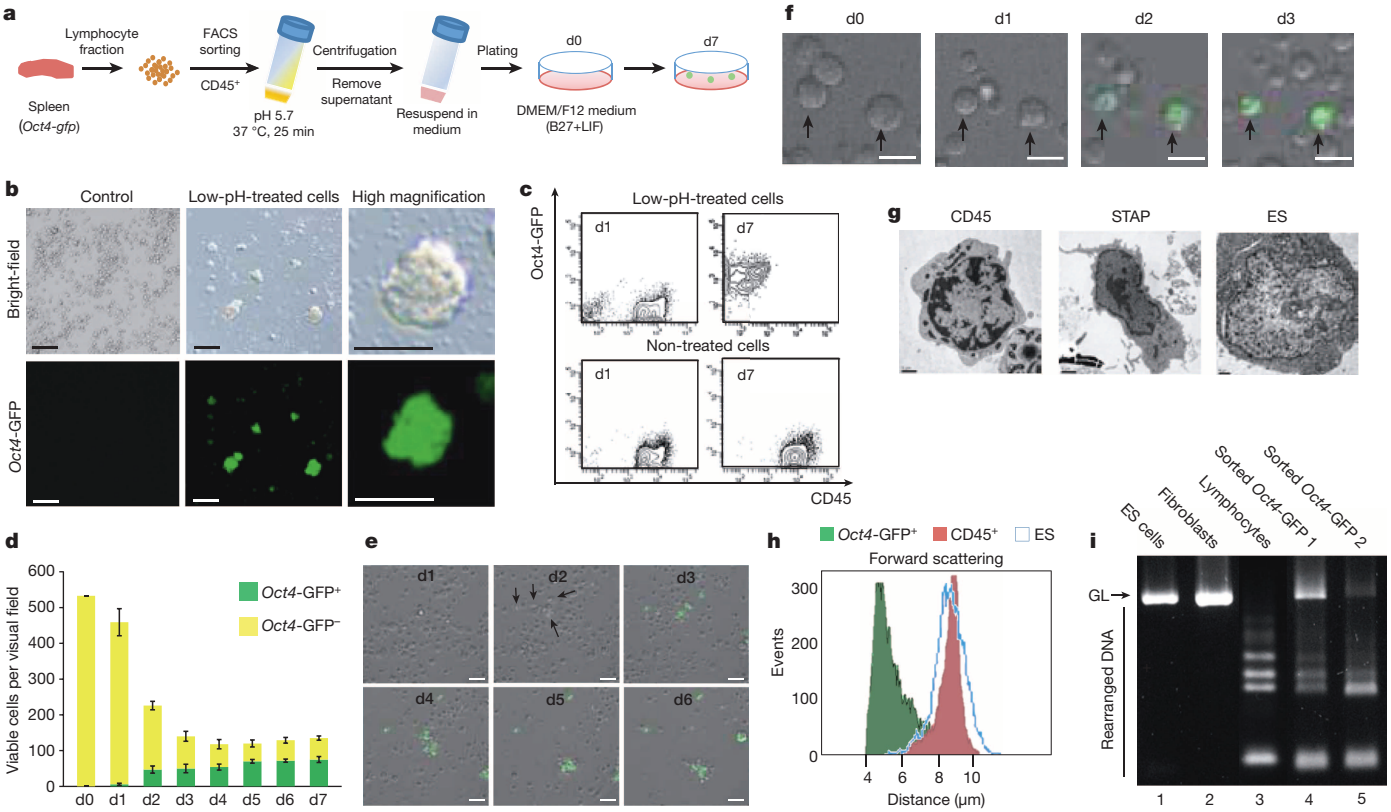


Figure 1 | Stimulus-triggered conversion of lymphocytes into *Oct4*-GFP$^+$ cells. **a**, Schematic of low-pH treatment. **b**, *Oct4*-GFP$^+$ cell clusters appeared in culture of low-pH-treated CD45$^+$ cells (middle; high magnification, right) on day 7 (d7) but not in culture of control CD45$^+$ cells (left). Top: bright-field view; bottom, GFP signals. Scale bar, 100 μm. **c**, FACS analysis. The *x* axis shows CD45 epifluorescence level; *y* axis shows *Oct4*-GFP level. Non-treated, cultured in the same medium but not treated with low pH. **d**, GFP$^+$ (green) and GFP$^-$ (yellow) cell populations (average cell numbers per visual field; ×10 objective lens). $n = 25$; error bars show average ± s.d. **e**, Snapshots of live imaging of culture of low-pH-treated CD45$^+$ cells (*Oct4-gfp*). Arrows indicate cells that started expressing *Oct4*-GFP. Scale bar, 50 μm. **f**, Cell size reduction in low-pH-treated CD45$^+$ cells on day 1 before turning on *Oct4*-GFP without cell division on day 2. In this live imaging, cells were plated at a half density for easier viewing of individual cells. Scale bar, 10 μm. **g**, Electron microscope analysis. Scale bar, 1 μm. **h**, Forward scattering analysis of *Oct4*-GFP$^-$CD45$^+$ cells (red) and *Oct4*-GFP$^+$CD45$^-$ cells (green) on day 7. Blue line, ES cells. **i**, Genomic PCR analysis of (D)J recombination at the *Tcrb* gene. GL is the size of the non-rearranged germline type, whereas the smaller ladders correspond to the alternative rearrangements of J exons. Negative controls, lanes 1, 2; positive controls, lane 3; FACS-sorted *Oct4*-GFP$^+$ cells (two independent preparations on day 7), lanes 4, 5.

Most of the surviving cells on day 1 were still CD45$^+$ and *Oct4*-GFP$^-$. On day 3, the total cell numbers were reduced to between one-third to one-half of the day 0 population (Fig. 1d; see Extended Data Fig. 1g, h for apoptosis analysis), and a substantial number of total surviving cells became *Oct4*-GFP$^+$ (Fig. 1d), albeit with relatively weak signal intensity. On day 7, a significant number of *Oct4*-GFP$^+$CD45$^-$ cells (one-half to two-thirds of total surviving cells) constituted a distinct population from the *Oct4*-GFP$^-$CD45$^-$ cells (Fig. 1c, top, day 7, and Fig. 1d). No obvious generation of *Oct4*-GFP$^+$CD45$^-$ populations was seen in non-treated CD45$^+$ cells cultured similarly but without low-pH treatment (Fig. 1c, bottom).

Low-pH-treated CD45$^+$ cells, but not untreated cells, gradually turned on GFP signals over the first few days (Fig. 1e, Supplementary Videos 1 and 2 and Extended Data Fig. 2a), whereas CD45 immunoreactivity became gradually reduced in the cells that demonstrated *Oct4*-GFP expression (Fig. 1f and Extended Data Fig. 2b). By day 5, the *Oct4*-GFP$^+$ cells attached together and formed clusters by accretion. These GFP$^+$ clusters (but not GFP$^-$ cells) were quite mobile and often showed cell processes on moving (Supplementary Video 1).

The *Oct4*-GFP$^+$ cells demonstrated a characteristic small cell size with little cytoplasm and also showed a distinct fine structure of the nucleus compared with that of parental CD45$^+$ lymphocytes (Fig. 1g). The *Oct4*-GFP$^+$ cells on day 7 were smaller than non-treated CD45$^+$ cells (Fig. 1g, h and Extended Data Fig. 2c) and embryonic stem (ES) cells (Fig. 1h), both of which are generally considered to be small in size. The diameter of low-pH-treated CD45$^+$ cells became reduced during the first 2 days, even before they started *Oct4*-GFP expression (Fig. 1f), whereas the onset of GFP expression was not accompanied by cell divisions. Consistent with this, no substantial 5-ethynyl-2′-deoxyuridine (EdU) uptake was observed in the *Oct4*-GFP$^+$ cells after the stressor (Extended Data Fig. 2d).

The lack of substantial proliferation argues against the possibility that CD45$^-$ cells, contaminating as a very minor population in the FACS-sorted CD45$^+$ cells, quickly grew and formed a substantial *Oct4*-GFP$^+$ population over the first few days after the low-pH treatment. In addition, genomic rearrangements of *Tcrb* (T-cell receptor gene) were observed in *Oct4*-GFP$^+$ cells derived from FACS-purified CD45$^+$ cells and CD90$^+$CD45$^+$ T cells (Fig. 1i, lanes 4, 5, and Extended Data Fig. 2e–g), indicating at least some contribution from lineage-committed T cells. Thus, *Oct4*-GFP$^+$ cells were generated *de novo* from low-pH-treated CD45$^+$ haematopoietic cells by reprogramming, rather than by simple selection of stress-enduring cells[23].

Low-pH-induced Oct4$^+$ cells have pluripotency

On day 7, the *Oct4*-GFP$^+$ spheres expressed pluripotency-related marker proteins[22] (Oct4, SSEA1, Nanog and E-cadherin; Fig. 2a) and marker genes (*Oct4*, *Nanog*, *Sox2*, *Ecat1* (also called *Khdc3*), *Esg1* (*Dppa5a*), *Dax1* (*Nrob1*) and *Rex1* (*Zfp42*); Fig. 2b and Extended Data Fig. 3a) in a manner comparable to those seen in ES cells[24]. Moderate levels of expression of these pluripotency marker genes were observed on day 3

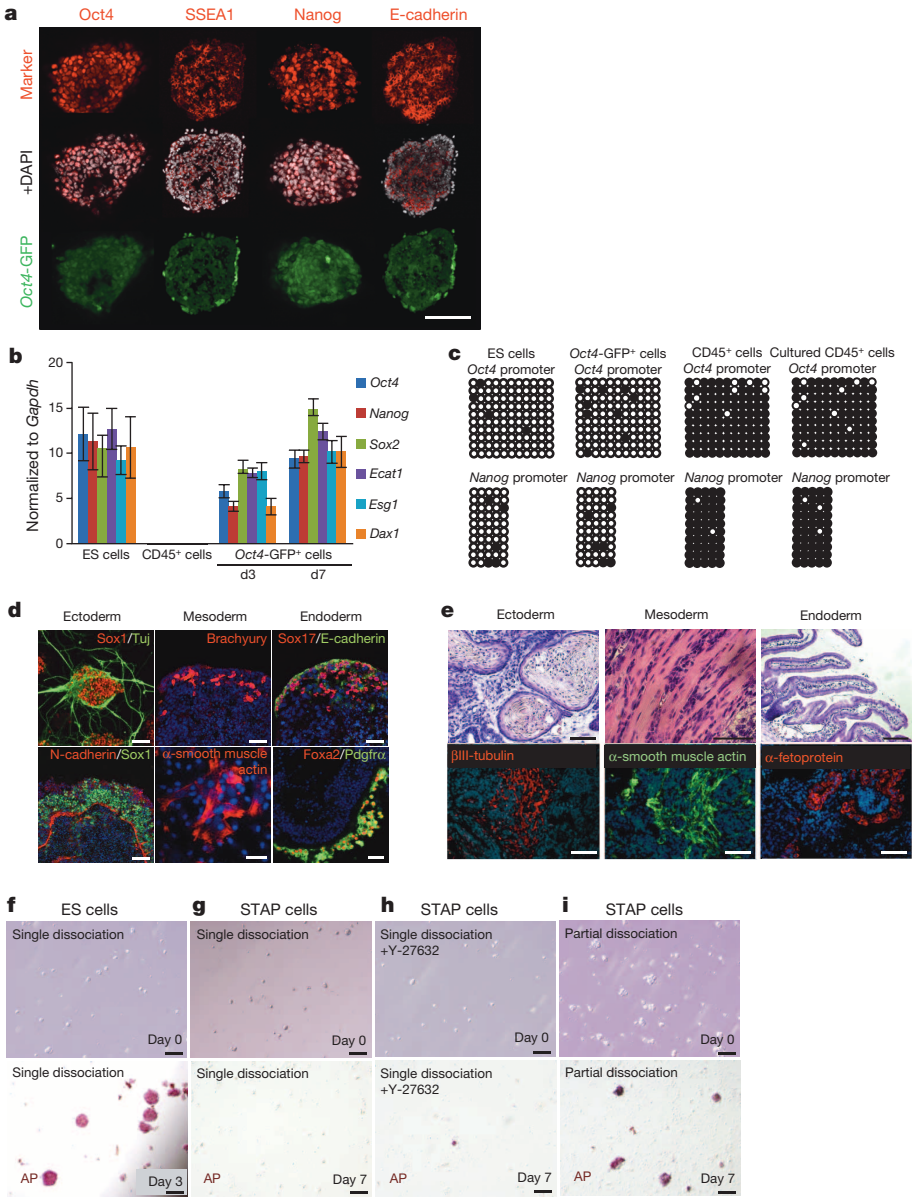


Figure 2 | Low-pH-induced *Oct4*-GFP$^+$ cells represent pluripotent cells. **a**, Immunostaining for pluripotent cell markers (red) in day 7 *Oct4*-GFP$^+$ (green) clusters. DAPI, white. Scale bar, 50 μm. **b**, qPCR analysis of pluripotency marker genes. From left to right, mouse ES cells; parental CD45$^+$ cells; low-pH-induced *Oct4*-GFP$^+$ cells on day 3; low-pH-induced *Oct4*-GFP$^+$ cells on day 7. $n = 3$; error bars show average ± s.d. **c**, DNA methylation study by bisulphite sequencing. Filled and open circles indicate methylated and non-methylated CpG, respectively. **d**, Immunostaining analysis of *in vitro* differentiation capacity of day 7 *Oct4*-GFP$^+$ cells. Ectoderm: the neural markers Sox1/Tuj1 (100%, $n = 8$) and N-cadherin (100%, $n = 5$). Mesoderm: smooth muscle actin (50%, $n = 6$) and brachyury (40%, $n = 5$). Endoderm: Sox17/E-cadherin (67%, $n = 6$) and Foxa2/Pdgfrα (67%, $n = 6$). Scale bar, 50 μm. **e**, Teratoma formation assay of day 7 clusters of *Oct4*-GFP$^+$ cells. Haematoxylin and eosin staining showed keratinized epidermis (ectoderm), skeletal muscle (mesoderm) and intestinal villi (endoderm), whereas immunostaining showed expression of Tuj1 (neurons), smooth muscle actin and α-fetoprotein. Scale bar, 100 μm. **f**–**i**, Dissociation culture of ES cells and STAP cells (additional 7 days from day 7; **f**, **g**) on gelatin-coated dishes. Top, bright-field; bottom, alkaline phosphatase (AP) staining. Partially dissociated STAP cells slowly generated small colonies (**i**), whereas dissociated STAP cells did not, even in the presence of the ROCK inhibitor (**g**, **h**), which allows dissociation culture of EpiSCs[29].

(Fig. 2b and Extended Data Fig. 3b). Notably, the *Oct4*-GFP$^+$ cells on day 3, but not on day 7, expressed early haematopoietic marker genes such as *Flk1* (also called *Kdr*) and *Tal1* (Extended Data Fig. 3c), indicating that *Oct4*-GFP$^+$ cells on day 3, as judged by their expression pattern at the population level, were still in a dynamic process of conversion.

On day 7, unlike CD45$^+$ cells and like ES cells, low-pH-induced *Oct4*-GFP$^+$ cells displayed extensive demethylation at the *Oct4* and *Nanog* promoter areas (Fig. 2c), indicating that these cells underwent a substantial reprogramming of epigenetic status in these key genes for pluripotency.

In vitro differentiation assays[25–27] demonstrated that low-pH-induced *Oct4*-GFP$^+$ cells gave rise to three-germ-layer derivatives (Fig. 2d) as well as visceral endoderm-like epithelium (Extended Data Fig. 3d). When grafted into mice, low-pH-induced *Oct4*-GFP$^+$ cell clusters formed teratomas (40%, $n = 20$) (Fig. 2e and Extended Data Fig. 4a–c) but no teratocarcinomas that persistently contained *Oct4*-GFP$^+$ cells ($n = 50$). Because some cellular variation was observed in the signal levels of *Oct4*-GFP within the clusters, we sorted GFP-strong cells (a major population) and GFP-dim cells (a minor population) by FACS on day 7 and separately injected them into mice. In this case, only GFP-strong cells formed teratomas (Extended Data Fig. 4d). In quantitative polymerase chain reaction (qPCR) analysis, the GFP-strong population expressed pluripotency marker genes but not early lineage-specific marker genes, whereas the GFP-dim cells showed substantial expression of some early lineage-specific marker genes (*Flk1*, *Gata2*, *Gata4*, *Pax6* and *Sox17*; Extended Data Fig. 4e) but not *Nanog* and *Rex1*. These observations indicate that three-germ-layer derivatives were generated from the GFP-strong cells expressing pluripotency marker genes, rather than from GFP-dim cells that seem to contain partially reprogrammed cells.

Collectively, these findings show that the differentiation state of a committed somatic cell lineage can be converted into a state of pluripotency by strong stimuli given externally. Hereafter, we refer to the fate conversion from somatic cells into pluripotent cells by strong external stimuli such as low pH as 'stimulus-triggered acquisition of pluripotency' (STAP) and the resultant cells as STAP cells. Under their establishment conditions, these STAP cells were rarely proliferative (Extended Data Figs 2d and 5a, b). Comparative genomic hybridization array analysis of STAP cells indicated no major global changes in chromosome number (Extended Data Fig. 5c).

STAP cells compared to ES cells

STAP cells, unlike mouse ES cells, showed a limited capacity for self-renewal in the LIF-containing medium and did not efficiently form

colonies in dissociation culture (Fig. 2f, g), even in the presence of the ROCK inhibitor Y-27632, which suppresses dissociation-induced apoptosis[28,29] (Fig. 2h). Also, even under high-density culture conditions after partial dissociation (Fig. 2i), STAP cell numbers started to decline substantially after two passages. Furthermore, expression of the ES cell marker protein Esrrβ was low in STAP cells (Extended Data Fig. 5d, e). In general, female ES cells do not show X-chromosomal inactivation[30] and contain no H3K27me3-dense foci (indicative of inactivated X chromosomes), unlike female CD45$^+$ cells and EpiSCs. In contrast, H3K27me3-dense foci were found in ~40% of female STAP cells strongly positive for *Oct4*-GFP (Extended Data Fig. 5f, g).

STAP cells were also dissimilar to mouse EpiSCs, another category of pluripotent stem cell[21,22,29,31], and were positive for Klf4 and negative for the epithelial tight junction markers claudin 7 and ZO-1 (Extended Data Fig. 5d, e).

STAP cells from other tissue sources

We next performed similar conversion experiments with somatic cells collected from brain, skin, muscle, fat, bone marrow, lung and liver tissues of 1-week-old *Oct4-gfp* mice. Although conversion efficacy varied, the low-pH-triggered generation of *Oct4*-GFP$^+$ cells was observed in day 7 culture of all tissues examined (Fig. 3a and Extended Data Fig. 6a–c), including mesenchymal cells of adipose tissues (Fig. 3a–c) and neonatal cardiac cells that were negatively sorted for CD45 by FACS (Fig. 3d–g; see Extended Data Fig. 6d for suppression of cardiac genes such as *Nkx2-5* and cardiac actin).

Chimaera formation and germline transmission in mice

We next performed a blastocyst injection assay with STAP cells that were generated from CD45$^+$ cells of neonatal mice constitutively expressing GFP (this C57BL/6 line with *cag-gfp* transgenes is referred to hereafter as B6GFP). We injected STAP cell clusters en bloc that were manually cut into small pieces using a microknife (Fig. 4a). A high-to-moderate contribution of GFP-expressing cells was seen in the chimaeric embryos (Fig. 4b and Extended Data Fig. 7a). These chimaeric mice were born at a substantial rate and all developed normally (Fig. 4c and Extended Data Fig. 7b).

CD45$^+$ cell-derived STAP cells contributed to all tissues examined (Fig. 4d). Furthermore, offspring derived from STAP cells were born to the chimaeric mice (Fig. 4e and Extended Data Fig. 7c), demonstrating their germline transmission, which is a strict criterion for pluripotency as well as genetic and epigenetic normality[32,33]. Furthermore, in a tetraploid (4N) complementation assay, which is considered to be the most rigorous test for developmental potency[34,35] (Fig. 4a, bottom), CD45$^+$ cell-derived STAP cells (from F_1 mice of B6GFP × 129/Sv or DBA/2) generated all-GFP$^+$ embryos on embryonic day (E)10.5 (Fig. 4f, Extended Data Fig. 7d and Supplementary Video 3), demonstrating that STAP cells alone are sufficient to construct an entire embryonic structure. Thus, STAP cells have the developmental capacity to differentiate into all somatic-cell lineages as well as germ-cell lineages *in vivo*.

Expandable pluripotent cell lines from STAP cells

STAP cells have a limited self-renewal capacity under the conditions used for establishment (Fig. 2g and Extended Data Figs 2e and 5a). However, in the context of the embryonic environment, a small fragment of a STAP cell cluster could grow even into a whole embryo (Fig. 4f). With this in mind, we next examined whether STAP cells have the potential to generate expandable pluripotent cell lines *in vitro* under certain conditions.

STAP cells could not be efficiently maintained for additional passages in conventional LIF+FBS-containing medium or 2i medium[20] (most STAP cells died in 2i medium within 7 days; Extended Data Fig. 8a). Notably, an adrenocorticotropic hormone (ACTH)+LIF-containing medium (hereafter called ACTH medium) known to facilitate clonal expansion of ES cells[36] supported outgrowth of STAP cell colonies. When cultured in this medium on a MEF feeder or gelatin, a portion of STAP cell clusters started to grow (Fig. 5a, bottom; such outgrowth was typically found in 10–20% of wells in single cluster culture using 96-well plates and in >75% when 12 clusters were plated

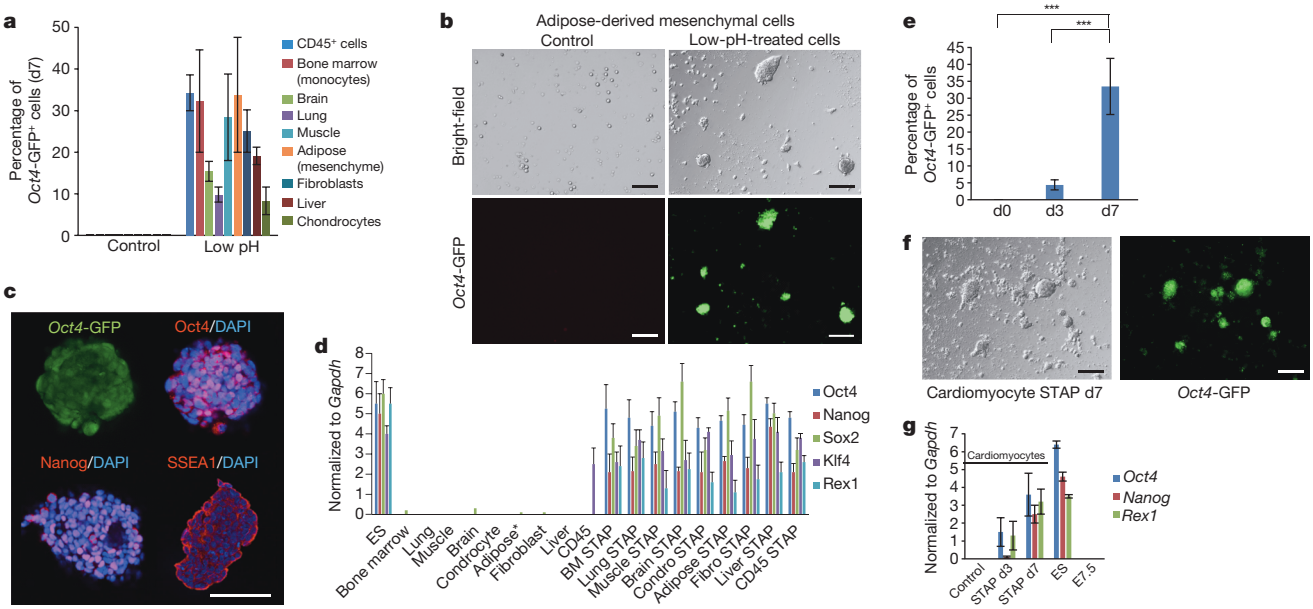


Figure 3 | STAP cell conversion from a variety of cells by low-pH treatment. **a**, Percentage of *Oct4*-GFP$^+$ cells in day 7 culture of low-pH-treated cells from different origins (1×10^5 cells per ml × 3 ml). The number of surviving cells on day 7 compared to the plating cell number was 20–30%, except for lung, muscle and adipose cells, for which surviving cells were ~10% (n=3, average ± s.d.). **b**, *Oct4*-GFP$^+$ cell clusters were induced by low-pH treatment from adipose-tissue-derived mesenchymal cells on day 7. Scale bar, 100 μm. **c**, Expression of pluripotent cell markers in day 7 clusters of low-pH-treated adipose-tissue-derived mesenchymal cells. Scale bar, 50 μm. **d**, Expression of pluripotency marker genes in STAP cells derived from various tissues. Gene expressions were normalized by *Gapdh* (n = 3, average ± s.d.). Asterisk indicates adipose tissue-derived mesenchymal cells. **e**, Quantification of *Oct4*-GFP$^+$ cells in culture of low-pH-treated neonatal cardiac muscle cells. ***$P < 0.001$; Tukey's test (n = 3). **f**, Generation of *Oct4*-GFP$^+$ cell clusters (d7) from CD45$^-$ cardiac muscle cells. **g**, qPCR analysis of pluripotency marker genes in STAP cells from CD45$^-$ cardiac muscle cells.

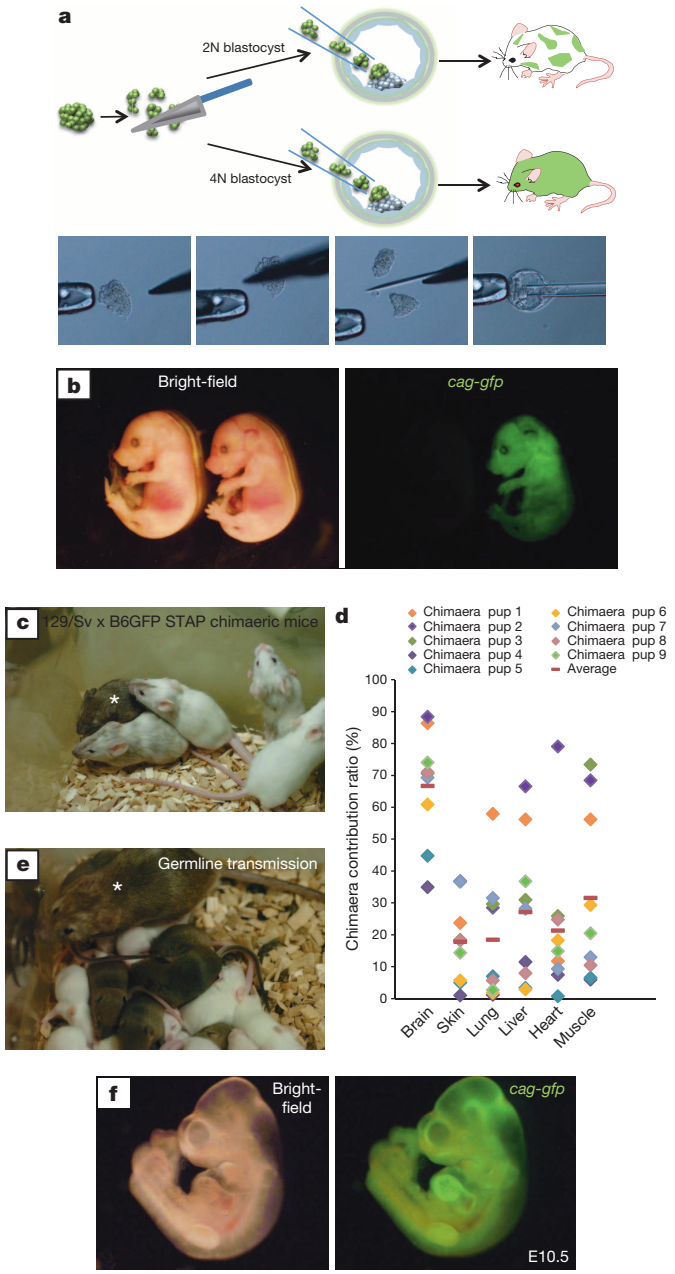


Figure 4 | Chimaeric mouse generation from STAP cells. **a**, Schematic of chimaeric mouse generation. **b**, E13.5 chimaera fetuses from 2N blastocytes injected with STAP cells (derived from B6GFP CD45$^+$ cells carrying *cag-gfp*). **c**, Adult chimaeric mice generated by STAP-cell (B6GFP × 129/Sv; agouti) injection into blastocysts (ICR strain; albino). Asterisk indicates a highly contributed chimaeric mouse. **d**, Chimaera contribution analysis. Tissues from nine pups were analysed by FACS. **e**, Offspring of chimaeric mice derived from STAP cells. Asterisk indicates the same chimaeric mouse shown in **c**. **f**, E10.5 embryo generated in the tetraploid complementation assay with STAP cells (B6GFP × 129/Sv).

per well). These growing colonies looked similar to those of mouse ES cells and expressed a high level of *Oct4*-GFP.

After culturing in ACTH medium for 7 days, this growing population of cells, unlike parental STAP cells, could be passaged as single cells (Fig. 5a, bottom, and Fig. 5b), grow in 2i medium (Extended Data Fig. 8a) and expand exponentially, up to at least 120 days of culture (Fig. 5c; no substantial chromosomal abnormality was seen; Extended Data Fig. 8b, c). Hereafter, we refer to the proliferative cells derived from STAP cells as STAP stem cells.

STAP stem cells expressed protein and RNA markers for pluripotent cells (Fig. 5d, e), showed low DNA methylation levels at the *Oct4* and *Nanog* loci (Extended Data Fig. 8d), and had a nuclear fine structure similar to that of ES cells (Extended Data Fig. 8e; few electron-dense areas corresponding to heterochromatin). In differentiation culture[25–27], STAP stem cells generated ectodermal, mesodermal and endodermal derivatives *in vitro* (Fig. 5f–h and Extended Data Fig. 8f, g), including beating cardiac muscles (Supplementary Video 4), and formed teratomas *in vivo* (Fig.5i and Extended Data Fig. 8h; no teratocarcinomas, $n = 40$). After blastocyst injection, STAP stem cells efficiently contributed to chimaeric mice (Fig. 5j), in which germline transmission was seen (Extended Data Fig. 8i). Even in tetraploid complementation assays, injected STAP stem cells could generate mice capable of growing to adults and producing offspring (Fig. 5k, l; in all eight independent lines, Extended Data Fig. 8j).

In addition to their expandability, we noticed at least two other differences between STAP stem cells and parental STAP cells. First, the expression of the ES cell marker protein Esrrβ, which was undetectable in STAP cells (Extended Data Fig. 5d, e), was clearly seen in STAP stem cells (Fig. 5e). Second, the presence of H3K27me3 foci, which was found in a substantial proportion of female STAP cells, was no longer observed in STAP stem cells (Extended Data Figs 5f and 8k). Thus, STAP cells have the potential to give rise to expandable cell lines that exhibit features similar to those of ES cells.

Discussion

This study has revealed that somatic cells latently possess a surprising plasticity. This dynamic plasticity—the ability to become pluripotent cells—emerges when cells are transiently exposed to strong stimuli that they would not normally experience in their living environments.

Low-pH treatment was also used in the 'autoneuralization' experiment[15–17] by Holtfreter in 1947, in which exposure to acidic medium caused tissue-autonomous neural conversion of salamander animal caps *in vitro* in the absence of Spemann's organizer signals. Although the mechanism has remained elusive, Holtfreter hypothesized that the strong stimulus releases the animal cap cells from some intrinsic inhibitory mechanisms that suppress fate conversion or, in his words, they pass through 'sublethal cytolysis' (meaning stimulus-evoked lysis of the cell's inhibitory state)[15,37]. Although Holtfreter's study and ours differ in the direction of fate conversion—orthograde differentiation and nuclear reprogramming, respectively—these phenomena may share some common aspects, particularly with regard to sublethal stimulus-evoked release from a static (conversion-resisting) state in the cell.

A remaining question is whether cellular reprogramming is initiated specifically by the low-pH treatment or also by some other types of sublethal stress such as physical damage, plasma membrane perforation, osmotic pressure shock, growth-factor deprivation, heat shock or high Ca^{2+} exposure. At least some of these stressors, particularly physical damage by rigorous trituration and membrane perforation by streptolysin O, induced the generation of *Oct4*-GFP$^+$ cells from CD45$^+$ cells (Extended Data Fig. 9a; see Methods). These findings raise the possibility that certain common regulatory modules, lying downstream of these distantly related sublethal stresses, act as a key for releasing somatic cells from the tightly locked epigenetic state of differentiation, leading to a global change in epigenetic regulation. In other words, unknown cellular functions, activated by sublethal stimuli, may set somatic cells free from their current commitment to recover the naive cell state.

Our present finding of an unexpectedly large capacity for radical reprogramming in committed somatic cells raises various important questions. For instance, why, and for what purpose, do somatic cells latently possess this self-driven ability for nuclear reprogramming, which emerges only after sublethal stimulation, and how, then, is this reprogramming mechanism normally suppressed? Furthermore, why isn't teratoma (or pluripotent cell mass) formation normally seen in *in vivo* tissues that may receive strong environmental stress? In our preliminary study, experimental reflux oesophagitis locally induced moderate

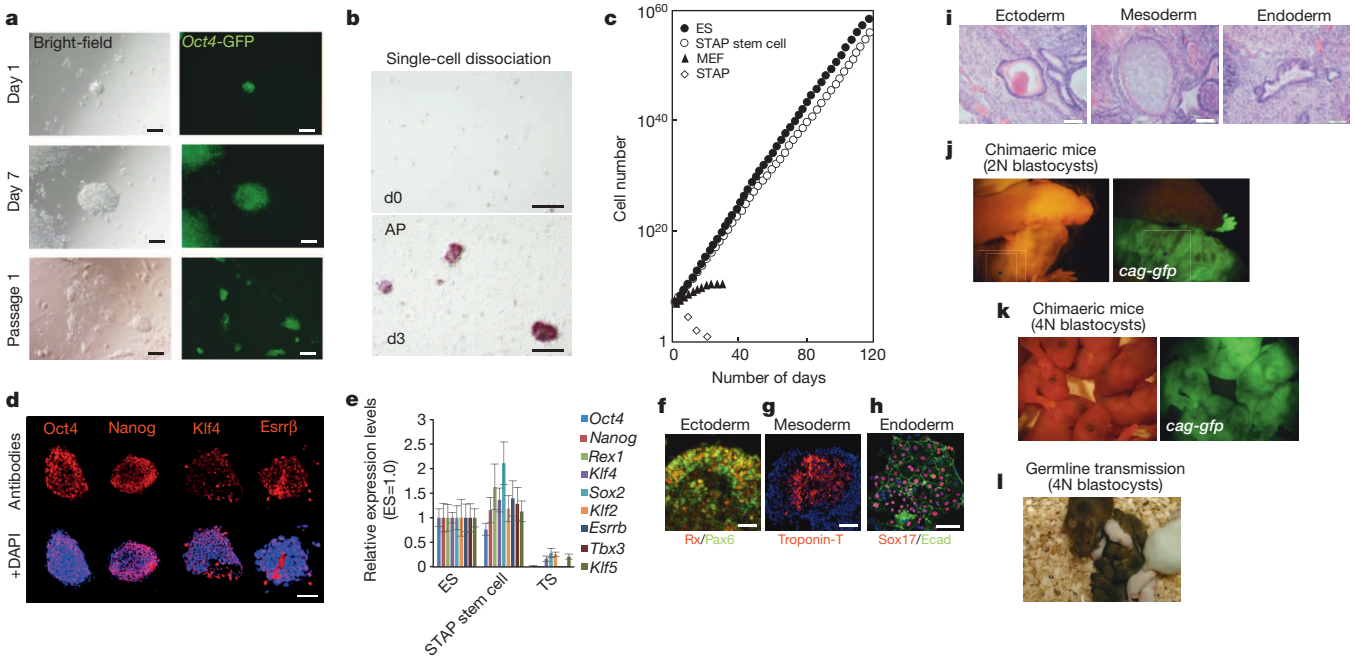


Figure 5 | ES-cell-like stem cells can be derived from STAP cells. **a**, Growth of STAP stem cells carrying *Oct4-gfp*. Scale bar, 50 μm. **b**, Dissociation culture of STAP stem cells to form colonies. Scale bar, 100 μm. **c**, Robust growth of STAP stem cells in maintenance culture. Similar results were obtained with eight independent lines. In contrast, parental STAP cells decreased in number quickly. **d**, Immunostaining of STAP stem cells for pluripotency markers (red). Scale bar, 50 μm. **e**, qPCR analysis of pluripotency marker gene expression. **f**–**h**, *In vitro* differentiation assays into three-germ-layer derivatives. **f**, Ectoderm: $Rx^+/Pax6^+$ (retinal epithelium; 83%, $n = 6$). **g**, Mesoderm: troponin-T^+ (cardiac muscle; 50%, $n = 6$). **h**, Endoderm: $Sox17^+$/E-$cadherin^+$ (endodermal progenitors; 67%, $n = 6$). Scale bar, 50 μm. **i**, Teratoma formation assays. Formation of keratinized epidermis (ectoderm; left), cartilage (mesoderm; middle) and bronchial-like epithelium (endoderm; right) is shown. Scale bar, 100 μm. **j**, Blastocyst injection assays. These pictures of live animals were taken serially (asterisk indicates the same chimaeric pup). **k**, **l**, Tetraploid complementation assay. 'All-GFP^+' pups were born (**k**) and germline transmission was observed (**l**).

expression of *Oct4*-GFP but not endogenous Nanog in the mouse oesophageal mucosa (Extended Data Fig. 9b). Therefore, an intriguing hypothesis for future research is that the progression from initial *Oct4* activation to further reprogramming is suppressed by certain inhibitory mechanisms *in vivo*.

The question of why and how this self-driven reprogramming is directed towards the pluripotent state is fundamentally important, given that STAP reprogramming takes a remarkably short period, only a few days for substantial expression of pluripotency marker genes, unlike transgene- or chemical-induced iPS cell conversion[38]. Thus, our results cast new light on the biological meaning of diverse cellular states in multicellular organisms.

METHODS SUMMARY

Tissue collection and low-pH exposure. To isolate haematopoietic cells, spleens were excised from 1-week-old *Oct4-gfp* C57BL/6 mice, minced by scissors and mechanically dissociated with pasture pipettes. Dissociated cells were collected, re-suspended in DMEM medium and added to the same volume of lympholyte (Cedarlane), then centrifuged at 1,000*g* for 20 min. CD45-positive cells were sorted by FACS Aria (BD Biosciences), and treated with low-pH HBSS solution (pH 5.7 for 25 min at 37 °C), centrifuged for 5 min to remove supernatant, and plated to non-adhesive culture plates in DMEM/F12 medium supplemented with 1,000 U LIF (Sigma) and B27 (Invitrogen). Although *Oct4*-GFP^+ cells (expressing pluripotency-related protein and gene markers and capable of differentiating into three germ-layer derivatives) were also generated from lymphocytes of young adult mice (for example, 6-week-old) under the same culture conditions, their proportion in culture was reduced by several to ten folds as compared to neonatal lymphocytes when lymphocytes were isolated from 1-month-old mice or older. Live imaging was performed using specially assembled confocal microscope systems with a CO_2 incubator[39], and CD45 immunoreactivity in live cells was examined as described[40].

***In vivo* and *in vitro* differentiation assay.** STAP cells were seeded onto a sheet $3 \times 3 \times 1$ mm, composed of a non-woven mesh of polyglycolic acid fibres and implanted subcutaneously into the dorsal flanks of 4-week-old NOD/SCID mice. To examine *in vitro* differentiation, STAP cells and STAP stem cells were collected at 7 days and subjected to SDIA or SFEBq culture[25,26] for neural differentiation and to embryoid body culture for mesodermal and endodermal[27] differentiation.

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 10 March; accepted 20 December 2013.

1. Gurdon, J. B. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* **10,** 622–640 (1962).
2. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394,** 369–374 (1998).
3. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
4. Thorpe, T. A. History of plant tissue culture. *Mol. Biotechnol.* **37,** 169–180 (2007).
5. Jiang, Y. *et al.* Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* **418,** 41–49 (2002).
6. D'Ippolito, G. *et al.* Marrow-isolated adult multilineage inducible (MIAMI) cells, a unique population of postnatal young and old human cells with extensive expansion and differentiation potential. *J. Cell Sci.* **117,** 2971–2981 (2004).
7. Kucia, M. *et al.* A population of very small embryonic-like (VSEL) $CXCR4^+SSEA\text{-}1^+Oct\text{-}4^+$ stem cells identified in adult bone marrow. *Leukemia* **20,** 857–869 (2006).
8. Kuroda, Y. *et al.* Unique multipotent cells in adult human mesenchymal cell populations. *Proc. Natl Acad. Sci. USA* **107,** 8639–8643 (2010).
9. Obokata, H. *et al.* The potential of stem cells in adult tissues representative of the three germ layers. *Tissue Eng. Part A* **17,** 607–615 (2011).
10. Lengner, C. J., Welstead, G. G. & Jaenisch, R. The pluripotency regulator Oct4: a role in somatic stem cells? *Cell Cycle* **7,** 725–728 (2008).
11. Berg, J. S. & Goodell, M. A. An argument against a role for Oct4 in somatic stem cells. *Cell Stem Cell* **1,** 359–360 (2007).
12. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460,** 1132–1135 (2009).
13. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–264 (2008).
14. Ohbo, K. *et al.* Identification and characterization of stem cells in prepubertal spermatogenesis in mice small star, filled. *Dev. Biol.* **258,** 209–225 (2003).
15. Holtfreter, J. Neural induction in explants which have passed through a sublethal cytolysis. *J. Exp. Zool.* **106,** 197–222 (1947).

16. Gerhart, J. Johannes Holtfreter's contributions to ongoing studies of the organizer. *Dev. Dyn.* **205,** 245–256 (1996).
17. Byrnes, W. M. Ernest Everett Just, Johannes Holtfreter, and the origin of certain concepts in embryo morphogenesis. *Mol. Reprod. Dev.* **76,** 912–921 (2009).
18. Gratama, J. W., Sutherland, D. R. & Keeney, M. Flow cytometric enumeration and immunophenotyping of hematopoietic stem and progenitor cells. *Semin. Hematol.* **38,** 139–147 (2001).
19. Inoue, K. *et al.* Inefficient reprogramming of the hematopoietic stem cell genome following nuclear transfer. *J. Cell Sci.* **119,** 1985–1991 (2006).
20. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
21. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448,** 196–199 (2007).
22. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448,** 191–195 (2007).
23. Kuroda, Y. *et al.* Isolation, culture and evaluation of multilineage-differentiating stress-enduring (Muse) cells. *Nature Protocols* **8,** 1391–1415 (2013).
24. Niwa, H. How is pluripotency determined and maintained? *Development* **134,** 635–646 (2007).
25. Watanabe, K. *et al.* Directed differentiation of telencephalic precursors from embryonic stem cells. *Nature Neurosci.* **8,** 288–296 (2005).
26. Kawasaki, H. *et al.* Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. *Neuron* **28,** 31–40 (2000).
27. Gouon-Evans, V. *et al.* BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nature Biotechnol.* **24,** 1402–1411 (2006).
28. Ohgushi, M. & Sasai, Y. Lonely death dance of human pluripotent stem cells: ROCKing between metastable cell states. *Trends Cell Biol.* **21,** 274–282 (2011).
29. Ohgushi, M. *et al.* Molecular pathway and cell state responsible for dissociation-induced apoptosis in human embryonic stem cells. *Cell Stem Cell* **7,** 225–239 (2010).
30. Murakami, K., Araki, K., Ohtsuka, S., Wakayama, T. & Niwa, H. Choice of random rather than imprinted X inactivation in female embryonic stem cell-derived extra-embryonic cells. *Development* **138,** 197–202 (2011).
31. Pera, M. F. & Tam, P. P. Extrinsic regulation of pluripotent stem cells. *Nature* **465,** 713–720 (2010).
32. Surani, M. A. & Barton, S. C. Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. *Science* **222,** 1034–1036 (1983).
33. Wakayama, S. *et al.* Successful serial recloning in the mouse over multiple generations. *Cell Stem Cell* **12,** 293–297 (2013).
34. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90,** 8424–8428 (1993).
35. Eakin, G. S., Hadjantonakis, A. K., Papaioannou, V. E. & Behringer, R. R. Developmental potential and behavior of tetraploid cells in the mouse embryo. *Dev. Biol.* **288,** 150–159 (2005).
36. Ogawa, K., Matsui, H., Ohtsuka, S. & Niwa, H. A novel mechanism for regulating clonal propagation of mouse ES cells. *Genes Cells* **9,** 471–477 (2004).
37. Hurtado, C. & De Robertis, E. M. Neural induction in the absence of organizer in salamander is mediated by MAPK. *Dev. Biol.* **307,** 282–289 (2007).
38. Hou, P. *et al.* Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* **341,** 651–654 (2013).
39. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472,** 51–56 (2011).
40. Eilken, M. H., Nishikawa, S. & Schroeder, T. Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature* **457,** 896–900 (2009).

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

Acknowledgements We thank S. Nishikawa for discussion and J. D. Ross, N. Takata, M. Eiraku, M. Ohgushi, S. Itoh, S. Yonemura, S. Ohtsuka and K. Kakiguchi for help with experiments and analyses. We thank A. Penvose and K. Westerman for comments on the manuscript. H.O. is grateful to T. Okano, S. Tsuneda and K. Kuroda for support and encouragement. Financial support for this research was provided by Intramural RIKEN Research Budget (H.O., T.W. and Y.S.), a Scientific Research in Priority Areas (20062015) to T.W., the Network Project for Realization of Regenerative Medicine to Y.S., and Department of Anesthesiology, Perioperative and Pain Medicine at Brigham and Women's Hospital to C.A.V.

Author Contributions H.O. and Y.S. wrote the manuscript. H.O., T.W. and Y.S. performed experiments, and K.K. assisted with H.O.'s transplantation experiments. H.O., T.W., Y.S., H.N. and C.A.V. designed the project. M.P.V. and M.Y. helped with the design and evaluation of the project.

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 H.O. (obokata@cdb.riken.jp) or C.A.V. (cvacanti@partners.org).

ARTICLE

doi:10.1038/nature12922

ANP32E is a histone chaperone that removes H2A.Z from chromatin

Arnaud Obri[1]*, Khalid Ouararhni[1]*, Christophe Papin[1]*, Marie-Laure Diebold[2], Kiran Padmanabhan[3], Martin Marek[2], Isabelle Stoll[1], Ludovic Roy[1], Patrick T. Reilly[4], Tak W. Mak[4,5], Stefan Dimitrov[3], Christophe Romier[2] & Ali Hamiche[1]

H2A.Z is an essential histone variant implicated in the regulation of key nuclear events. However, the metazoan chaperones responsible for H2A.Z deposition and its removal from chromatin remain unknown. Here we report the identification and characterization of the human protein ANP32E as a specific H2A.Z chaperone. We show that ANP32E is a member of the presumed H2A.Z histone-exchange complex p400/TIP60. ANP32E interacts with a short region of the docking domain of H2A.Z through a new motif termed H2A.Z interacting domain (ZID). The 1.48 Å resolution crystal structure of the complex formed between the ANP32E–ZID and the H2A.Z/H2B dimer and biochemical data support an underlying molecular mechanism for H2A.Z/H2B eviction from the nucleosome and its stabilization by ANP32E through a specific extension of the H2A.Z carboxy-terminal α-helix. Finally, analysis of H2A.Z localization in ANP32E$^{-/-}$ cells by chromatin immunoprecipitation followed by sequencing shows genome-wide enrichment, redistribution and accumulation of H2A.Z at specific chromatin control regions, in particular at enhancers and insulators.

H2A.Z is a histone variant that is highly conserved across all eukaryotes[1], and is required for survival in fly and mouse[2,3]. H2A.Z belongs to the H2A family but shares only limited conservation with conventional H2A[1]. Current available data show that H2A.Z is implicated in key biological processes ranging from transcription to DNA repair[4–20] as well as cancer initiation and progression[11,21].

The overall crystal structure of the H2A.Z nucleosome is similar to the conventional one[22]. However, the H2A.Z nucleosome shows an altered H2A.Z/H2B dimer interface and an extended accessible acidic surface, the acidic patch, determined by the negative charge of the docking domain of H2A.Z[22]. Intriguingly, a small part of this docking domain of H2A.Z, known as the M6 cassette and containing its carboxy-terminal α-helix (αC), is required for its specific function and *Drosophila* survival[2].

The proper incorporation of conventional and variant histones into the nucleosome is achieved by specific histone chaperones (for reviews see refs 23, 24). In yeast, the incorporation of H2A.Z (Htz1) into chromatin is mediated by the histone chaperone Chz1 (refs 25, 26) and the histone-exchanging nucleosome remodelling complex SWR1 (refs 27–29). Two SWR1-related multiprotein complexes, SRCAP and p400/TIP60, were described in higher eukaryotes[30–33]. However, no H2A.Z-specific histone chaperone has been identified so far in metazoans.

Using a double immunoaffinity procedure, we have isolated the H2A.Z predeposition complex from HeLa cells. We identified the acidic nuclear phosphoprotein 32 kilodalton e (ANP32E) as a member of this complex. We show that ANP32E is an H2A.Z chaperone able specifically to remove H2A.Z from the nucleosome. Our biochemical and structural data further provide the molecular basis for H2A.Z recognition and H2A.Z/H2B nucleosomal removal by ANP32E. Finally, genome-wide chromatin immunoprecipitation followed by sequencing (ChIP-seq) shows that ANP32E regulates H2A.Z deposition at promoters and strikingly preserves enhancers and insulator sites free of H2A.Z nucleosomes.

H2A.Z predeposition complex purification

To identify and analyse how mammalian H2A.Z chaperones function, we established stable HeLa cell lines expressing conventional H2A (e-H2A) or H2A.Z (e-H2A.Z) histones fused to amino (N)-terminal double haemagglutinin (HA) and Flag tags (Extended Data Fig. 1a). We next isolated the respective predeposition complexes (e-H2A.com and e-H2A.Z.com) by double immunoaffinity purification from these cell lines[34,35]. The complexes were separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and their components were identified by mass spectrometry (Fig. 1a, b).

The e-H2A.Z.com contained several specific protein subunits (p400, TRRAP, SRCAP, Brd8, Tip60, YL1, ING3), which were not present in the conventional e-H2A.complex (Fig. 1a, b and Extended Data Table 1a, b). These specific subunits are part of either the SRCAP or the p400/TIP60 complexes, which are thought to exchange H2A for H2A.Z in chromatin[27–30,32,33,36]. Intriguingly, the main interacting partner in e-H2A.Z.com was a 32 kilodalton protein, which we identified as the ANP32E protein (Fig. 1b, d and Extended Data Table 1b). Fractionation of e-H2A.Z.com on a glycerol gradient confirmed that ANP32E is a stable component of this complex (Extended Data Fig. 1b, c).

Purification of the nuclear e-ANP32E.com from HeLa cell lines stably expressing epitope-tagged ANP32E (e-ANP32E) further showed that e-ANP32E.com was associated, as expected, with large amounts of H2A.Z (Fig. 1c, e and Extended Data Table 1c, d). Several other proteins (p400, TRRAP, Brd8, TIP60, Tip49 a, b, ING3) were identified as members of the e-ANP32E.com (Fig. 1c and Extended Data Table 1c). SRCAP was not found in the e-ANP32E.com. We conclude that ANP32E is a member of the p400/TIP60 complex, but not of the SRCAP complex. In agreement with this, we found that ANP32E interacts through its N-terminal domain with MRGBP, a component of the p400/TIP60 complex (Extended Data Fig. 1d).

[1]Département de Génomique Fonctionnelle et Cancer, Institut de Génétique et Biologie Moléculaire et Cellulaire (IGBMC), Université de Strasbourg, CNRS, INSERM, 1 rue Laurent Fries, B.P. 10142, 67404 Illkirch Cedex, France. [2]Département de Biologie Structurale Intégrative, Institut de Génétique et Biologie Moléculaire et Cellulaire (IGBMC), Université de Strasbourg, CNRS, INSERM, 1 rue Laurent Fries, B.P. 10142, 67404 Illkirch Cedex, France. [3]Equipe labélisée Ligue contre le Cancer, INSERM/Université Joseph Fourier , Institut Albert Bonniot, U823, Site Santé-BP 170, 38042 Grenoble Cedex 9, France. [4]Laboratory of Inflammation Biology, Division of Cellular and Molecular Research, National Cancer Centre, Singapore, Singapore. [5]The Campbell Family Institute for Breast Cancer Research, University Health Network, Toronto, Ontario, Canada.
*These authors contributed equally to this work.

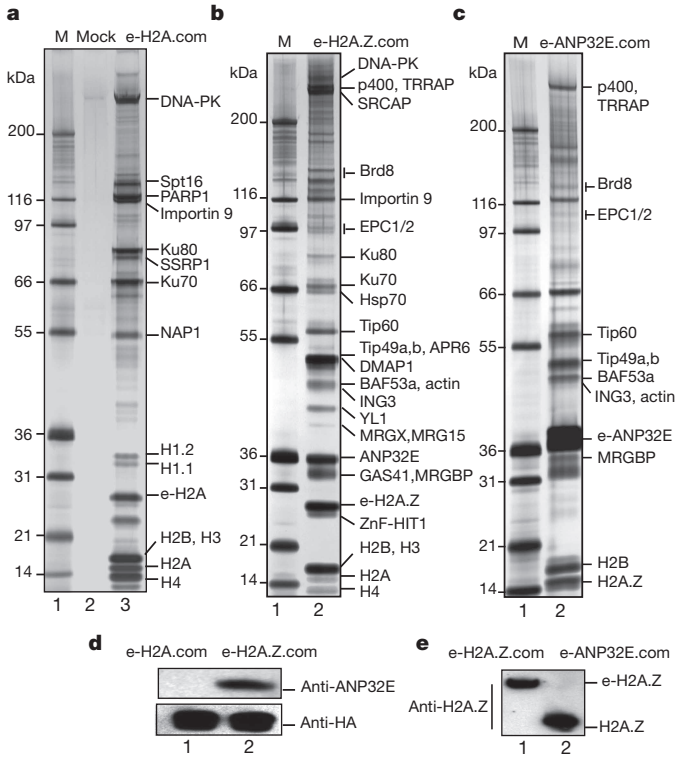


Figure 1 | Immunopurification of e-H2A, e-H2A.Z and e-ANP32E predeposition complexes from soluble nuclear fractions. a–c, Silver staining of proteins associated with (**a**) e-H2A, (**b**) e-H2A.Z or (**c**) e-ANP32E. 'Mock' indicates purification from a non-tagged HeLa cell line. **d**, Western blot analysis of ANP32E in e-H2A and e-H2A.Z predeposition complexes. **e**, Western blot analysis of H2A.Z in either e-H2A.Z.com or e-Anp32.com. All complex purifications shown are representative of $n \geq 3$ for each experiment.

ANP32E specifically interacts with H2A.Z

ANP32E is a highly conserved vertebrate protein belonging to the ANP32 family[37]. The ANP32 proteins are involved in several biological processes, ranging from transcription to apoptosis[38,39]. ANP32 proteins have conserved N-terminal domains that are composed of leucine-rich repeats, and divergent acidic carboxy (C)-terminal domains of unknown functions (Fig. 2a)[37]. The ANP32B N-terminal domain has been shown to chaperone the H3/H4 pair[40].

Pull-down experiments using purified glutathione *S*-transferase (GST)–ANP32E as bait show that ANP32E binds to the H2A.Z/H2B dimer, but not to the conventional H2A/H2B dimer (Fig. 2b). This interaction is highly specific and very stable because it is not affected by treatment with 1 M NaCl (Fig. 2b). Deletion analyses of the ANP32E protein showed that only its C-terminal domain (amino acids 151–268), notably its very C-terminal part (amino acids 195–268), was able to interact strongly with the H2A.Z/H2B pair (Extended Data Fig. 2a, b).

Further deletion experiments showed that a small region (amino acids 215–240) in the ANP32E C terminus was sufficient for H2A.Z/H2B binding (Extended Data Fig. 2c). Accordingly, deletion of this region in the full-length ANP32E protein completely abolished binding of this protein to the H2A.Z/H2B dimer (Fig. 2c). We therefore termed this region ZID (H2A.Z interacting domain) (Fig. 2a, upper panel). Alignment of ANP32E sequences from various organisms shows that the ANP32E-ZID corresponds to the primary conserved region of the ANP32E C terminus (Extended Data Fig. 3a). Specifically, this region encompasses an insertion not found in the other members of the ANP32 family (Extended Data Fig. 3b).

H2A.Z αC–helix specifies ANP32E binding

The M6 cassette encompassing amino acids 89–100 of H2A.Z has been shown to be highly specific and crucial for its function (Fig. 2a, middle panel, and refs 2, 41, 42). On the basis of the very strong and specific interaction between H2A.Z and ANP32E, we considered that the M6 cassette could be crucial for driving this interaction. Accordingly, pull-down experiments using GST–ANP32E showed that the H2A.Z region encompassing amino acids 92–114, and thus containing the M6 cassette (Fig. 2a, middle panel), is necessary for the binding of ANP32E to H2A.Z (Extended Data Fig. 2d).

The main differences between the M6 cassettes of conventional H2A and H2A.Z reside in the small αC-helix (Fig. 2a, lower panel), suggesting that the H2A.Z αC-helix is required for binding to ANP32E. Indeed, in GST–ANP32E pull-down experiments using $H2A.Z_{NKLLG}$, an H2A.Z mutant that contains the αC-helix of H2A (Fig. 2a, lower panel), we did not detect any interaction (Fig. 2d).

To analyse if the αC-helix of H2A.Z is also essential for its binding to ANP32E in cells, we generated stable HeLa cell lines expressing the double HA and Flag epitope tagged $H2A.Z_{NKLLG}$ (e-$H2A.Z_{NKLLG}$), and purified the nuclear e-$H2A.Z_{NKLLG}$ complex (e-$H2A.Z_{NKLLG}$.com). The components of the e-$H2A.Z_{NKLLG}$.com were identified both by mass spectrometry and by western blotting (Fig. 2e and Extended Data Table 1e). As expected, both techniques showed that ANP32E was absent from the e-$H2A.Z_{NKLLG}$.com. Of note, p400, TRRAP, TIP60 and Brd8 were also missing from the e-$H2A.Z_{NKLLG}$.com (Fig. 2e, right panel and Extended Data Table 1e). These data demonstrate that the αC-helix of H2A.Z is essential for the *in vivo* interaction of ANP32E with H2A.Z and further indicate that ANP32E is the main link between H2A.Z and the p400/TIP60 complex.

Structural basis for H2A.Z recognition

All the data presented above provide evidence that ANP32E is an H2A.Z-specific chaperone. We next investigated the molecular basis for ANP32E specificity and function using X-ray crystallography.

Crystals were obtained from the complex formed by the ANP32E-ZID (amino acids 215–240) and H2A.Z and H2B deleted from their N-terminal unstructured tails (amino acids 18–127 and 30–125, respectively). These crystals diffracted to 1.48 Å resolution and the structure of the complex was solved by molecular replacement (Extended Data Table 1f). The structure showed that the ANP32E-ZID interacts with one tip of the H2A.Z/H2B dimer (Fig. 3a). The ANP32E-ZID has no defined secondary structure elements, except for a small N-terminal α-helix (αN). Strikingly, in the ANP32E-ZID–H2A.Z/H2B structure, the H2A.Z αC-helix is extended twice compared with its canonical length in the nucleosome (Fig. 3a–d).

This H2A.Z αC conformational change impedes the remaining part of the docking domain, which mediates the main nucleosomal interactions of H2A and H2A.Z with the H3/H4 pair[22,43], from adopting the conformation observed in the nucleosome (Fig. 3a–c, e and Extended Data Fig. 4). Specifically, the residues that participate in the extended part of the H2A.Z αC-helix do not fold back anymore onto the H2A.Z α3-helix and do not interact with the H4 C terminus. Instead, this position is now occupied by the ANP32E-ZID αN-helix (Fig. 3e and Extended Data Fig. 4).

The ANP32E-ZID binding seems to be the driving force for H2A.Z αC conformational change. Consistent with this, upon H2A.Z binding the ANP32E-ZID αN-helix provides three residues (L218, L221 and M222) that form a network of hydrophobic interactions and salt bridges with residues of helices α3 and αC of H2A.Z as well as residues of helix α2 of H2B (Extended Data Fig. 5a).

Residues belonging to the extended part of H2A.Z αC, in particular T103 and I104, contribute to these interactions and stabilize this extended conformation (Fig. 3d and Extended Data Fig. 5a). Surprisingly, these residues are fully conserved in H2A (T101 and I102), indicating that this canonical histone could form the same extended αC-helix (Fig. 3d). However, a main difference between H2A and H2A.Z is the presence of an extra conserved glycine (H2A G98) two residues before the conserved threonine–isoleucine motif (Fig. 3d). Apart from its helix-breaker propensity, this glycine would shift by one residue the

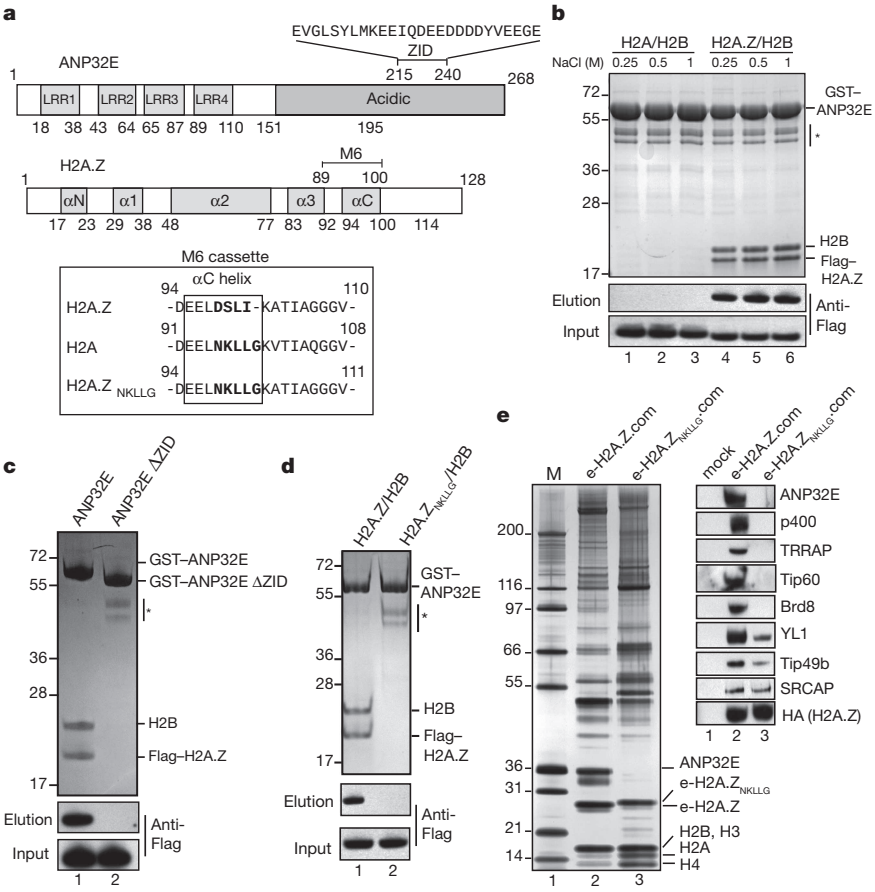


Figure 2 | The ANP32E ZID domain interacts with H2A.Z αC-helix. **a**, Upper panel, ANP32E schematic view. LRR, leucine-rich repeat domain. ZID, H2A.Z interaction domain. Middle panel, H2A.Z schematic view. Lower panel, M6 cassette sequences (with H2A.Z αC-helix boxed) of H2A.Z, H2A and the H2A.Z_{NKLLG} swap mutant. **b**, GST–ANP32E pull-down assays. Upper panel, ANP32E specifically interacts with H2A.Z/H2B but not with H2A/H2B dimers. Lower panels, immunoblotting using anti-Flag antibody of Flag-tagged H2A and H2A.Z. **c**, Same as in **b**, using GST–ANP32E full-length or lacking the ZID domain (GST–ANP32EΔZID). **d**, Same as in **b**, using H2A.Z/H2B and H2A.Z_{NKLLG}/H2B dimers as substrates. **e**, Analysis of e-H2A.Z and e-H2A.Z_{NKLLG} complexes by silver staining (left panel) and by immunoblotting (right panel). Stars, degradation products. All data are representative of $n \geq 3$ for each experiment.

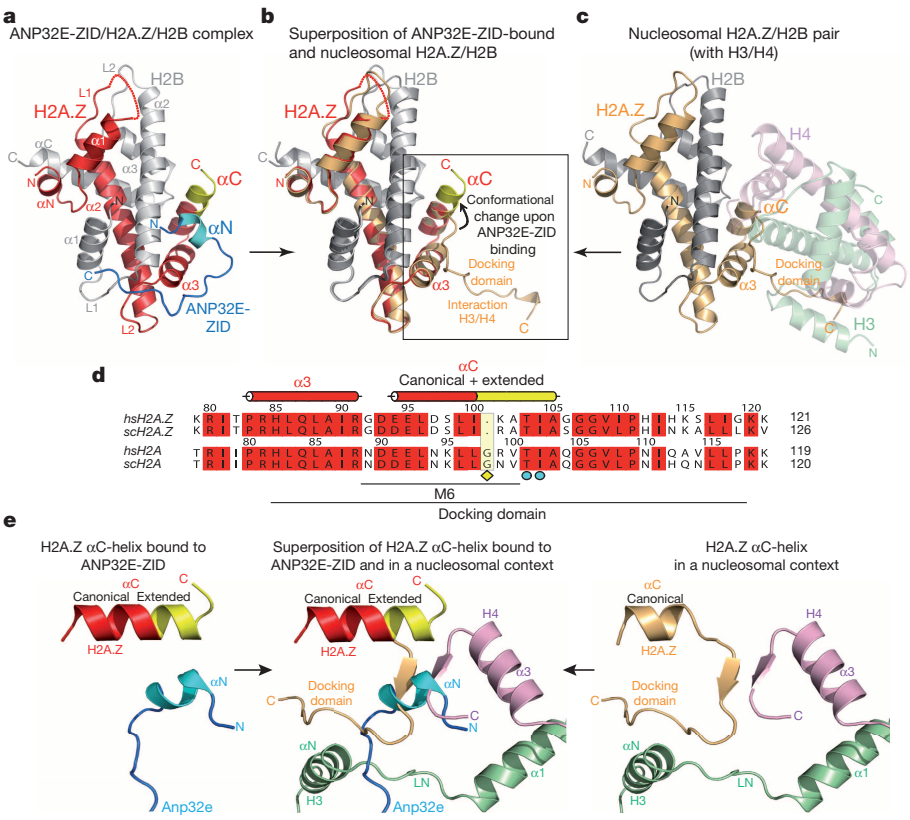


Figure 3 | Specific recognition of the H2A.Z/H2B pair by ANP32E. **a**–**c**, Structural comparison (**b**) of the H2A.Z/H2B pair bound to the ANP32E-ZID (**a**) with the same pair in a nucleosomal context (**c**). ANP32E-ZID binding induces a main conformational change of H2A.Z αC-helix whose length is doubled (**a**, **b**), thus preventing the C-terminal docking domain of H2A.Z to form extensive interactions with a H3/H4 nucleosomal pair (**c**). **d**, Alignment of human and yeast H2A and H2A.Z docking domains highlighting the specific residues forming the extended part of H2A.Z αC-helix. **e**, Model for H2A.Z/H2B nucleosomal eviction by ANP32E. The ANP32E mediated H2A.Z αC-helix extension (left panel) results in several steric hindrances at the H3/H4 interface (middle panel) that are incompatible with stable H2A.Z/H2B binding to the nucleosome (right panel).

position of T101 and I102 in an extended H2A αC-helix and would prevent these residues from forming the network of interactions observed in our structure.

Thus, binding of the ANP32E-ZID N-terminal helix to H2A.Z/H2B would enable the specific recognition of this dimer over the H2A/H2B pair by creating and stabilizing a conformational change in H2A.Z that cannot be accommodated by H2A. Accordingly, insertion of a glycine in H2A.Z at the equivalent position of H2A G98 completely abrogates the recognition of the H2A.Z/H2B pair by ANP32E (Extended Data Fig. 6a). The other H2A histone variant αC-helices also have an extra residue before their conserved threonine–isoleucine motif, suggesting that ANP32E also discriminates H2A.Z from the other H2A variants (Extended Data Fig. 6b).

The C-terminal part of the ANP32E-ZID also forms strong interactions with the H2A.Z/H2B pair at the surface formed by the H2A.Z loop L2 and the H2B α1-loop L1-α2 region (Extended Data Fig. 5b, c). Specifically, the ANP32E-ZID main chain and side chains of D232 and D234 form a hydrogen bond network with residues from H2A.Z and H2B, whereas Y235 is involved in hydrophobic interactions with H2B residues (Extended Data Fig. 5b, c). These interactions do not allow further discrimination of H2A.Z over H2A, but shield the H2A.Z loop L2/H2B loop L1 region that is known to interact with DNA close to the entry/exit sites of the nucleosome.

To investigate the importance of the ANP32E-ZID for the functional role of ANP32E, three mutants were designed in the context of the full-length protein, targeting both H2A.Z/H2B binding regions of the ANP32E-ZID (ANP32E-m1, L218A/L221A/M222A; ANP32E-m2, D232A/D234A/Y235A; ANP32E-m1m2, L218A/L221A/M222A/D232A/D234A/Y235A). Pull-down experiments showed that all three mutants lost their ability to interact with the H2A.Z/H2B dimer (Extended Data Fig. 6c). We next expressed the ANP32E-m1m2 mutant in HeLa cells and purified the complex it forms (e-ANP32E-m1m2.com). In contrast to wild-type ANP32E, ANP32E-m1m2 was unable to bind the H2A.Z/H2B dimer but did not affect the incorporation of ANP32E to the rest of the p400/Tip60 complex (Extended Data Fig. 6d, e).

H2A.Z/H2B nucleosomal removal by ANP32E

Collectively, our structural data suggest that ANP32E is able specifically to remove H2A.Z from the nucleosome. To analyse this ability of ANP32E directly, we performed *in vitro* experiments. Conventional and H2A.Z nucleosomes were assembled on 256 base pair (bp) 5S DNA, immobilized on magnetic beads and incubated with increasing amounts of recombinant ANP32E. The bound protein/DNA complexes were purified using the magnetic particle concentrator and the unbound material was collected. Bound and unbound material were run on SDS–PAGE, transferred to PVD membrane and blotted either with anti-H2A or anti-H2A.Z antibodies. The incubation of ANP32E with the H2A.Z immobilized nucleosomes resulted in the dissociation of H2A.Z from the nucleosome (Fig. 4a, left and right panels).

To get more insight in the mechanism of H2A.Z removal from the nucleosome, conventional and H2A.Z nucleosomes were reconstituted on a mini-circle DNA[44] and both samples were incubated with increasing amounts of highly purified ANP32E. As seen in Fig. 4b (left panel), the presence of ANP32E in the H2A.Z nucleosome reaction mixture resulted in the appearance of $(H3/H4)_2$ tetramer particle (tetrasome), the amounts of which increased upon increasing amounts of ANP32E. We attributed this effect to the ANP32E-induced eviction of the H2A.Z/H2B dimer from the H2A.Z nucleosome. Accordingly, two-dimensional SDS–PAGE analyses showed that the nucleosome-like particle with lower mobility generated by the treatment with ANP32E only contains H3 and H4 (Fig. 4b, right panel).

No such effect was, however, observed when either conventional H2A nucleosomes or nucleosomes reconstituted with the H2A.Z-$_{NKLLG}$ mutant were used as substrates (Fig. 4b, left and right panels). Note that the ANP32E-induced removal of H2A.Z is not affected by competitor DNA (Extended Data Fig. 7a). The conventional histone chaperone

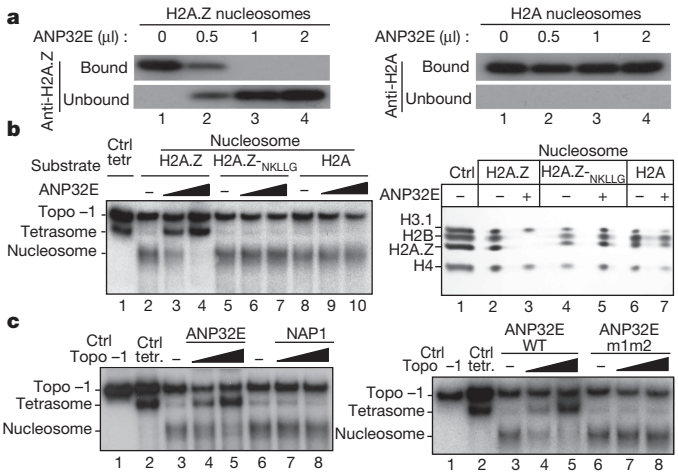


Figure 4 | Specific removal of H2A.Z from the nucleosome by ANP32E. **a**, Effect of increasing amount of ANP32E on bead-immobilized H2A.Z (left panel) or H2A (right panel) nucleosomes. **b**, Effects of increasing amount of ANP32E on H2A.Z, H2A.Z$_{NKLLG}$ and conventional H2A nucleosomes reconstituted on negatively supercoiled human α-satellite 360 bp DNA minicircle corresponding to topoisomer −1 (Topo −1); left panel, native PAGE; right panel, two-dimensional SDS–PAGE. **c**, Effects of increasing amount of either ANP32E (wild type (WT) or mutant) or NAP1 on the stability of H2A.Z nucleosomes. Only wild-type ANP32E promotes H2A.Z eviction. Tetr., tetrasome.

NAP-1 and the ANP32E-m1m2 mutant were also unable to evict H2A.Z/H2B dimers from the nucleosome (Fig. 4c, left and right panels). Thus, ANP32E, through its specific H2A.Z interaction domain, is able to remove H2A.Z directly from the nucleosome particle. ANP32E, in contrast to Nap-1, was unable to deposit H2A.Z/H2B or H2A/H2B dimers on reconstituted $(H3/H4)_2$ tetrasome particles (Extended Data Fig. 7b, c). In addition, ANP32E.com was not capable of exchanging H2A for H2A.Z in the nucleosome (Extended Data Fig. 7d, e). We conclude that ANP32E is involved in the specific removal of H2A.Z, but not in its deposition, and possibly stabilizes the evicted H2A.Z/H2B dimer, thus shifting the equilibrium towards dissociation.

H2A.Z ChIP–seq profile in ANP32E$^{-/-}$ cells

If ANP32E, as strongly suggested by our data, is a histone chaperone implicated in the removal of H2A.Z *in vivo*, its depletion would result in accumulation of H2A.Z in chromatin. To study the effect of ANP32E depletion on H2A.Z localization *in vivo*, we performed a genome-wide comparative ChIP-seq analysis using both ANP32E$^{+/+}$ (wild type) and ANP32E$^{-/-}$ (knockout) mouse embryonic fibroblasts (MEFs). Our data show that the absence of ANP32E results in the appearance of a significant number of new H2A.Z binding sites (4245 new knockout-specific peaks) at distal elements from promoters (distance to nearest transcription start site (TSS) >2 kilobases; Figs 5a, b).

Interestingly, although no new H2A.Z peaks were observed at promoters in knockout cells (Figs 5c–e), the net amount of promoter-associated H2A.Z increased by 20% and showed similar distribution around TSS compared with wild-type cells (Fig. 5f and Extended Data Fig. 8a, b). As observed previously[5–12,45–47], two histone H2A.Z–containing nucleosomes are positioned on either side of the TSS (Fig. 5g and Extended Data Fig. 8b) and H2A.Z was primarily found on CpG-containing promoters (Extended Data Fig. 8c).

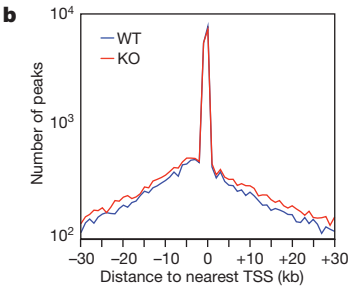
To map the genomic distribution of the distal knockout-specific new peaks, we compared their distribution to the *cis*-regulatory sequences identified in MEF cells by the Encode project[48]. We were able to map the new H2A.Z peaks to two distinct functional regions: (1) enhancer regions (peaks outside promoters, lacking H3K4me3) that correlated with the presence of either H3K4me1 or H3K27ac marks, or both (Fig. 5d), and (2) insulator regions (CTCF-binding sites that do not overlap with either promoters or enhancers) (Fig. 5e). Notably, 42% of the new H2A.Z binding sites overlapped with enhancers and 11% with

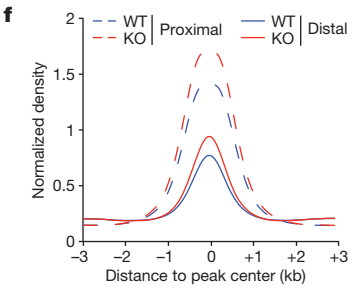
a

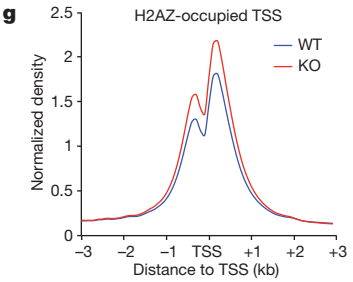
Distance to nearest TSS	Proximal		Distal		Total	
	< 2 kb		> 2 kb			
MEFs	WT	KO	WT	KO	WT	KO
	14,058	14,093	23,609	27,819	37,667	41,912

c

	H2A.Z peaks		Overlapping with			
	WT	KO	H3K4me3	H3K4me1 and/or H3K27ac	CTCF	KO-specific peaks
Promoters	16,914	16,854	+			
Enhancers	11,601	13,391	–	+		42%
Insulators	2,116	2,563	–	–	+	11%
Others	7,036	9,104	–	–	–	47%
Total	37,667	41,912				







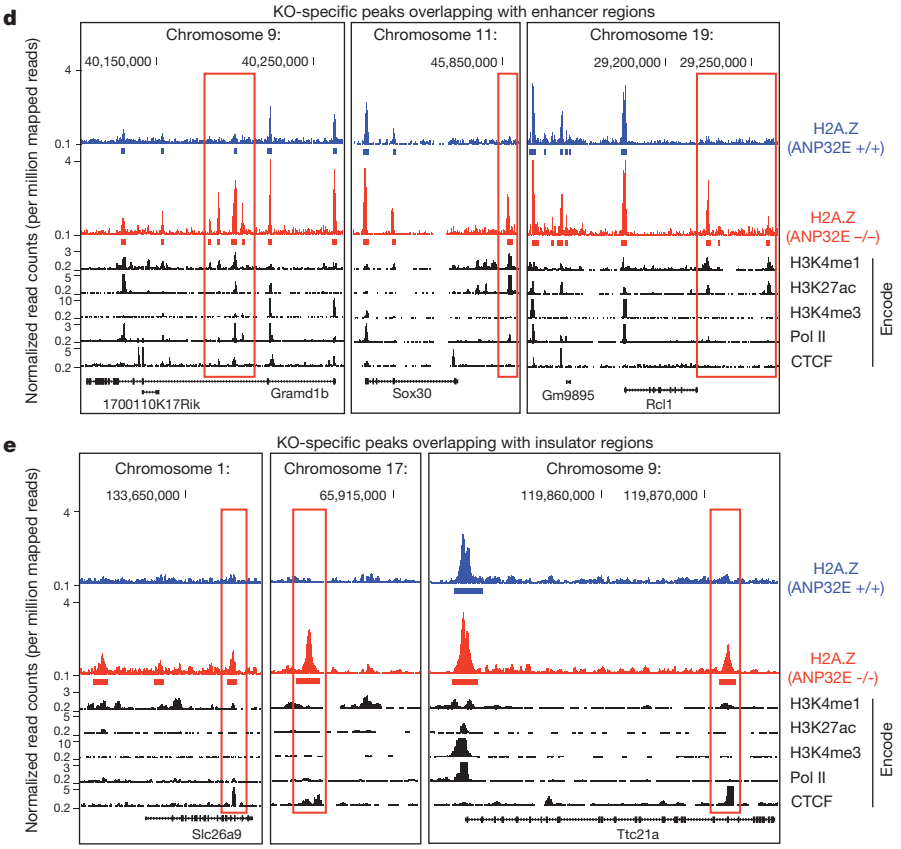


Figure 5 | Genomic localization and chromatin enrichment of H2A.Z is dependent on ANP32E. **a**, Number of H2A.Z peaks observed in ANP32E wild type (WT) and knockout (KO) MEFs (significant differences highlighted in red). **b**, Peak distribution of H2A.Z relative to TSS for ANP32E WT (blue) and knockout (red) in MEFs. **c**, Genomic location of H2A.Z peaks. Table shows overlap of H2A.Z peaks with those of H3K4me1, H3K27ac, H3K4me3 and CTCF (data from ENCODE consortium). **d**, **e**, Genome browser views indicating ectopic presence of H2A.Z in ANP32E-knockout MEFs compared with WT MEFs at enhancers (**d**) and insulators (**e**). **f**, Normalized density of H2A.Z in ANP32E WT and knockout MEFs. **g**, Density of H2A.Z deposition across TSS in ANP32E knockout MEFs (red) compared with WT (blue).

insulators (Fig. 5c–e). These data strongly implicate ANP32E as an H2A.Z chaperone specialized in the genome-wide removal of H2A.Z from its normal sites of deposition and in particular from enhancer and insulator regions. This suggests that at least part of the remaining non-annotated H2A.Z binding sites in MEFs depleted for ANP32E (47% of the knockout-specific peaks) could also show regulatory features.

In conclusion, we show here that human ANP32E is an H2A.Z-specific chaperone able to remove H2A.Z *in vitro* and to promote establishment of H2A.Z depleted chromatin loci. Our data show that the ANP32E-ZID region is essential for H2A.Z-specific recognition and destabilization of H2A.Z/H2B binding to nucleosomes, notably through H2A.Z αC-helix extension. The absence of ANP32E is associated with increased accumulation of H2A.Z around the TSS as well as at other chromatin regions. Absence of ANP32E leads to H2A.Z incorporation into enhancers and insulators, which are otherwise normally depleted of H2A.Z. These data demonstrate that ANP32E regulates the specific H2A.Z genome-wide localization pattern. Nevertheless, unlike H2A.Z knockout mice that are embryonically lethal[3], mice lacking ANP32E have no notable phenotype[49] indicating the presence of other chaperones that also function independently to deplete H2A.Z from chromatin.

On the basis of our data, we propose the following mechanism for the ANP32E assisted removal of H2A.Z/H2B from the nucleosomes (Fig. 3e and Extended Data Fig. 9). Initial recognition of the H2A.Z loop L2/H2B loop L1 region by the ANP32E-ZID DNA shielding domain leads to the destabilization of H2A.Z/H2B binding to DNA at the nucleosomal entry/exit sites. This destabilization is further increased by the central acidic linker of the ANP32E-ZID, which tends to occupy the space of the LN region of H3 and the C-terminal part of the docking domain, thus altering the interaction of H3 with the H2A.Z docking domain. As a result, the H2A.Z αC-helix becomes accessible to the ANP32E-ZID αN. The positioning of this latter helix as well as the H2A.Z αC-helix extension leads to large steric clashes and subsequent eviction of the H2A.Z/H2B dimer. This situation implies initial accessibility of the H2A.Z loop L2/H2B loop L1 region to ANP32E, which could be achieved thanks to nucleosome end breathing[50]. *In vivo*, however, p400/TIP60-specific ANP32E targeting and H2A.Z nucleosome remodelling should also greatly facilitate this mechanism and H2A.Z/H2B eviction. Once ANP32E evicts H2A.Z, it could also potentially stabilize the released H2A.Z/H2B dimer and thus shift the equilibrium towards dissociation and the off-chromatin state.

METHODS SUMMARY

e-H2A, e-H2A.Z and e-ANP32E nuclear complexes and their respective mutants were purified by double immunoaffinity as previously described[34,35]. Crystals of

the complex formed by human ANP32E (residues 215–240), H2A.Z (residues 18–127) and H2B (residues 30–125) were obtained by vapour diffusion, and crystallographic data on this complex were collected at 1.48 Å resolution. ChIP experiments were performed as described in Methods. ChIP-seq was performed on the Illumina Hiseq 2500 as single-end 50 base reads following Illumina's instructions. Reads were mapped onto the mm9 assembly of the mouse genome.

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 2 August 2012; accepted 28 November 2013.
Published online 22 January 2014.

1. Iouzalen, N., Moreau, J. & Mechali, M. H2A.ZI, a new variant histone expressed during *Xenopus* early development exhibits several distinct features from the core histone H2A. *Nucleic Acids Res.* **24,** 3947–3952 (1996).
2. Clarkson, M. J., Wells, J. R., Gibson, F., Saint, R. & Tremethick, D. J. Regions of variant histone His2AvD required for *Drosophila* development. *Nature* **399,** 694–697 (1999).
3. Faast, R. *et al.* Histone variant H2A.Z is required for early mammalian development. *Curr. Biol.* **11,** 1183–1187 (2001).
4. Billon, P. & Cote, J. Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance. *Biochim. Biophys. Acta* **1819,** 290–302 (2012).
5. Mavrich, T. N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18,** 1073–1083 (2008).
6. Raisner, R. M. *et al.* Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. *Cell* **123,** 233–248 (2005).
7. Zhang, H., Roberts, D. N. & Cairns, B. R. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123,** 219–231 (2005).
8. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
9. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456,** 125–129 (2008).
10. Weber, C. M., Henikoff, J. G. & Henikoff, S. H2A.Z nucleosomes enriched over active genes are homotypic. *Nature Struct. Mol. Biol.* **17,** 1500–1507 (2010).
11. Conerly, M. L. *et al.* Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Res.* **20,** 1383–1390 (2010).
12. Guillemette, B. *et al.* Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol.* **3,** e384 (2005).
13. Santisteban, M. S., Kalashnikova, T. & Smith, M. M. Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes. *Cell* **103,** 411–422 (2000).
14. Kalocsay, M., Hiller, N. J. & Jentsch, S. Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. *Mol. Cell* **33,** 335–343 (2009).
15. Rangasamy, D., Berven, L., Ridgway, P. & Tremethick, D. J. Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *EMBO J.* **22,** 1599–1607 (2003).
16. Rangasamy, D., Greaves, I. & Tremethick, D. J. RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. *Nature Struct. Mol. Biol.* **11,** 650–655 (2004).
17. Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C. & Tremethick, D. J. The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nature Struct. Biol.* **9,** 172–176 (2002).
18. Meneghini, M. D., Wu, M. & Madhani, H. D. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112,** 725–736 (2003).
19. Papamichos-Chronakis, M., Watanabe, S., Rando, O. J. & Peterson, C. L. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144,** 200–213 (2011).
20. van Attikum, H., Fritsch, O. & Gasser, S. M. Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks. *EMBO J.* **26,** 4113–4125 (2007).
21. Hua, S. *et al.* Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression. *Mol. Syst. Biol.* **4,** 188 (2008).
22. Suto, R. K., Clarkson, M. J., Tremethick, D. J. & Luger, K. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nature Struct. Biol.* **7,** 1121–1124 (2000).
23. Hamiche, A. & Shuaib, M. Chaperoning the histone H3 family. *Biochim. Biophys. Acta* **1819,** 230–237 (2012).
24. Henikoff, S. & Ahmad, K. Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* **21,** 133–153 (2005).
25. Luk, E. *et al.* Chz1, a nuclear chaperone for histone H2AZ. *Mol. Cell* **25,** 357–368 (2007).
26. Zhou, Z. *et al.* NMR structure of chaperone Chz1 complexed with histones H2A.Z-H2B. *Nature Struct. Mol. Biol.* **15,** 868–869 (2008).
27. Mizuguchi, G. *et al.* ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303,** 343–348 (2003).
28. Kobor, M. S. *et al.* A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biol.* **2,** E131 (2004).
29. Krogan, N. J. *et al.* A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Mol. Cell* **12,** 1565–1576 (2003).
30. Cai, Y. *et al.* The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. *J. Biol. Chem.* **280,** 13665–13670 (2005).
31. Ikura, T. *et al.* Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* **102,** 463–473 (2000).
32. Choi, J., Heo, K. & An, W. Cooperative action of TIP48 and TIP49 in H2A.Z exchange catalyzed by acetylation of nucleosomal H2A. *Nucleic Acids Res.* **37,** 5993–6007 (2009).
33. Ruhl, D. D. *et al.* Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. *Biochemistry* **45,** 5671–5677 (2006).
34. Nakatani, Y. & Ogryzko, V. Immunoaffinity purification of mammalian protein complexes. *Methods Enzymol.* **370,** 430–444 (2003).
35. Drane, P., Ouararhni, K., Depaux, A., Shuaib, M. & Hamiche, A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev.* **24,** 1253–1265 (2010).
36. Doyon, Y., Selleck, W., Lane, W. S., Tan, S. & Cote, J. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol. Cell. Biol.* **24,** 1884–1896 (2004).
37. Matilla, A. & Radrizzani, M. The Anp32 family of proteins containing leucine-rich repeats. *Cerebellum* **4,** 7–18 (2005).
38. Kular, R. K., Cvetanovic, M., Siferd, S., Kini, A. R. & Opal, P. Neuronal differentiation is regulated by leucine-rich acidic nuclear protein (LANP), a member of the inhibitor of histone acetyltransferase complex. *J. Biol. Chem.* **284,** 7783–7792 (2009).
39. Jiang, X. *et al.* Distinctive roles of PHAP proteins and prothymosin-α in a death regulatory pathway. *Science* **299,** 223–226 (2003).
40. Tochio, N. *et al.* Solution structure of histone chaperone ANP32B: interaction with core histones H3–H4 through its acidic concave domain. *J. Mol. Biol.* **401,** 97–114 (2010).
41. Jensen, K., Santisteban, M. S., Urekar, C. & Smith, M. M. Histone H2A.Z acid patch residues required for deposition and function. *Mol. Genet. Genom.* **285,** 287–296 (2011).
42. Wu, W. H. *et al.* Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nature Struct. Mol. Biol.* **12,** 1064–1071 (2005).
43. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–260 (1997).
44. Hamiche, A. & Richard-Foy, H. Characterization of specific nucleosomal states by use of selective substitution reagents in model octamer and tetramer structures. *Methods* **19,** 457–464 (1999).
45. Ku, M. *et al.* H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome Biol.* **13,** R85 (2012).
46. Kelly, T. K. *et al.* H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. *Mol. Cell* **39,** 901–911 (2010).
47. Nekrasov, M. *et al.* Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. *Nature Struct. Mol. Biol.* **19,** 1076–1083 (2012).
48. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).
49. Reilly, P. T. *et al.* Generation and characterization of the Anp32e-deficient mouse. *PLoS ONE* **5,** e13597 (2010).
50. Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nature Struct. Mol. Biol.* **12,** 46–53 (2005).

Acknowledgements We thank I. Davidson, P. Antony, J. Cavarelli and D. Moras for reading the manuscript, and V. Cura and A. McEwen for help during data collection. This work was supported by CNRS, INSERM, Université de Strasbourg, and by grants from INCa (INCa_4496 and INCa_4454), ANR (VariZome, ANR-12-BSV8-0018-01; Nucleoplat, NT09_476241), the Association pour la Recherche sur le Cancer, La Fondation pour la Recherche Médicale, La Ligue Nationale contre le Cancer (Equipe labellisée, to A.H. and S.D.), the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INSB-05-01) and by Instruct (ESFRI). A.O. acknowledges the Association pour la Recherche sur le Cancer for Financial support. K.P. was supported by la Ligue Nationale contre le Cancer.

Author Contributions A.H. conceived and supervised the project. A.O. and K.O. built constructs and performed complex purifications and chromatin assembly/eviction assays. C.R., M.L.D. and M.M. solved the ANP32E-ZID–H2A.Z/H2B structure. C.P. and K.P. performed ChIP-seq experiments. P.T.R. and T.W.M. provided the ANP32E$^{-/-}$ mice. I.S. and K.O. generated the ANP32E$^{-/-}$ MEF cells. C.P. and L.R. conducted all bioinformatics analyses. A.H., C.R. and S.D. designed experiments, analysed data and wrote the paper.

Author Information ChIP-Seq datasets have been deposited in GEO under accession number GSE51579. X-ray crystallographic coordinates and structure factor files have been deposited in Protein Data Bank under accession number 4CAY. 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 A.H. (hamiche@igbmc.fr), C.R. (romier@igbmc.fr) or S.D. (dimitrov@ujf-grenoble.fr).

LETTER

doi:10.1038/nature12955

A global cloud map of the nearest known brown dwarf

I. J. M. Crossfield[1], B. Biller[1,2], J. E. Schlieder[1], N. R. Deacon[1], M. Bonnefoy[1,3], D. Homeier[4], F. Allard[4], E. Buenzli[1], Th. Henning[1], W. Brandner[1], B. Goldman[1] & T. Kopytova[1,5]

Brown dwarfs—substellar bodies more massive than planets but not massive enough to initiate the sustained hydrogen fusion that powers self-luminous stars[1,2]—are born hot and slowly cool as they age. As they cool below about 2,300 kelvin, liquid or crystalline particles composed of calcium aluminates, silicates and iron condense into atmospheric 'dust'[3,4], which disappears at still cooler temperatures (around 1,300 kelvin)[5,6]. Models to explain this dust dispersal include both an abrupt sinking of the entire cloud deck into the deep, unobservable atmosphere[5,7] and breakup of the cloud into scattered patches[6,8] (as seen on Jupiter and Saturn[9]). However, hitherto observations of brown dwarfs have been limited to globally integrated measurements[10], which can reveal surface inhomogeneities but cannot unambiguously resolve surface features[11]. Here we report a two-dimensional map of a brown dwarf's surface that allows identification of large-scale bright and dark features, indicative of patchy clouds. Monitoring suggests that the characteristic timescale for the evolution of global weather patterns is approximately one day.

The recent discovery of the Luhman 16AB system (also called WISE J104915.57-531906.1AB; ref. 12) revealed two brown dwarfs only two parsecs away, making these the closest objects to the Solar System after the Alpha Centauri system and Barnard's star. Both of these newly discovered brown dwarfs are near the dust-clearing temperature[13,14], and one (Luhman 16B) exhibits strong temporal variability of its thermal radiation consistent with a rotation period of 4.9 hours (ref. 15). Luhman 16AB's proximity to Earth makes these brown dwarfs the first substellar objects bright enough to be studied at high precision and high spectral resolution on short timescales. We observed both of the brown dwarfs for five hours (about one rotation period of Luhman 16B) using the CRyogenic high-resolution InfraRed Echelle Spectrograph[16] (CRIRES) at the European Southern Observatory's Very Large Telescope to search for spectroscopic variability.

Absorption features from CO and H_2O dominate the spectra of the brown dwarfs, as shown in Fig. 1. The two objects have similar spectra but the absorption lines are broader for the B component: it exhibits a projected equatorial rotational velocity of $26.1 \pm 0.2\,\mathrm{km\,s^{-1}}$, versus $17.6 \pm 0.1\,\mathrm{km\,s^{-1}}$ for Luhman 16A. Taking Luhman 16B's rotation period[15] and considering that evolutionary models predict these objects to be 1.0 ± 0.2 times the radius of Jupiter[17], Luhman 16B's rotation axis must be inclined less than about 30 degrees from the plane of the sky; that is, we are viewing this brown dwarf nearly equator-on. If the axes of the two brown dwarfs are closely aligned (like those of the planets in our Solar System) then Luhman 16A rotates more slowly than Luhman 16B and the objects either formed with different initial angular momenta or experienced different accretion or spin-braking histories. Alternatively, if the two brown dwarfs have comparable rotation periods (as tentatively indicated by recent observations[18]) then the two components' rotation axes must be misaligned, which would imply either an initially aligned system (like the Solar System) that was subsequently perturbed or a primordial misalignment (in contrast to the close alignment more typically observed for pre-main-sequence stellar binaries[19]). Measuring Luhman 16A's rotation period is the best way to determine whether the axes of the brown dwarfs are currently aligned or misaligned.

Our data clearly show spectroscopic variability intrinsic to Luhman 16B, and this brown dwarf's rapid rotation allows us to produce the global surface map shown in Fig. 2 using Doppler imaging techniques[20,21]. This produces a map that shows a large, dark, mid-latitude region, a

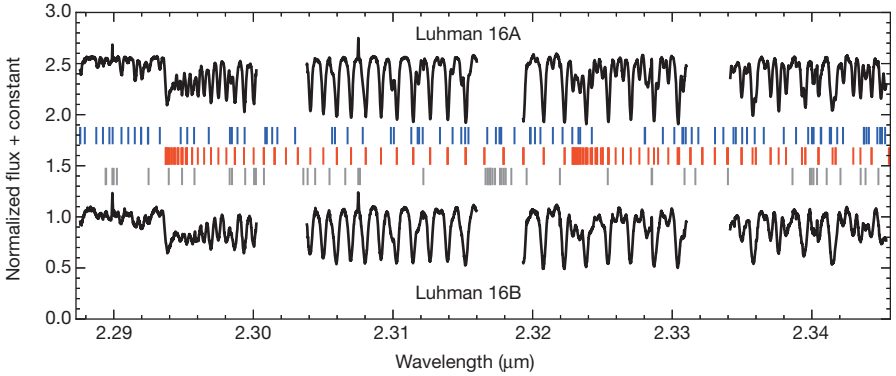


Figure 1 | High-resolution, near-infrared spectra of the Luhman 16AB brown dwarfs (black curves). All absorption features seem to be real and not artefacts: the vertical ticks indicate absorption features in the spectra of the brown dwarfs from H_2O (blue) and CO (red), and residual features from the Earth's atmospheric absorption (grey). The lines of the B component are broader, indicating a higher projected rotational velocity: thus either the brown dwarfs' rotation axes are misaligned or Luhman 16B formed with or developed a shorter rotation period than did its companion. The gaps in the spectra correspond to physical spaces between the four infrared array detectors. The plotted data represent the mean of all spectra. Luhman 16A's spectrum has been offset vertically for clarity by 1.5 units. The flux is normalized so that the continuum level (the flux level outside of absorption features, as seen, for example, in the far left regions of the spectra) is unity (1.0).

[1]Max Planck Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany. [2]Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK. [3]UJF-Grenoble 1/CNRS-INSU, Institut de Planétologie et d'Astrophysique de Grenoble (IPAG) UMR 5274, 38041 Grenoble, France. [4]CRAL-ENS, 46 Allée d'Italie, 69364 Lyon, Cedex 07, France. [5]International Max-Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg, Königstuhl 17, 69117 Heidelberg, Germany.

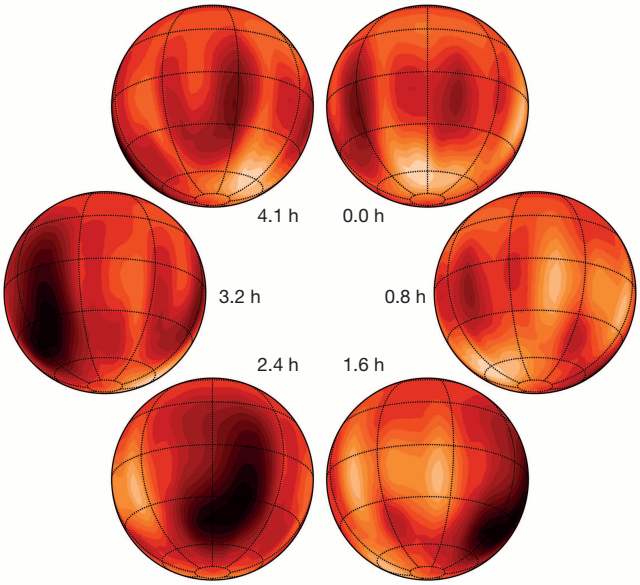


Figure 2 | Surface map of brown dwarf Luhman 16B. A bright near-polar region can clearly be seen in the upper-right panels. A darker mid-latitude area visible in the lower-left panels is consistent with large-scale cloud inhomogeneities. The lightest and darkest regions shown correspond to brightness variations of roughly 10%. The time index of each projection is indicated; the rotation period of the brown dwarf is 4.9 hours.

brighter area on the opposite hemisphere located close to the pole and mottling at equatorial latitudes.

A natural explanation for the features seen in our map of Luhman 16B is that we are directly mapping the patchy global clouds inferred to exist from observations of multiwavelength variability[15,18]. If this is true, the dark areas of our map represent thicker clouds that obscure deeper, hotter parts of the atmosphere and present a higher-altitude (and thus colder) emissive surface, whereas bright regions correspond to holes in the upper cloud layers that provide a view of the hotter, deeper interior. This result is also consistent with previous suggestions of multiple stratified cloud layers in brown dwarf atmospheres[4,10,11]. Because our mapping is mostly sensitive to CO, the map could in principle show a combination of surface brightness (that is, brightness temperature) and chemical abundance variations. Coupled models of global circulation and atmospheric chemistry[22], maps obtained via simultaneous observations of multiple molecular tracers and simultaneous Doppler imaging and broadband photometric monitoring could distinguish between these hypotheses.

The high-latitude bright spot could be similar to the polar vortices seen on Jupiter and Saturn and predicted to exist on highly irradiated gas giants in short-period orbits around other stars[23]; in this case, the high-latitude feature should still be visible in future maps of Luhman 16B. Jupiter and Saturn exhibit prominent circumplanetary banding, but (as described in the Methods) our analysis is not sufficiently sensitive to detect banding on Luhman 16B. Furthermore, assuming a mean horizontal windspeed of about 300 m s^{-1} (as predicted by global circulation models of brown dwarfs at these temperatures[24]) the Rhines relation[25] predicts that Luhman 16B should exhibit roughly ten bands from pole to pole—too many to resolve with our 18-degrees-wide map cells.

Long-term monitoring of Luhman 16B suggests that its weather conditions change rapidly but remain at least partly coherent from one night to the next[15], a result that indicates that the characteristic timescale for evolution of global weather patterns is of the order of one day. In this case, successive full nights of Doppler imaging could observe the formation, evolution and breakup of global weather patterns—the first opportunity for such a study outside the Solar System. Such measurements would provide a new benchmark against which to compare global circulation models of dusty atmospheres[24,26], and could perhaps measure differential rotation in Luhman 16B's atmosphere[27].

Future mapping efforts should reveal whether we are mapping variations in temperature, cloud properties or atmospheric abundances: high-resolution spectrographs with broader wavelength coverage than CRIRES should provide better sensitivity and spatial resolution[28], perhaps sufficient to search for banded cloud structures. Instruments with broader wavelength coverage will also allow maps to be made at multiple wavelengths and using independent molecular tracers (for example, H_2O). In addition, a few other variable brown dwarfs may be bright enough for these techniques to be applied. Although the day sides of hot, short-period gas giant planets can also be mapped using occultations under favourable conditions[29], model degeneracies may prevent these efforts from achieving a spatial resolution comparable to that achievable with Doppler imaging[30]. Thus, Doppler imaging in general, and Luhman 16B in particular, represent the best opportunity to challenge and improve our current understanding of the processes that dominate the atmospheres of brown dwarfs and of giant extrasolar planets.

METHODS SUMMARY

We extract and calibrate our spectroscopic data using standard techniques (Extended Data Figs 1 and 2) and look for temporal changes in the mean spectral line profiles. Luhman 16B exhibits strong spectroscopic variability but we see no evidence for similar variations in our simultaneously acquired observations of Luhman 16A (Extended Data Fig. 3). A simplified analysis using a parameterized spot model verifies that our observations are consistent with rotationally induced variations (Extended Data Fig. 4). We then produce our global map of brown dwarf Luhman 16B using Doppler imaging.

The technique of Doppler imaging relies on the varying Doppler shifts across the face of a rotating object and has been widely used to map the inhomogeneous surfaces of many rapidly rotating stars[20,21]. As darker regions rotate across the visible face of the brown dwarf, the Doppler-broadened absorption line profiles exhibit deviations at the projected radial velocities of the darker areas. Features near the equator cause changes across the entire line profile and move across the full span of velocities; features at higher latitudes move more slowly, experience smaller Doppler shifts, and affect a narrower range of velocities.

Our modelling framework is based on that described in ref. 20. We break the brown dwarf's surface into a 10×20 grid, giving an effective equatorial cell size of roughly 20,000 km. The recovered maps do not change significantly if we use finer resolution. We verify our analysis by constructing a number of Doppler images using simulated data. These simulations demonstrate that we can robustly detect large, isolated features with strong brightness temperature contrasts (Extended Data Fig. 5) but that we are not sensitive to axially symmetric features such as zonal banding (Extended Data Fig. 6).

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 October; accepted 5 December 2013.

1. Kumar, S. S. The structure of stars of very low mass. *Astrophys. J.* **137,** 1121–1125 (1963).
2. Becklin, E. E. & Zuckerman, B. A low-temperature companion to a white dwarf star. *Nature* **336,** 656–658 (1988).
3. Lunine, J. I., Hubbard, W. B. & Marley, M. S. Evolution and infrared spectra of brown dwarfs. *Astrophys. J.* **310,** 238–260 (1986).
4. Kirkpatrick, J. D. *et al.* Dwarfs cooler than "M": the definition of spectral type "L" using discoveries from the 2 Micron All-Sky Survey (2MASS). *Astrophys. J.* **519,** 802–833 (1999).
5. Stephens, D. C. *et al.* The 0.8–14.5 μm spectra of mid-L to mid-T dwarfs: diagnostics of effective temperature, grain sedimentation, gas transport, and surface gravity. *Astrophys. J.* **702,** 154–170 (2009).
6. Burgasser, A. J. *et al.* Evidence of cloud disruption in the L/T dwarf transition. *Astrophys. J.* **571,** L151–L154 (2002).
7. Tsuji, T., Nakajima, T. & Yanagisawa, K. Dust in the photospheric environment. II. Effect on the near-infrared spectra of L and T dwarfs. *Astrophys. J.* **607,** 511–529 (2004).
8. Ackerman, A. S. & Marley, M. S. Precipitating condensation clouds in substellar atmospheres. *Astrophys. J.* **556,** 872–884 (2001).
9. Fletcher, L. N. *et al.* Retrievals of atmospheric variables on the gas giants from ground-based mid-infrared imaging. *Icarus* **200,** 154–175 (2009).
10. Buenzli, E. *et al.* Vertical atmospheric structure in a variable brown dwarf: pressure-dependent phase shifts in simultaneous Hubble Space Telescope-Spitzer light curves. *Astrophys. J.* **760,** L31–L36 (2012).

11. Apai, D. *et al.* HST spectral mapping of L/T transition brown dwarfs reveals cloud thickness variations. *Astrophys. J.* **768,** 121–136 (2013).
12. Luhman, K. L. Discovery of a binary brown dwarf at 2 pc from the Sun. *Astrophys. J.* **767,** L1–L6 (2013).
13. Kniazev, A. Y. *et al.* Characterization of the nearby L/T binary brown dwarf WISE J104915.57–531906.1 at 2 pc from the Sun. *Astrophys. J.* **770,** 124–128 (2013).
14. Burgasser, A. J., Sheppard, S. S. & Luhman, K. L. Resolved near-infrared spectroscopy of WISE J104915.57–531906.1AB: a flux-reversal binary at the L dwarf/T dwarf transition. *Astrophys. J.* **772,** 129–135 (2013).
15. Gillon, M. *et al.* Fast-evolving weather for the coolest of our two new substellar neighbours. *Astron. Astrophys.* **555,** L5–L8 (2013).
16. Käufl, H.-U. *et al.* CRIRES: a high-resolution infrared spectrograph for ESO's VLT. *Proc. SPIE* **5492,** 1218–1227 (2004).
17. Burrows, A., Heng, K. & Nampaisarn, T. The dependence of brown dwarf radii on atmospheric metallicity and clouds: theory and comparison with observations. *Astrophys. J.* **736,** 47–60 (2011).
18. Biller, B. A. *et al.* Weather on the nearest brown dwarfs: resolved simultaneous multi-wavelength variability monitoring of WISE J104915.57–531906.1AB. *Astrophys. J.* **778,** L10–L16 (2013).
19. Wheelwright, H. E., Vink, J. S., Oudmaijer, R. D. & Drew, J. E. On the alignment between the circumstellar disks and orbital planes of Herbig Ae/Be binary systems. *Astron. Astrophys.* **532,** A28 (2011).
20. Vogt, S. S., Penrod, G. D. & Hatzes, A. P. Doppler images of rotating stars using maximum entropy image reconstruction. *Astrophys. J.* **321,** 496–515 (1987).
21. Rice, J. B., Wehlau, W. H. & Khokhlova, V. L. Mapping stellar surfaces by Doppler imaging—technique and application. *Astron. Astrophys.* **208,** 179–188 (1989).
22. Agúndez, M. *et al.* The impact of atmospheric circulation on the chemistry of the hot Jupiter HD 209458b. *Astron. Astrophys.* **548,** A73 (2012).
23. Cho, J. Y.-K., Menou, K., Hansen, B. M. S. & Seager, S. The changing face of the extrasolar giant planet HD 209458b. *Astrophys. J.* **587,** L117–L120 (2003).
24. Showman, A. P. & Kaspi, Y. Atmospheric dynamics of brown dwarfs and directly imaged giant planets. *Astrophys. J.* **776,** 85–103 (2013).
25. Vasavada, A. R. & Showman, A. P. Jovian atmospheric dynamics: an update after Galileo and Cassini. *Rep. Prog. Phys.* **68,** 1935–1996 (2005).
26. Freytag, B., Allard, F., Ludwig, H.-G., Homeier, D. & Steffen, M. Radiation-hydrodynamics simulations of cool stellar and substellar atmospheres. *ASP Conf. Ser.* **448,** 855–862 (2011).
27. Collier Cameron, A. & Unruh, Y. C. Doppler Images of AB Doradus. *Mon. Not. R. Astron. Soc.* **269,** 814–836 (1995).
28. Artigau, É., Donati, J.-F. & Delfosse, X. Planet detection, magnetic field of protostars and brown dwarfs meteorology with SPIRou. *ASP Conf. Ser.* **448,** 771–778 (2011).
29. Majeau, C., Agol, E. & Cowan, N. B. A two-dimensional infrared map of the extrasolar planet HD 189733b. *Astrophys. J.* **747,** L20–L24 (2012).
30. de Wit, J., Gillon, M., Demory, B.-O. & Seager, S. Towards consistent mapping of distant worlds: secondary-eclipse scanning of the exoplanet HD 189733b. *Astron. Astrophys.* **548,** A128 (2012).

Acknowledgements We thank A. Hatzes for advice on Doppler imaging analysis, J. Smoker for helping to plan and execute our observations and E. Mills for help in designing figures. This work is based on Director's Discretionary observations made with the European Southern Observatory Telescopes at the Paranal Observatory under programme ID 291.C-5006(A); data are available in the European Southern Observatory Data Archive at http://archive.eso.org/. D.H. received support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement number 247060). M.B., D.H. and F.A. acknowledge support from the French National Research Agency (ANR) through project grant ANR10-BLANC0504-01. E.B. is supported by the Swiss National Science Foundation (SNSF). Atmosphere models have been computed on the Pôle Scientifique de Modélisation Numérique at the Ecole Normale Supérieure (ENS) de Lyon, and at the Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen in co-operation with the Institut für Astrophysik Göttingen. IRAF (the Image Reduction and Analysis Facility) is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under cooperative agreement with the National Science Foundation. PyRAF (Python-based IRAF) is a product of the Space Telescope Science Institute, which is operated by AURA for NASA. We also thank contributors to SciPy, Matplotlib, AstroPy and the Python programming language.

Author Contributions I.J.M.C. coordinated the project and conducted all analyses. I.J.M.C., B.B., J.E.S., N.R.D., M.B., W.B., B.G. and T.K. assisted in obtaining the spectroscopic observations. I.J.M.C., B.B., J.E.S., N.R.D., M.B., D.H., F.A., E.B., Th.H. and W.B. contributed to the manuscript. J.E.S. provided an independent analysis of the projected rotational velocity and radial velocity of both brown dwarfs. B.B. and N.R.D. provided advice on binary dynamics. D.H. and F.A. provided advice on brown dwarf atmospheric processes and the spectral models used for the data calibration, least squares deconvolution and Doppler imaging.

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 I.J.M.C. (ianc@mpia.de).

METHODS

Observation and data reduction. We observed the Luhman 16AB system for five hours with the Very Large Telescope/CRIRES[16] on 5 May 2013 Universal Time. Our spectra span wavelengths from 2.288–2.345 μm to cover the CO (3,1) and CO (2,0) bandheads. During our observations the spectrograph slit was aligned to the binary position angle, so that both brown dwarfs were observed contemporaneously. The telescope was nodded along the binary axis to subtract the emission from the infrared-bright sky using standard techniques. A small, random offset was applied to each nod position to mitigate bad detector pixels. Observing conditions were good: seeing was roughly 0.5″, humidity was <10%, air mass ranged from 1.2 to 1.6, and the Moon was down during our observations. We spatially resolved the two brown dwarfs and over five hours obtained 56 spectra with exposures of 300 s each. We calibrate the raw CRIRES data frames using the standard CRIRES esorex data reduction routines, combining spectra in sets of four to boost the signal to noise ratio. We extracted one-dimensional spectra from both brown dwarfs in each combined frame using the standard astronomical IRAF data analysis package.

We used these extracted, raw spectral data to measure precisely the physical parameters of both brown dwarfs. We followed past work using high-precision infrared spectroscopy and used a forward-modelling approach[31,32] to calibrate our data. This analysis method transforms high-resolution spectra of the telluric transmission[33] and of model brown dwarf atmospheres[34] into a simulated CRIRES spectrum using an appropriate set of instrumental and astrophysical parameters. We used a model whose free parameters are the radial velocity of the model brown dwarf, rotational broadening ($v\sin i$) and linear limb-darkening coefficients[35], two multiplicative scaling factors for the telluric and brown dwarf models, polynomial coefficients that convert pixel number into wavelength, and coefficients for a low-order polynomial with which to normalize the continuum. As weights in the fits we used the uncertainties reported by IRAF after scaling these so that the weighted sum of the residuals equals the number of data points. The effect of all this is to place all observations on a common wavelength scale, to remove the effects of variable telluric absorption and spectrograph slit losses (from slight guiding errors or changes in seeing) and to estimate the astrophysical parameters listed above. Extended Data Fig. 1 shows examples of the raw and modelled data in this approach, and all calibrated spectra are shown in Extended Data Fig. 2. For each brown dwarf, we took as uncertainties the standard deviation on the mean of the measurements from each of the 14 spectra.

We applied the above modelling approach using a wide range of model brown dwarf spectra from the BT-Settl library[34], a set of models computed with the PHOENIX code that spans effective temperatures ($T_{\rm eff}$) of 1,000–1,600 K and surface gravities ($\log_{10}g$) of 4.0–5.5. We performed a separate fit to data from each of the four CRIRES detectors and found that the same model does not always give the best fit to the data from all four detectors. Judging from the residuals to the fits (see Extended Data Fig. 1), this effect resulted from inaccuracies in both the adopted telluric spectrum and in the brown dwarf atmospheric models. Considering all these ambiguities, we found that the models with $\log_{10}g = 5.0$ and $T_{\rm eff} = 1{,}500$ and 1,450 K (for components A and B, respectively) gave the best fits to data from all four detectors. There is some degeneracy between temperature and surface gravity, with greater $T_{\rm eff}$ allowing somewhat higher $\log_{10}g$. Brown dwarf atmospheres have never before been tested at this level of precision and so we did not interpolate between models to improve marginally the quality of the fit. The effective temperatures estimated from our analysis moderately exceeded the values reported by previous studies[13,14], and we attribute this difference to the well-known phenomenon that the effective temperature estimated from fitting model spectra to the CO bandheads typically exceeds the temperature derived from integrating the broadband spectral energy distribution[36,37]. Comparison of future models to these data should be highly instructive in refining substellar atmospheric models. In the analyses that follow, we use the BT-Settl models with the parameters given above; using slightly different model parameters does not change our conclusions.

To conduct a Doppler imaging analysis properly we had to account for the radial velocity shift of the brown dwarfs. We measure radial velocities for the A and B components of 20.1 ± 0.5 km s^{-1} and 17.4 ± 0.5 km s^{-1}, respectively, relative to the Solar System barycentre; the uncertainties in these absolute measurements are dominated by systematic uncertainties in our instrument model. Although the radial velocities of Luhman 16A exhibited little internal scatter during our observations, we saw an anomalous deviation (lasting from 1.5 h to 3 h after the start of observations) of roughly 1 km s^{-1} in the radial velocity measurements of Luhman 16B. Assuming that the systematic effects in measuring radial velocities are common to our observations of both brown dwarfs, and examining only the spectra taken outside the time of anomalous radial velocities, we obtained a relative radial velocity between the components of 2,800 ± 50 m s^{-1}. This measurement is consistent with the orbital velocity expected between two old brown dwarfs in an orbit of a few decades[14], and indicates that it will eventually be possible to test brown dwarf evolutionary models[17] by measuring the individual component masses via a full three-dimensional orbital solution using the system's radial velocities and astrometry[38,39].

To enhance our sensitivity, we used the technique of least squares deconvolution[40] (LSD) to transform each spectrum into a single mean absorption line, with high signal-to-noise ratio. Deviations in the resulting mean line profiles are difficult to see with the unaided eye, but after subtraction of the night's mean line profile, variations were apparent. Extended Data Fig. 3 shows the resulting temporal evolution in the deviations from the global mean line profile: the rotational signature of Luhman 16B's inhomogeneous surface is clearly visible, dominated by rotation of a darker region into and then out of view. Hints of brighter regions are visible at other times. We found that the total absorption depth of the mean line profile decreased by about 4% during this period. No such coherent signatures are observable beyond Luhman 16B's projected rotational velocity of ±26.1 km s^{-1}, and we do not see any such time-variable phenomena in our simultaneously acquired spectra of Luhman 16A.

Spot modelling. To interpret our LSD line profiles, we first implemented a simple spot model similar to that used to interpret photometry of variable brown dwarfs[11]. This initial toy model assumes that Luhman 16B's surface is dominated by a single spot. We divided the surface into a grid, regularly spaced in latitude and longitude. A 10 × 20 grid (18 degrees across each cell) is sufficient for the analysis to converge. The spot was assumed to be circular and the remainder of the photosphere was assumed to have uniform surface brightness with a linear limb-darkening law. The free parameters are the brightness of the spot relative to the photosphere and the spot's radius, latitude and longitude. For a given set of parameters we generated a surface map with the specified surface brightness distribution. For each grid cell we then used the projected visible area and apparent flux, and the cell's rotational Doppler shift, to generate a set of rotationally broadened line profiles corresponding to the time of each observation. Each line profile was continuum-normalized, and the resulting set of simulated data is compared to the observed LSD line profiles.

To estimate the uncertainty on the spot parameters we used the emcee tool[41], which implements an affine-invariant Markov Chain Monte Carlo approach. We initialized 150 chains near a set of reasonable guess parameters (final results are insensitive to this guess) and run all chains for 1,500 steps. After this initial 'burn-in' phase the Markov chains were randomized and had lost any memory of their initial starting conditions; we discarded the initial steps and ran the chains for an additional 1,500 steps, afterwards verifying that they were well mixed both by examination of the autocorrelation of the individual chains and by visual inspection of the likelihood and parameter values of the chains. Extended Data Fig. 4 shows the resulting posterior distributions of the spot's latitude, radius and surface brightness assuming $i = 30$ degrees; the results do not change significantly for smaller values of i. In this model the dark spot lies ≤31 degrees from the equator, has a radius of 33 ± 7 degrees, and is 88 ± 3% as bright as the surrounding photosphere. This result implies a photometric variation of about 3%, consistent with the wide range of variability seen from this system[15,18]. However, such parametrized models typically exhibit strong degeneracies and tend not to lead to unique maps of the surface brightness distributions of brown dwarfs[11].

Doppler imaging. We constructed our Doppler imaging model as described by ref. 20 using the 10 × 20 grid and line profile simulation techniques described above. Our results did not change significantly if we increased the model's spatial resolution. Instead of an arbitrarily parametrized spot, in the Doppler imaging model there are 200 free parameters: the contributions to the line profile from each grid cell. Because there are approximately 35 pixels across each of 14 mean line profiles, we nominally had 490 constraints; thus the problem appears well posed and simple matrix techniques (for example, singular value decomposition) would seem to be sufficient. However, it has long been recognized that such an approach yields extremely noisy maps[20,21,42], often with nonphysical values (for example, negative surface brightness in some cells). Regularization was needed, so we used a maximum entropy approach[43] in which the merit function is $Q = \chi^2 - aS$ (where χ^2 has its usual meaning, S is the image entropy of the map's grid cells, and a is a hyperparameter that determines the balance between goodness of fit and entropy). We minimized Q using a standard multivariate optimizer, and we sped up convergence by calculating the analytical gradients of Q relative to the brightnesses of the map cells.

Our data are noisier, and the spectroscopic variations weaker, than in typical Doppler imaging analyses of stars, so we tuned a to minimize the appearance of spurious features while maintaining the fidelity of the resulting map. We did this by generating a number of synthetic surface maps, simulating their line profiles and adding Gaussian noise of the same amplitude as we found in our observed LSD line profiles, and minimizing Q for various choices of a. An example of one such simulation and recovery is shown in Extended Data Fig. 5, using the same value of a as in the analysis leading to Fig. 2. This analysis demonstrates that the prominent mid-latitude and polar features are probably real, whereas the lower-contrast equatorial features may be more affected by noise. The longitudinal elongation of equatorial

features is a known feature of Doppler imaging maps[44], so features near the equator may be narrower than they appear. The main features in our recovered map do not change for small variations in the Doppler imaging modelling parameters. Finally, we find that although we cannot yet directly measure i with the current data[45], our maps do not change much for expected values of i (0–30 degrees).

Zonal banding and brown dwarf line profiles. The detection of axisymmetric features (such as zonal banding) via Doppler Imaging is more challenging than the detection of features lacking such symmetry: the latter are easily seen via their time-variable effects on the line profiles, but the former can only be distinguished by discerning deviations of the mean line shape from the modelled profile. We ran a number of simulated Doppler imaging analyses on brown dwarfs with various levels of banding. Extended Data Fig. 6 shows one such example, which is typical insofar as it demonstrates our inability to recover even strong, large-scale zonal bands. Even if the band contrast were 100%, our simulations show that recovery of such features would be tentative only, given the current precision of our data. Future observations at higher precision should have greater sensitivity to such features, and these efforts will therefore become more susceptible to the spurious axisymmetric bands that can result from Doppler imaging analyses performed with inappropriate line profile shapes[21,27,46]. We therefore consider below possible sources of uncertainty in modelling the detailed line shapes probed by our analysis.

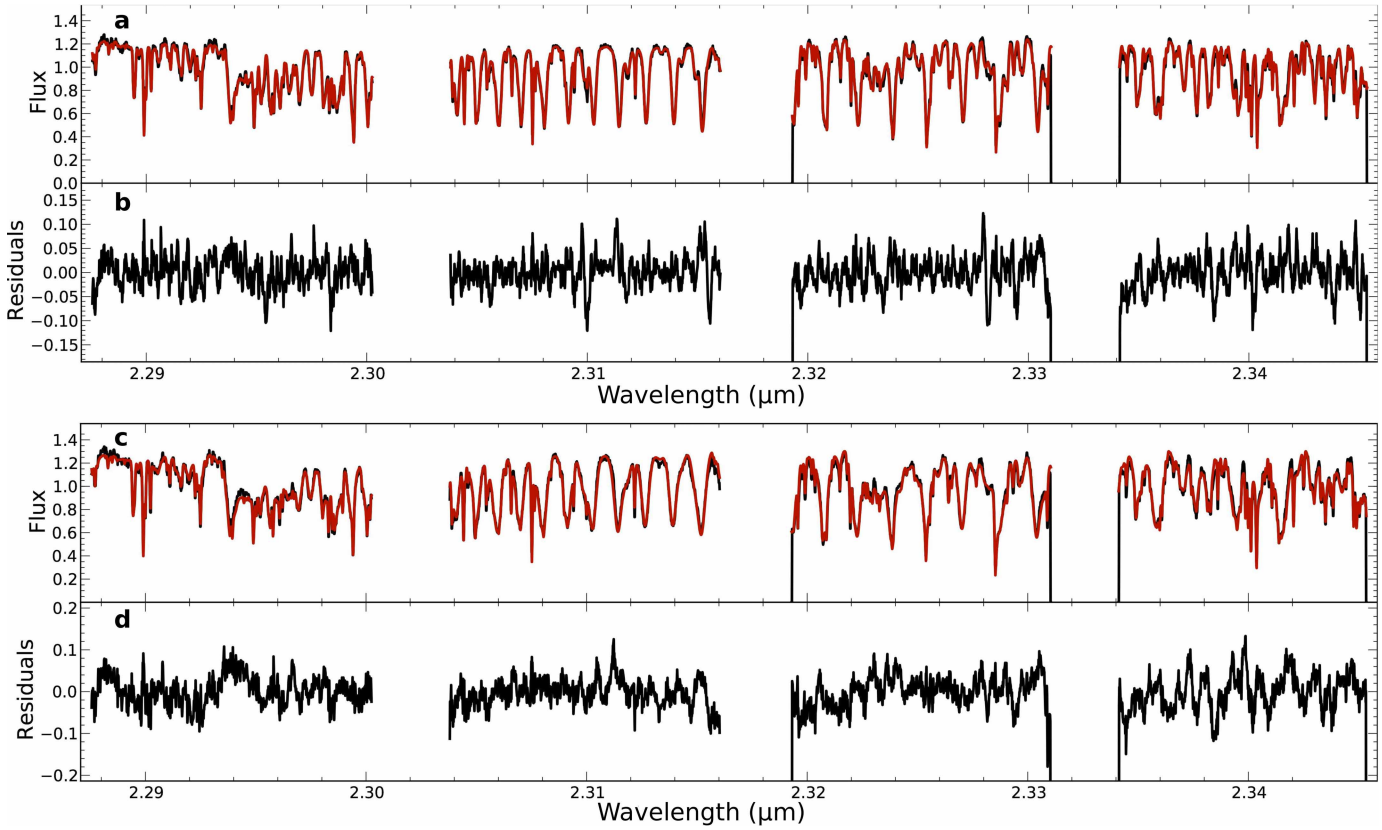
In the PHOENIX synthetic atmosphere and spectral model employed in our analysis, the strongest molecular lines are modelled as regular, symmetric Voigt profiles extending out to a maximum half width of 20 cm^{-1}. Beyond this detuning, effects of asymmetry and mixing of neighbouring lines no longer allow an adequate representation of the wings by a simple Lorentzian. A generic half width at half maximum (HWHM) of 0.08 $cm^{-1}\,bar^{-1}$ at 296 K and a temperature exponent of 0.5 was assumed for the Lorentz (pressure broadening) part of all molecular lines[34], and Doppler broadening was calculated for the thermal velocity plus an isotropic microturbulence of 0.8 km s^{-1}. The part of the spectrum covered by our observations formed mainly at pressure levels of 1–2 bar (ref. 18). Specifically, the wings of the strongest CO lines would form as deep as 4 bar, whereas the most central portion of the line cores formed as high as the 10 mbar level. The corresponding atmospheric temperatures in these layers (1,000–1,500 K) yielded a total Doppler broadening of the order of 1.1–1.25 km s^{-1}, whereas the half width of the collisional profile ranges from a few hundred metres per second in the cores up to nearly 10 km s^{-1} in the outer wings.

The true line profiles might deviate in several respects from the assumptions used in the PHOENIX model. Microturbulence, which in stellar atmosphere modelling simply denotes a random Gaussian velocity distribution on scales small compared to the photon mean free path, has not been constrained very tightly for brown dwarfs yet. A badly estimated microturbulence, particularly in the case of an anisotropic distribution with a stronger horizontal component, may affect the retrieval of surface features[21,27,46].

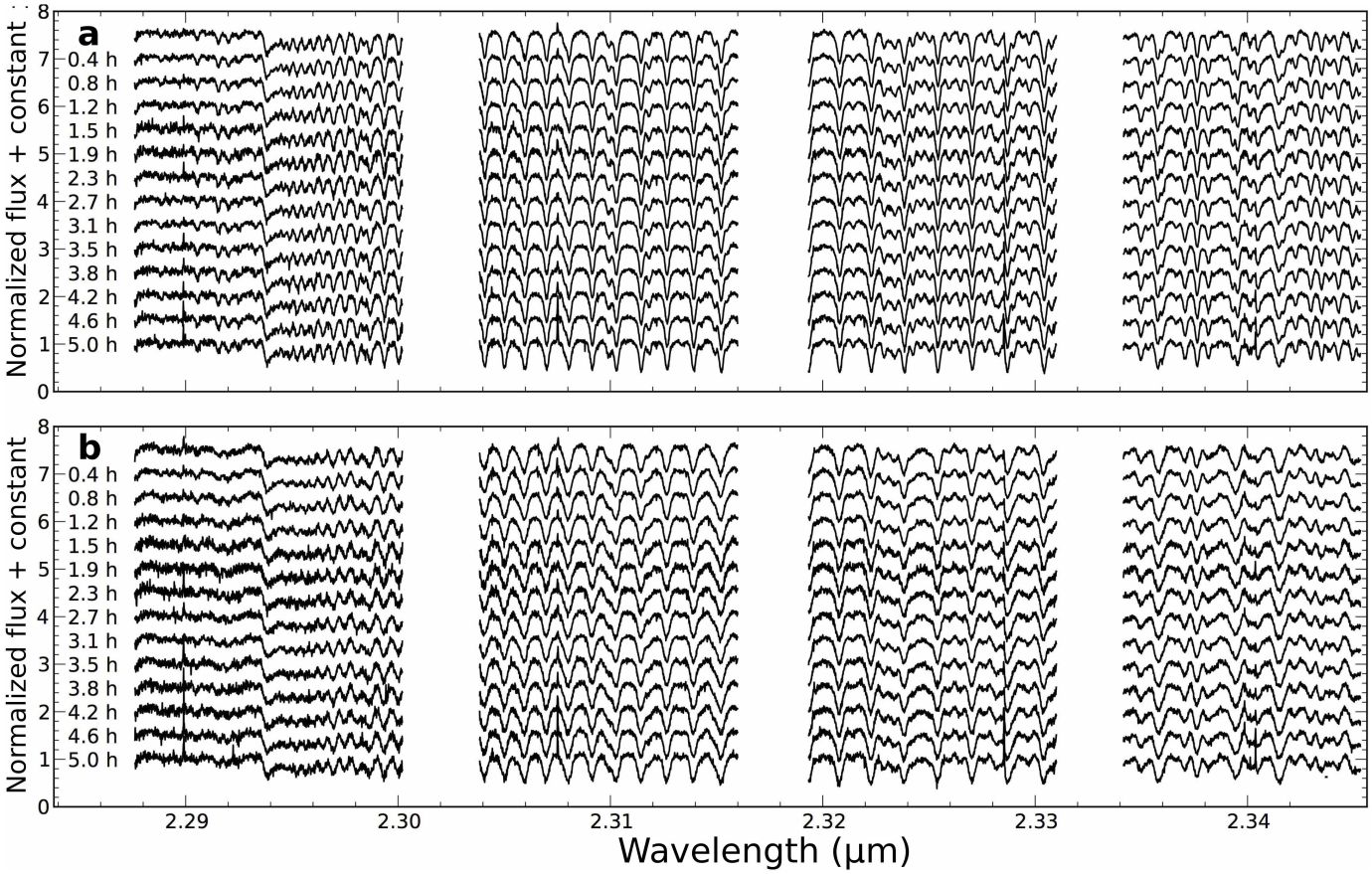
Radiation hydrodynamic simulations do allow us some insight into the dynamic structure of brown dwarf atmospheres, predicting horizontal root-mean-square velocities of the order of 0.3 km s^{-1} for our case, compared to values 3–5 times smaller for the vertical component[47]. However, unlike for typical stars mapped by Doppler imaging, in our case the total broadening is always dominated by the thermal velocity, so given the constant microturbulent velocity of 0.8 km s^{-1} of the PHOENIX models, any realistic changes are unlikely to have a noticeable impact on the line shapes. Pressure broadening of molecular lines, in contrast, has been poorly studied for stellar and substellar atmosphere conditions, that is, for temperatures of 1,000 K and higher and with molecular hydrogen (H_2) and helium (He) as main perturbers. Measurements of the broadening of CO lines in the fundamental band at 4.6 μm by noble gases and various other perturbers have yielded a HWHM of approximately 0.07 $cm^{-1}\,bar^{-1}$ at 296 K for H_2 (ref. 48). A study of the overtone band at 2.3 μm perturbed by various noble gases showed very similar widths to those in the fundamental[49], so it may be safe to assume that the H_2 broadening in this band is also comparable, and only about 12% smaller than our model value. The temperature dependence, however, could be stronger, with a possible temperature exponent of 0.5–0.75 (refs 50, 51). With all these effects combined, we may expect the actual damping widths to be up to a factor of two smaller than assumed in our model. On the other hand, the actual atmospheric conditions also remain poorly constrained, without a detailed spectral analysis or tighter limits on age and mass of the system. For an older and more massive brown dwarf, a surface gravity up to three times higher with correspondingly larger atmospheric pressures is possible, which would affect the collisional damping part of the line wings, but not the Doppler cores. Finally, collisional perturbations are also known to shift molecular lines. This effect, and in particular its temperature dependence, is even less well studied for perturbers other than H_2 and He[49,52], but the shifts should be around an order of magnitude smaller than the HWHM and thus have little effect on the position of the line cores.

In conclusion, it seems feasible that future observations at higher precision could determine whether Luhman 16B exhibits zonal banding. At present, our current data are not sufficiently sensitive to address this issue.

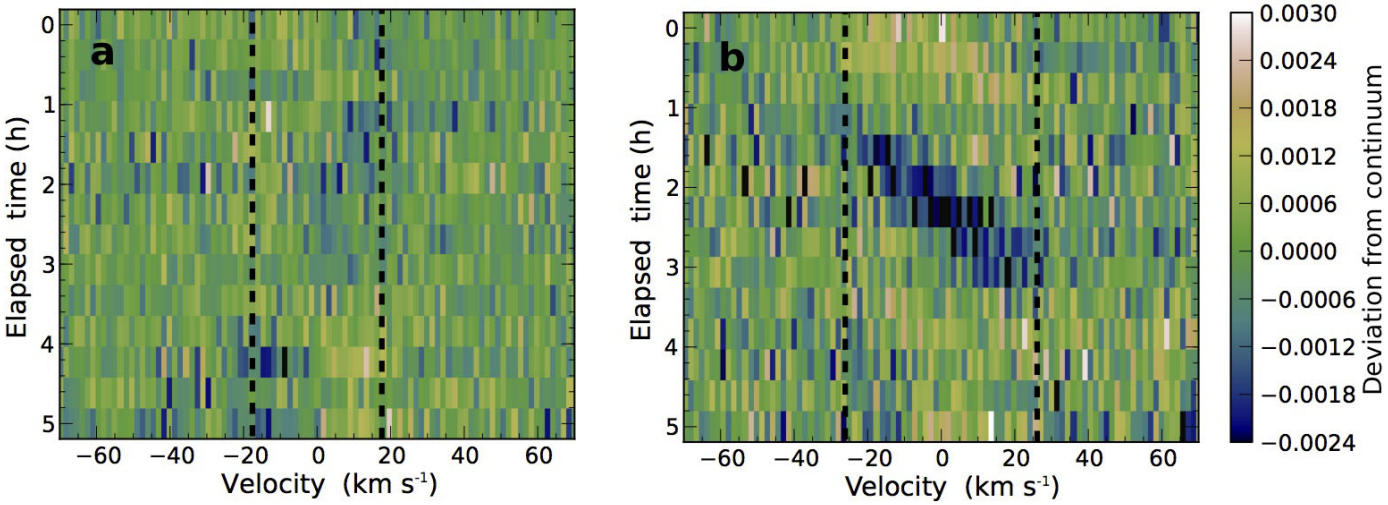
31. Blake, C. H., Charbonneau, D., White, R. J., Marley, M. S. & Saumon, D. Multiepoch radial velocity observations of L dwarfs. *Astrophys. J.* **666,** 1198–1204 (2007).
32. Bean, J. L. *et al.* The CRIRES search for planets around the lowest-mass stars. I. High-precision near-infrared radial velocities with an ammonia gas cell. *Astrophys. J.* **713,** 410–422 (2010).
33. Hinkle, K. H., Wallace, L. & Livingston, W. Atmospheric transmission above Kitt Peak, 0.5 to 5.5 microns. *Bull. Am. Astron. Soc.* **35,** 1260 (2003).
34. Allard, F., Homeier, D., Freytag, B., Schaffenberger, W. & Rajpurohit, A. S. Progress in modeling very low mass stars, brown dwarfs, and planetary mass objects. *Mem. Soc. Astron. Ital.* **24,** (Suppl.), 128–139 (2013).
35. Grey, D. F. *The Observation and Analysis of Stellar Photospheres* 3rd edn, Ch. 18 (Cambridge Univ. Press, 2005).
36. Jones, H. R. A. *et al.* Carbon monoxide in low-mass dwarf stars. *Mon. Not. R. Astron. Soc.* **358,** 105–112 (2005).
37. Schweitzer, A. *et al.* Effective temperatures of late L dwarfs and the onset of methane signatures. *Astrophys. J.* **566,** 435–441 (2002).
38. Konopacky, Q. M. *et al.* High-precision dynamical masses of very low mass binaries. *Astrophys. J.* **711,** 1087–1122 (2010).
39. Dupuy, T. J. & Liu, M. C. On the distribution of orbital eccentricities for very low-mass binaries. *Astrophys. J.* **733,** 122–135 (2011).
40. Donati, J.-F., Semel, M., Carter, B. D., Rees, D. E. & Collier Cameron, A. Spectropolarimetric observations of active stars. *Mon. Not. R. Astron. Soc.* **291,** 658–682 (1997).
41. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: The MCMC hammer. *Publ. Astron. Soc. Pacif.* **125,** 306–312 (2013).
42. Vogt, S. S. Doppler images of spotted late-type stars. In *The Impact of Very High S/N Spectroscopy on Stellar Physics* 253–272 (Proc. 132nd Symp. Int. Astron. Union, 1988).
43. Narayan, R. & Nityananda, R. Maximum entropy image restoration in astronomy. *Annu. Rev. Astron. Astrophys.* **24,** 127–170 (1986).
44. Rice, J. B. Doppler imaging of stellar surfaces—techniques and issues. *Astron. Nachr.* **323,** 220–235 (2002).
45. Rice, J. B. & Strassmeier, K. G. Doppler imaging from artificial data. Testing the temperature inversion from spectral-line profiles. *Astron. Astrophys.* **147** (Suppl.), 151–168 (2000).
46. Unruh, Y. C. & Collier Cameron, A. The sensitivity of Doppler imaging to line profile models. *Mon. Not. R. Astron. Soc.* **273,** 1–16 (1995).
47. Freytag, B., Allard, F., Ludwig, H.-G., Homeier, D. & Steffen, M. The role of convection, overshoot, and gravity waves for the transport of dust in M dwarf and brown dwarf atmospheres. *Astron. Astrophys.* **513,** A19 (2010).
48. Draegert, D. A. & Williams, D. Collisional broadening of CO absorption lines by foreign gases. *J. Opt. Soc. Am.* **58,** 1399–1403 (1968).
49. Bouanich, J. Détermination expérimentale des largeurs et des déplacements des raies de la bande 0 → 2 de co pertubé par les gaz rares (He, Ne, Ar, Kr, Xe). *J. Quant. Spec. Radiat. Transf.* **12,** 1609–1615 (1972).
50. Malathy Devi, V. *et al.* Spectral line parameters including temperature dependences of self- and air-broadening in the 2 → 0 band of CO at 2.3 micron. *J. Quant. Spec. Radiat. Transf.* **113,** 1013–1033 (2012).
51. Faure, A., Wiesenfeld, L., Drouin, B. J. & Tennyson, J. Pressure broadening of water and carbon monoxide transitions by molecular hydrogen at high temperatures. *J. Quant. Spec. Radiat. Transf.* **116,** 79–86 (2013).
52. Predoi-Cross, A., Bouanich, J. P., Benner, D. C., May, A. D. & Drummond, J. R. Broadening, shifting, and line asymmetries in the 2 ← 0 band of CO and CO–N_2: experimental results and theoretical calculations. *J. Chem. Phys.* **113,** 158–168 (2000).



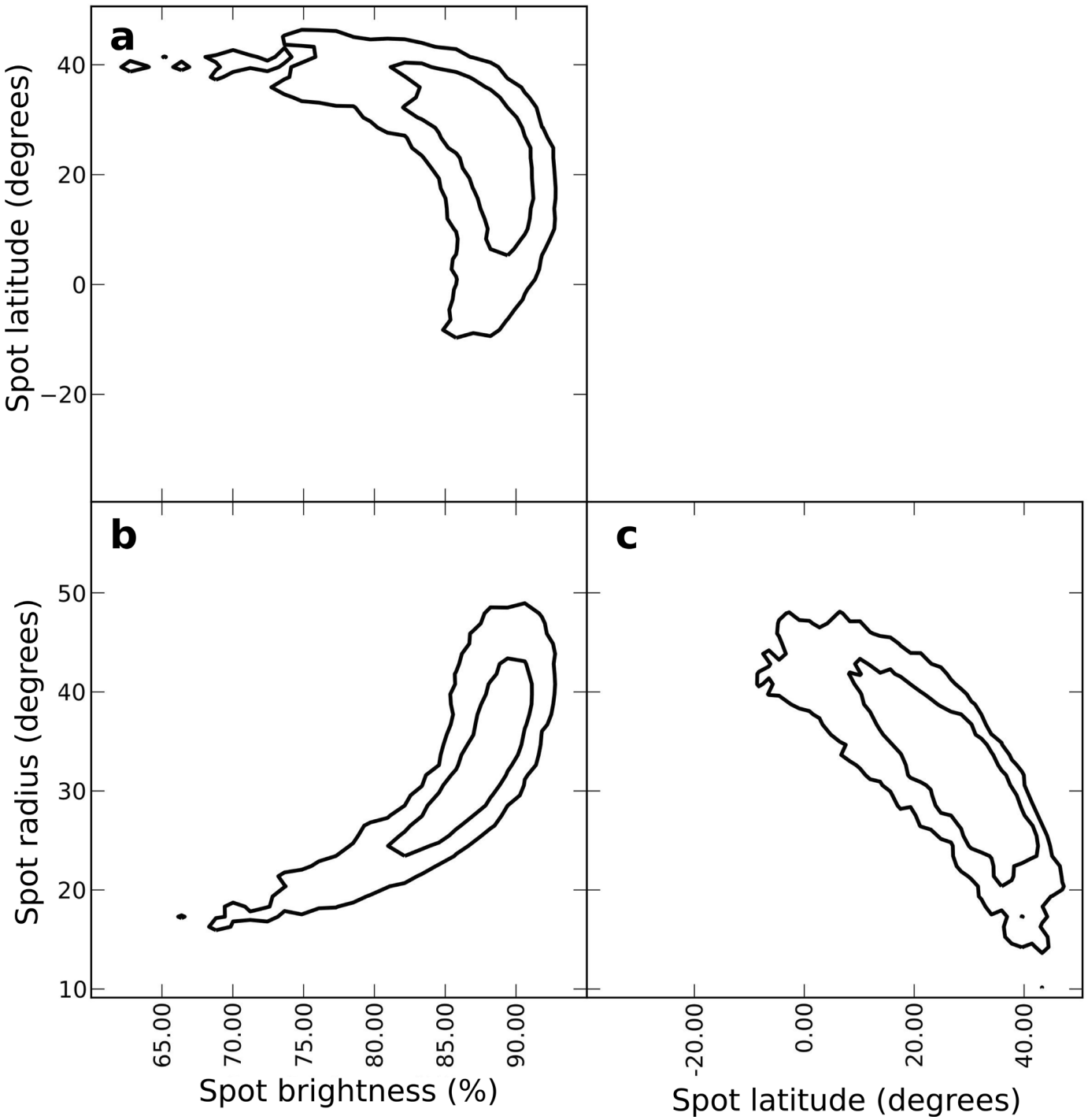
Extended Data Figure 1 | Spectral calibration for Luhman 16A (a and b) and Luhman 16B (c and d). The red curves (**a** and **c**) show the modelled spectra, which mostly overlap the observed spectra (plotted in black). The gaps in the spectra correspond to physical spaces between the four infrared array detectors. The residuals to the fits (**b** and **d**) are generally a few per cent, with larger deviations apparent near CO bandheads (for example, at 2.294 μm and 2.323 μm) and strong telluric absorption lines (for example, at 2.290 μm and 2.340 μm).



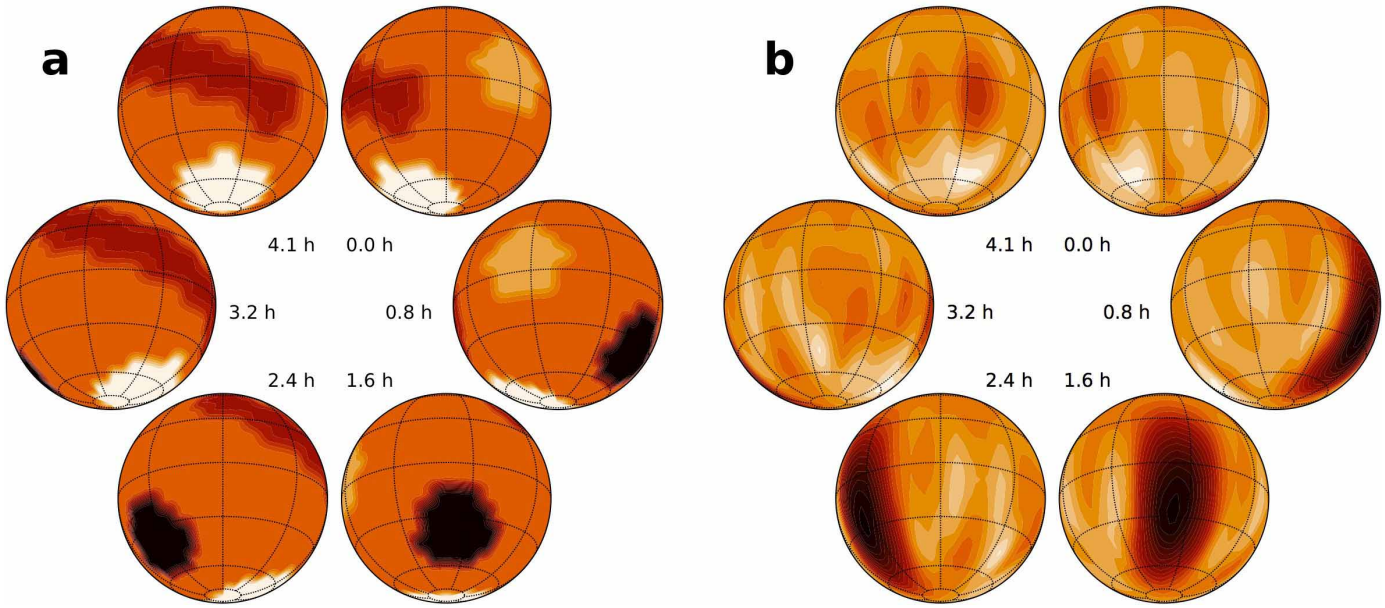
Extended Data Figure 2 | Calibrated spectra of the brown dwarfs, showing the individual calibrated spectra of Luhman 16A (a) and Luhman 16B (b). The time of each observation is indicated.



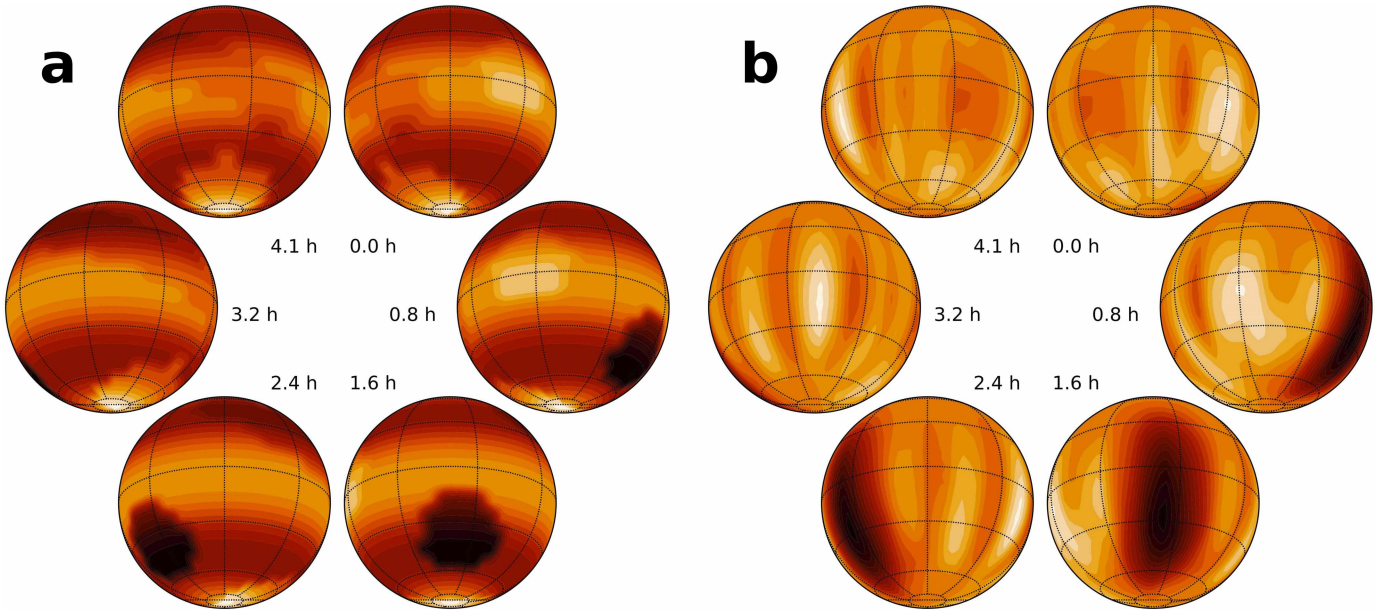
Extended Data Figure 3 | Luhman 16B shows strong rotationally induced variability (b) whereas Luhman 16A does not (a). The colour scale indicates the deviations from a uniform line profile as measured relative to the line continuum. Luhman 16B's variations are dominated by a dark region (diagonal streak, corresponding to a decrease of roughly 4% in equivalent width) that comes into view at 1.5 h heading towards the observer, rotates across the brown dwarf to the receding side, and is again hidden behind the limb at 3 h. Brighter regions are visible at earlier and later times, but are less prominent. No significant spectroscopic variability is apparent for Luhman 16A, and no coherent features are seen beyond Luhman 16B's projected rotational velocity (enclosed between the vertical dashed lines). All these points indicate that we are detecting intrinsic spectroscopic variability from Luhman 16B.



Extended Data Figure 4 | Posterior parameter distributions from our single-spot toy model, showing a large mid-latitude spot. The inner and outer curves in panels **a**–**c** indicate the 68.3% and 95.4% confidence regions. The plot shown assumes $i = 30$ degrees; smaller inclinations result in a slightly more equatorial spot, but the best-fit values remain within the inner 68.3% confidence regions.



Extended Data Figure 5 | Simulated brown dwarf with spots, and the map recovered from Doppler imaging. **a**, Simulated variable brown dwarf seen at an inclination of $i = 30$ degrees. The dark and light mid-latitude spots are, respectively, 40% darker and 10% brighter than the photosphere; the dark streak is 10% darker and the polar spot is 20% brighter. **b**, Surface map recovered from Doppler imaging, assuming noise levels similar to that seen in our observed data, after tuning the hyperparameter a to minimize spurious features. High-contrast features are recovered: the dark spot is in the correct location and the polar spot is only moderately distorted. The equatorial bright spot is visible in the recovered map, but it cannot be reliably distinguished from image artefacts that preferentially cluster near the equator. The dark stripe is not recovered. Thus our analysis can accurately recover strong features, but data quality precludes us from discerning smaller or fainter features.



Extended Data Figure 6 | Simulated brown dwarf with spots and bands, and the map recovered from Doppler imaging. **a**, Simulated variable brown dwarf with the same surface features as in Extended Data Fig. 5, but now also exhibiting zonal bands with an amplitude of ±20% of the mean photospheric brightness level. **b**, Surface map recovered from Doppler imaging under the same assumptions as in Extended Data Fig. 5. High-contrast, non-axisymmetric features are recovered as before, but we cannot recover even prominent global bands with the current precision of our data.

LETTER

doi:10.1038/nature12954

Observation of Dirac monopoles in a synthetic magnetic field

M. W. Ray[1], E. Ruokokoski[2], S. Kandel[1]†, M. Möttönen[2,3] & D. S. Hall[1]

Magnetic monopoles—particles that behave as isolated north or south magnetic poles—have been the subject of speculation since the first detailed observations of magnetism several hundred years ago[1]. Numerous theoretical investigations and hitherto unsuccessful experimental searches[2] have followed Dirac's 1931 development of a theory of monopoles consistent with both quantum mechanics and the gauge invariance of the electromagnetic field[3]. The existence of even a single Dirac magnetic monopole would have far-reaching physical consequences, most famously explaining the quantization of electric charge[3,4]. Although analogues of magnetic monopoles have been found in exotic spin ices[5,6] and other systems[7–9], there has been no direct experimental observation of Dirac monopoles within a medium described by a quantum field, such as superfluid helium-3 (refs 10–13). Here we demonstrate the controlled creation[14] of Dirac monopoles in the synthetic magnetic field produced by a spinor Bose–Einstein condensate. Monopoles are identified, in both experiments and matching numerical simulations, at the termini of vortex lines within the condensate. By directly imaging such a vortex line, the presence of a monopole may be discerned from the experimental data alone. These real-space images provide conclusive and long-awaited experimental evidence of the existence of Dirac monopoles. Our result provides an unprecedented opportunity to observe and manipulate these quantum mechanical entities in a controlled environment.

Maxwell's equations refer neither to magnetic monopoles nor to the magnetic currents that arise from their motion. Although a simple symmetrization with respect to the electric and magnetic fields, respectively E and B, leads to equations that involve these magnetic charges, it also seemingly prevents their description in terms of the familiar scalar and vector potentials, respectively V and A, alone. Because the quantum mechanical Hamiltonian is expressed in terms of potentials, rather than electromagnetic fields, this modification immediately leads to serious theoretical challenges.

In a celebrated paper that combined arguments from quantum mechanics and classical electrodynamics[3], Dirac identified electromagnetic potentials consistent with the existence of magnetic monopoles. His derivation relies on the observation that in quantum mechanics the potentials V and A influence charged-particle dynamics either through the Hamiltonian or, equivalently, through modifications of the complex phase of the particle wavefunction. Armed with these equivalent perspectives, Dirac then considered the phase properties of a wavefunction pierced by a semi-infinite nodal line with non-zero phase winding. He discovered that the corresponding electromagnetic potentials yield the magnetic field of a monopole located at the endpoint of the nodal line. The vector potential in this case also exhibits a nonphysical line singularity, or 'Dirac string', that terminates at the monopole.

We experimentally create Dirac monopoles in the synthetic electromagnetic field that arises in the context of a ferromagnetic spin-1 ^{87}Rb Bose–Einstein condensate (BEC) in a tailored excited state[14]. The BEC is described by a quantum mechanical order parameter that satisfies a nonlinear Schrödinger equation, and the synthetic gauge potentials describing a north magnetic pole (Fig. 1) are generated by the spin texture. This experiment builds on studies of synthetic electric and magnetic fields, respectively E^* and B^*, in atomic BECs, which is an emerging topic of intense interest in the simulation of condensed-matter systems with ultracold atoms[15,16]. Unlike monopole experiments in spin ices[5,6], liquid crystals[7], skyrmion lattices[9] and metallic ferromagnets[8], our experiments demonstrate the essential quantum features of the monopole envisioned by Dirac[3].

Physically, the vector potential, A^*, and synthetic magnetic field, $B^* = \hbar\nabla \times A^*$, are related to the superfluid velocity, v_s, and vorticity, $\Omega = \nabla \times v_s$, respectively. (Here $\hbar$ denotes Planck's constant divided by 2π.) Our primary evidence for the existence of the monopole comes from images of the condensate density taken after the creation of these fields (Figs 2 and 3), which reveal a nodal vortex line with 4π phase winding terminating within the condensate. The images also display a three-dimensional spin structure that agrees well with the results of

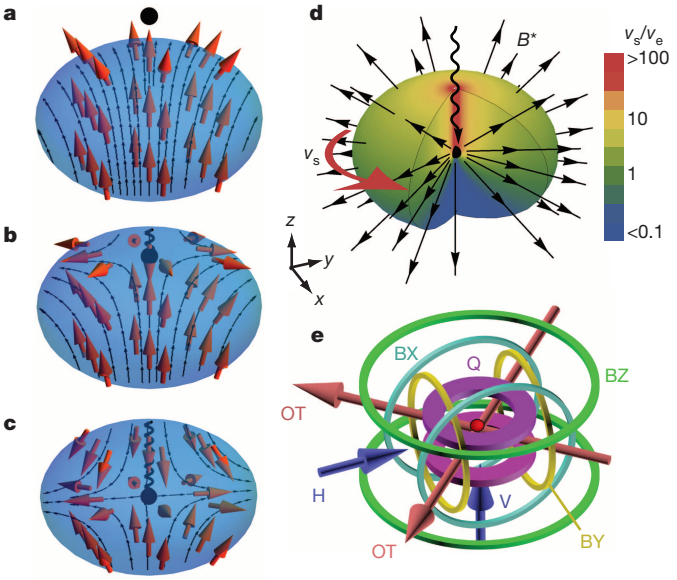


Figure 1 | Schematic representations of the monopole creation process and experimental apparatus. a–c, Theoretical spin orientation (red arrows) within the condensate when the magnetic field zero (black dot) is above (**a**), entering (**b**) and in the middle of (**c**) the condensate. The helix represents the singularity in the vorticity. **d**, Azimuthal superfluid velocity, v_s (colour scale and red arrow), scaled by equatorial velocity, v_e. Black arrows depict the synthetic magnetic field, B^*. **e**, Experimental set-up showing magnetic quadrupole (Q) and bias field (BX, BY and BZ) coils. Red arrows (OT) show beam paths of the optical dipole trap, and blue arrows indicate horizontal (H) and vertical (V) imaging axes. Gravity points in the $-z$ direction.

[1]Department of Physics, Amherst College, Amherst, Massachusetts 01002–5000, USA. [2]QCD Labs, COMP Centre of Excellence, Department of Applied Physics, Aalto University, PO Box 13500, 00076 Aalto, Finland. [3]Low Temperature Laboratory (OVLL), Aalto University, PO Box 13500, 00076 Aalto, Finland. †Present address: City of Hope National Medical Center, 1500 East Duarte Road, Duarte, California 91010, USA.

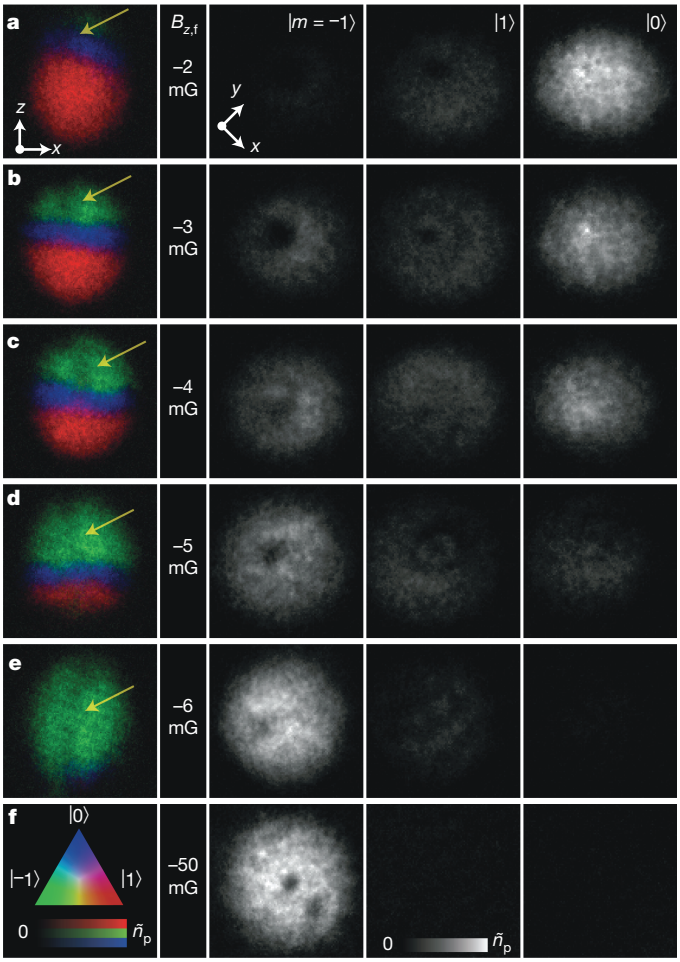


Figure 2 | Experimental creation of Dirac monopoles. Images of the condensate showing the integrated particle densities in different spin components as $B_{z,f}$ is decreased. Each row **a**–**f** contains images of an individual condensate. The leftmost column shows colour composite images of the column densities taken along the horizontal axis for the three spin states $\{|1\rangle, |0\rangle, |-1\rangle\}$; the colour map is given in **f**. Yellow arrows indicate the location of the nodal lines. The rightmost three columns show images taken along the vertical axis. The scale is 285 μm × 285 μm (horizontal) and 220 μm × 220 μm (vertical), and the peak column density is $\tilde{n}_p = 1.0 \times 10^9$ cm^{-2}.

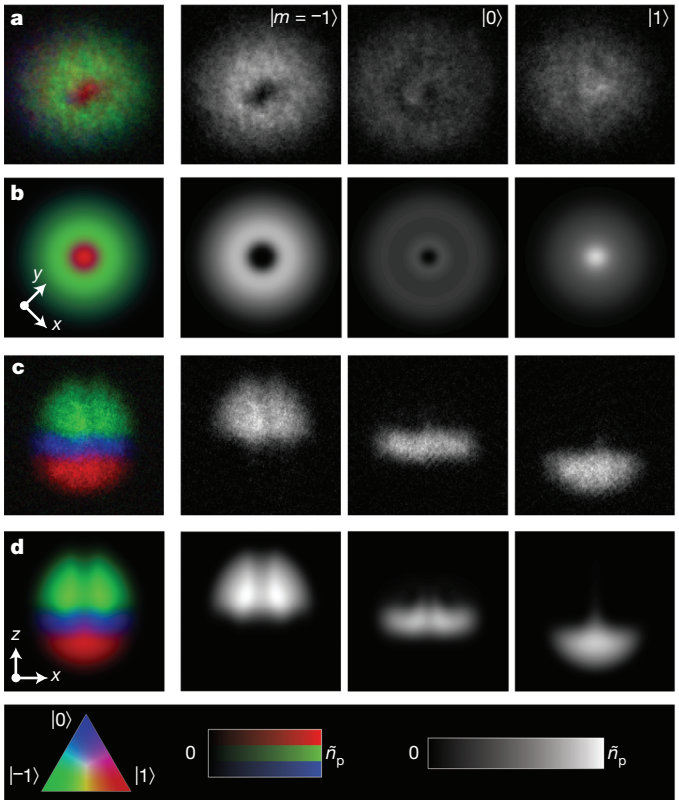


Figure 3 | Comparison between experiment and simulation. Experimental (**a**, **c**) and simulated (**b**, **d**) condensate particle densities with the monopole near the centre of the condensate. Comparisons along the vertical axis are shown in rows **a** and **b**, and those along the horizontal axis are shown in rows **c** and **d**. The hole observed in the $|-1\rangle$ component (row **a**) is discernible as a line of diminished density in row **c**. The field of view is 220 μm × 220 μm in **a** and **b** and 285 μm × 285 μm in **c** and **d**. The colour composite images and $\tilde{n}_p$ are as in Fig. 2.

numerical simulations (Fig. 4). We analyse these findings and discuss their implications below.

The spinor order parameter corresponding to the Dirac monopole[14,17] is generated by an adiabatic spin rotation in response to a time-varying magnetic field, $B(r, t)$. Similar spin rotations have been used to create multiply quantized vortices[18] and skyrmion spin textures[19]. The order parameter $\Psi(r, t) = \psi(r, t)\zeta(r, t)$ is the product of a scalar order parameter, ψ, and a spinor, $\zeta = (\zeta_{+1}, \zeta_0, \zeta_{-1})^{\mathrm{T}} \widehat{=} |\zeta\rangle$, where $\zeta_m = \langle m|\zeta\rangle$ represents the mth spinor component along z. The condensate is initially spin-polarized along the z axis, that is, $\zeta = (1, 0, 0)^{\mathrm{T}}$. Following the method introduced in ref. 14, a magnetic field $B(r,t) = b_q(x\hat{x} + y\hat{y} - 2z\hat{z}) + B_z(t)\hat{z}$ is applied, where $b_q > 0$ is the strength of a quadrupole field gradient and $B_z(t)$ is a uniform bias field. The magnetic field zero is initially located on the z axis at $z = B_z(0)/(2b_q) \gg Z$, where Z is the axial Thomas–Fermi radius of the condensate. The spin rotation occurs as B_z is reduced, drawing the magnetic field zero into the region occupied by the superfluid.

Ideally, the condensate spin adiabatically follows the local direction of the field (Fig. 1a–c). Our numerical analysis indicates, and both simulations and experiment confirm, that the fraction of atoms undergoing non-adiabatic spin-flip transitions is of order 1% for our experimental parameters. The spin texture in the adiabatic case is conveniently expressed in a scaled and shifted coordinate system with $x' = x$, $y' = y$, $z' = 2z - B_z/b_q$, corresponding derivatives ∇', and spherical coordinates (r', θ', φ'). This transformation scales the z axis by a factor of two and shifts the origin of coordinates to coincide with the zero of the magnetic field. The applied magnetic field is then $B = b_q(x'\hat{x}' + y'\hat{y}' - z'\hat{z}')$. As B_z is reduced, each spin rotates by an angle $\pi - \theta'$ about an axis $\hat{n}(r', \theta', \varphi') = -\hat{x}' \sin\varphi' + \hat{y}' \cos\varphi'$. This spatially dependent rotation leads to a superfluid velocity

$$v_s = \frac{\hbar}{Mr'} \frac{1 + \cos\theta'}{\sin\theta'} \hat{\varphi}' \tag{1}$$

and vorticity

$$\Omega = -\frac{\hbar}{Mr'^2} \hat{r}' + \frac{4\pi\hbar}{M} \delta(x')\delta(y')\Theta(z')\hat{r}' \tag{2}$$

where M is the atomic mass, δ is the Dirac delta function and Θ is the Heaviside step function. The vorticity is that of a monopole attached to a semi-infinite vortex line singularity, of phase winding 4π, extending along the $+z'$ axis.

The synthetic vector potential arising from the spin rotation can be written as $A^* = -Mv_s/\hbar$, with the line singularity in A^* coincident with the nodal line in Ψ. However, this singularity is nonphysical, because it depends on the choice of gauge and can even be made to vanish[20] (Supplementary Information). The synthetic magnetic field of the monopole is therefore simply

$$B^* = \frac{\hbar}{r'^2} \hat{r}' \tag{3}$$

The fields v_s and B^* are depicted in Fig. 1d.

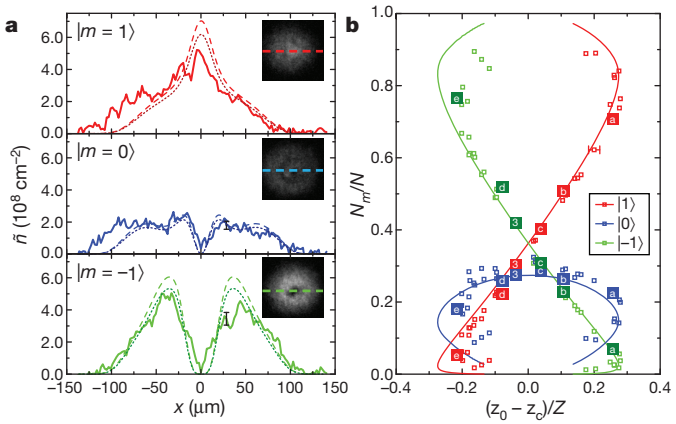


Figure 4 | Quantitative comparison between experiment and simulation. **a**, Experimental (solid lines) and simulated (dashed and dotted lines) column densities $\tilde{n}$ of the condensate from the vertical images in Fig. 3, with cross-sections taken as shown in the insets. Dotted lines show the approximate effect of three-body losses (see text). The origin $x = 0$ coincides with the hole in state $|0\rangle$. **b**, Fractions in each spin state for different positions of the centre of mass of the $|0\rangle$ state (z_0) relative to that of the condensate (z_c), in units of the axial Thomas–Fermi radius (Z). Solid lines are simulated values and points marked with letters and numbers correspond to panels **a**–**e** of Figs 2 and 3, respectively. Typical error bars that reflect uncertainties in the calibration of the imaging system are shown for several points.

The experimental set-up[21] is shown schematically in Fig. 1e. The optically trapped ^{87}Rb BEC consists of $N \approx 1.8(2) \times 10^5$ atoms in the $|F = 1, m = 1\rangle \equiv |1\rangle$ spin state, where the uncertainty reflects shot-to-shot variations and the calibration of the detection system. The calculated radial and axial Thomas–Fermi radii are $R = 6.5\,\mu$m and $Z = 4.6\,\mu$m, respectively, and the corresponding optical trap frequencies are respectively $\omega_r \approx 2\pi \times 160$ Hz and $\omega_z \approx 2\pi \times 220$ Hz. Four sets of coils are used to produce b_q, B_z and the transverse magnetic field components B_x and B_y, which are used to guide the applied magnetic field zero into the condensate. At the beginning of the monopole creation process, the bias field is $B_z = 10$ mG. The quadrupole field gradient is then linearly ramped from zero to $b_q = 3.7\,\mathrm{G\,cm^{-1}}$, placing the magnetic field zero approximately 30 μm above the condensate. The field zero is then brought down into the condensate by decreasing B_z linearly to $B_{z,f}$ at the rate $\dot{B}_z = -0.25\ \mathrm{G\ s^{-1}}$. We call this the 'creation ramp'.

The atomic density of each spinor component $|m\rangle$ is imaged as established by the local spin rotation during the creation ramp (Methods). As the field zero passes through the condensate (Fig. 2a–f), the distribution of particles in the three spin states changes in a manner indicative of the expected spin rotation shown in Fig. 1. The nodal line appears in the images taken along the vertical axis as holes in the $|-1\rangle$ and $|0\rangle$ components, and in the side images as regions of reduced density extending vertically from the top of the condensate towards, but not through, the $|1\rangle$ component. This nodal line extends more deeply into the condensate as $B_{z,f}$ is reduced. Ultimately it splits into two vortex lines (Fig. 2f; see also Extended Data Fig. 1)—the characteristic signature of the decay of a doubly quantized vortex[22]—illustrating its 4π phase winding.

We compare the experimental images of the vertically (Fig. 3a) and horizontally (Fig. 3c) imaged density profiles with those given by numerical simulations (Fig. 3b, d) in which the monopole is near the centre of the condensate. The simulation data are obtained by solving the full three-dimensional dynamics of the spinor order parameter (Methods). The locations of the doubly quantized and singly quantized vortices in spinor components $|-1\rangle$ and $|0\rangle$ are visible in the experimentally acquired density profiles, as are other structures discernible in the images obtained from the numerical simulations. The observed vertical spatial separation of the spinor components (Fig. 3c) confirms that the vortex line terminates within the bulk of the condensate.

The quantitative agreement between experiment and simulation is apparent in Fig. 4, which shows cross-sections of the density profiles taken through the centre of the condensate. The differences observed in the peak densities (Fig. 4a) of the experimental (solid lines) and simulated (dashed lines) data are due to effects not taken into account in the simulation, such as three-body losses that were observed to be ~10% in the experiment. To show their effect, we have scaled the simulated data accordingly (dotted lines). Noting the absence of free parameters, the experimental data are in very good agreement with the numerical simulation.

We also show the fraction of the condensate in each spinor component for different vertical monopole locations within the condensate (Fig. 4b), including data from images in which the nodal line of the order parameter does not necessarily coincide with the z axis. The physical observable is the position of the centre of mass of the $|0\rangle$ component, z_0, relative to the centre of mass of the whole condensate, z_c. Again, we find that the experiments and simulations are in very good quantitative agreement without any free parameters.

An alternative description of the origins of the velocity and vorticity profiles (equations (1) and (2)) can be presented in terms of the motion of the monopole (Supplementary Information). As the monopole approaches the condensate, it is a source not only of the synthetic magnetic field, B^* (equation (3)), but also of an azimuthal synthetic electric field, E^*, described by Faraday's law, $\nabla' \times E^* = -\partial B^*/\partial t$. Each mass element of the superfluid is given a corresponding azimuthal acceleration by E^*. The monopole motion thereby induces the appropriate superfluid velocity and vorticity profiles within the condensate, in a manner similar to the induction of electric current in a superconducting loop by the motion of a (natural) magnetic monopole[23]. In our case, the condensate itself is the monopole detector, analogous to the superconducting loop. Being three-dimensional, however, it is sensitive to the entire 4π solid angle surrounding the monopole.

The creation and manipulation of a Dirac monopole in a controlled environment opens up a wide range of experimental and theoretical investigations. The time evolution and decay[14] of the monopole are of particular interest because it is not created in the ground state[24]. Interactions between the monopole and other topological excitations, such as vortices, present another fundamental research avenue with a variety of unexplored phenomena. There exists also the possibility of identifying and studying condensate spin textures that correspond to other exotic synthetic electromagnetic fields, such as that of the non-Abelian monopole[25]. Finally, the experimental methods developed in this work can also be directly used in the realization of a vortex pump[26], which paves the way for the study of peculiar many-body quantum states, such as those related to the quantum Hall effect[27].

Note added in proof: The effects of the Lorentz force arising from an inhomogeneous synthetic magnetic field have recently been observed in condensate dynamics[28].

METHODS SUMMARY

Imaging. After the creation ramp, we non-adiabatically change B_z from $B_{z,f}$ to a large value (typically several hundred milligauss) to project the condensate spinor components $\{|m\rangle\}$ into the approximate eigenstates of the Zeeman Hamiltonian while preserving the monopole spin texture. We call this the 'projection ramp'. The condensate is then released from the trap and allowed to expand for 22.9 ms. The three spin states are separated along the x axis during the expansion by a 3.5-ms pulse of the magnetic field gradient with the magnetic bias field pointing in the x direction. We take images simultaneously along the horizontal and vertical axes.

Data. The images shown in Figs 2 and 3 are selected from among several dozen similar images taken under identical conditions, and hundreds of similar images taken under similar conditions (see also Extended Data Fig. 2 for representative examples). Not every experimental run yields an image of a monopole, because drifts in the magnetic field and location of the optical trap cause the magnetic field zero to pass outside the BEC. Under optimal conditions, five to ten consecutive images may be taken before drifts require adjustment of the bias fields.

Simulation. We solve the full three-dimensional Gross–Pitaevskii equation with simulation parameters chosen to match those of the experiment, excepting the effects of three-body losses and the magnetic forces arising from the gradient

during the spin state separation just before imaging. To show the effects of the expansion, we present integrated particle densities of the condensate from the numerical simulation immediately after the creation ramp, and while the magnetic field zero is still in the condensate, in Extended Data Fig. 3. The volume considered varies from $20 \times 20 \times 20 a_r^3$ to $320 \times 320 \times 320 a_r^3$, where $a_r = \sqrt{\hbar/M\omega_r} \approx 0.9$ μm is the radial harmonic oscillator length. The size of the computational grid changes from $180 \times 180 \times 180$ to $1{,}024 \times 1{,}024 \times 1{,}024$ points.

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 20 September; accepted 4 December 2013.

1. Goldhaber, A. S. & Trower, W. P. (eds) *Magnetic Monopoles* (American Association of Physics Teachers, 1990).
2. Milton, K. A. Theoretical and experimental status of magnetic monopoles. *Rep. Prog. Phys.* **69,** 1637–1711 (2006).
3. Dirac, P. A. M. Quantised singularities in the electromagnetic field. *Proc. R. Soc. Lond. A* **133,** 60–72 (1931).
4. Vilenkin, A. & Shellard, E. P. S. (eds) *Cosmic Strings and Other Topological Defects* (Cambridge Univ. Press, 1994).
5. Castelnovo, C., Moessner, R. & Sondhi, S. L. Magnetic monopoles in spin ice. *Nature* **451,** 42–45 (2008).
6. Morris, D. J. P. *et al.* Dirac strings and magnetic monopoles in the spin ice $Dy_2Ti_2O_7$. *Science* **326,** 411–414 (2009).
7. Chuang, I., Durrer, R., Turok, N. & Yurke, B. Cosmology in the laboratory: defect dynamics in liquid crystals. *Science* **251,** 1336–1342 (1991).
8. Fang, Z. *et al.* The anomalous Hall effect and magnetic monopoles in momentum space. *Science* **302,** 92–95 (2003).
9. Milde, P. *et al.* Unwinding of a skyrmion lattice by magnetic monopoles. *Science* **340,** 1076–1080 (2013).
10. Blaha, S. Quantization rules for point singularities in superfluid ^{3}He and liquid crystals. *Phys. Rev. Lett.* **36,** 874–876 (1976).
11. Volovik, G. & Mineev, V. P. Vortices with free ends in superfluid He^3-A. *JETP Lett.* **23,** 647–649 (1976).
12. Salomaa, M. M. Monopoles in the rotating superfluid helium-3 A–B interface. *Nature* **326,** 367–370 (1987).
13. Volovik, G. *The Universe in a Helium Droplet* 214–217 (Oxford Univ. Press, 2003).
14. Pietilä, V. & Möttönen, M. Creation of Dirac monopoles in spinor Bose-Einstein condensates. *Phys. Rev. Lett.* **103,** 030401 (2009).
15. Lin, Y.-J., Compton, R. L., Jiménez-Garcia, K., Porto, J. V. & Spielman, I. B. Synthetic magnetic fields for ultracold neutral atoms. *Nature* **462,** 628–631 (2009).
16. Dalibard, J., Gerbier, F., Juzeliūnas, G. & Öhberg, P. Artificial gauge potentials for neutral atoms. *Rev. Mod. Phys.* **83,** 1523–1543 (2011).
17. Savage, C. M. & Ruostekoski, J. Dirac monopoles and dipoles in ferromagnetic spinor Bose-Einstein condensates. *Phys. Rev. A* **68,** 043604 (2003).
18. Leanhardt, A. E. *et al.* Imprinting vortices in a Bose-Einstein condensate using topological phases. *Phys. Rev. Lett.* **89,** 190403 (2002).
19. Choi, J.-y., Kwon, W. J. & Shin, Y.-i. Observation of topologically stable 2D skyrmions in an antiferromagnetic spinor Bose-Einstein condensate. *Phys. Rev. Lett.* **108,** 035301 (2012).
20. Wu, T. T. & Yang, C. N. Concept of nonintegrable phase factors and global formulation of gauge fields. *Phys. Rev. D* **12,** 3845–3857 (1975).
21. Kaufman, A. M. *et al.* Radio-frequency dressing of multiple Feshbach resonances. *Phys. Rev. A* **80,** 050701 (2009).
22. Shin, Y. *et al.* Dynamical instability of a doubly quantized vortex in a Bose-Einstein condensate. *Phys. Rev. Lett.* **93,** 160406 (2004).
23. Cabrera, B. First results from a superconductive detector for moving magnetic monopoles. *Phys. Rev. Lett.* **48,** 1378–1381 (1982).
24. Ruokokoski, E., Pietilä, V. & Möttönen, M. Ground-state Dirac monopole. *Phys. Rev. A* **84,** 063627 (2011).
25. Pietilä, V. & Möttönen, M. Non-Abelian magnetic monopole in a Bose-Einstein condensate. *Phys. Rev. Lett.* **102,** 080403 (2009).
26. Möttönen, M., Pietilä, V. & Virtanen, S. M. M. Vortex pump for dilute Bose-Einstein condensates. *Phys. Rev. Lett.* **99,** 250406 (2007).
27. Roncaglia, M., Rizzi, M. & Dalibard, J. From rotating atomic rings to quantum Hall states. *Sci. Rep.* **1,** 43 (2011).
28. Choi, J.-y. *et al.* Observation of a geometric Hall effect in a spinor Bose-Einstein condensate with a skyrmion spin texture. *Phys. Rev. Lett.* **111,** 245301 (2013).

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

Acknowledgements We acknowledge funding by the National Science Foundation (grants PHY–0855475 and PHY–1205822), by the Academy of Finland through its Centres of Excellence Program (grant no. 251748) and grants (nos 135794, 272806 and 141015), and the Finnish Doctoral Programme in Computational Sciences. CSC – IT Center for Science Ltd is acknowledged for computational resources (project no. ay2090). We thank G. Volovik, M. Krusius, R. H. Romer, M. Nakahara and J. R. Friedman for their comments on the manuscript. We also thank H. Valja for his artistic input. M.W.R. and D.S.H. acknowledge discussions with R. P. Anderson and K. Jagannathan, and experimental assistance from N. B. Bern.

Author Contributions M.W.R., S.K. and D.S.H. developed and conducted the experiments, after which M.W.R. and D.S.H. analysed the data. E.R. performed the numerical simulations under the guidance of M.M., who also developed the gauge transformations presented in Supplementary Information. Interactive feedback between the experiments and simulations carried out by M.W.R., D.S.H., E.R. and M.M. was essential to achieving the reported results. All authors discussed both experimental and theoretical 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 D.S.H. (dshall@amherst.edu).

LETTER

doi:10.1038/nature12936

Site- and energy-selective slow-electron production through intermolecular Coulombic decay

Kirill Gokhberg[1], Přemysl Kolorenč[2], Alexander I. Kuleff[1] & Lorenz S. Cederbaum[1]

Irradiation of matter with light tends to electronically excite atoms and molecules, with subsequent relaxation processes determining where the photon energy is ultimately deposited and electrons and ions produced. In weakly bound systems, intermolecular Coulombic decay[1] (ICD) enables very efficient relaxation of electronic excitation through transfer of the excess energy to neighbouring atoms or molecules that then lose an electron and become ionized[2–9]. Here we propose that the emission site and energy of the electrons released during this process can be controlled by coupling the ICD to a resonant core excitation. We illustrate this concept with *ab initio* many-body calculations on the argon–krypton model system, where resonant photoabsorption produces an initial or 'parent' excitation of the argon atom, which then triggers a resonant-Auger-ICD cascade that ends with the emission of a slow electron from the krypton atom. Our calculations show that the energy of the emitted electrons depends sensitively on the initial excited state of the argon atom. The incident energy can thus be adjusted both to produce the initial excitation in a chosen atom and to realize an excitation that will result in the emission of ICD electrons with desired energies. These properties of the decay cascade might have consequences for fundamental and applied radiation biology and could be of interest in the development of new spectroscopic techniques.

Since its prediction[1] in 1997, ICD has been successfully investigated in a variety of systems[7]. It usually proceeds on a femtosecond timescale and becomes faster the more neighbours are present, often dominating most of the competing relaxation processes. ICD remains effective over considerable interatomic distances: in He dimers, the weakest bound systems known in nature, it is operative over distances of about 45 times the atomic radius[8,9]. The initial electronic excitation triggering ICD may be produced directly by photoabsorption, electron impact or even ion impact, as demonstrated recently[10]. It can also result from multistage processes such as Auger decay[11–13], with the overall Auger-ICD cascade initiated by core ionization of an atom (typically through X-ray absorption) that is part of a more complex system. In this case, however, there is little control over where exactly the Auger decay is triggered and where the subsequent ICD takes place. Indeed, in a polyatomic system, all atoms with core-ionization potentials below the energy of the impacting photon may become ionized and undergo an Auger transition.

Our proposal for realizing ICD with control over both the location of the process and the energies of the emitted ICD electrons exploits resonant Auger decay[14]. It uses photons with an energy just below the core-ionization threshold of a selected atom in a larger system, so that at a number of discrete energies the core electron will resonantly absorb the photon and be promoted to some bound, unoccupied orbital to give a highly energetic, core-excited state that can decay through the emission of an Auger electron. In this process, a valence electron fills the initial vacancy and another valence electron is ejected into the continuum, while the initially excited electron remains a spectator. This 'spectator resonant Auger' mechanism produces highly excited valence-ionized states (known as photoionization satellite states). The alternative, 'participator', process, in which the initially excited electron participates in the decay, is usually the much less efficient de-excitation pathway of core excitations.

As sketched in Fig. 1, the resonant Auger decay transforms the initially core-excited species into an excited, valence-ionized state with an excess energy of typically a few tens of electronvolts; the latter can then transfer its excess energy to the environment by continuing to decay electronically through ICD. In contrast to Auger-driven ICD, this resonant-Auger-driven ICD (RA–ICD) offers control over key features of the ICD process. First, in a given environment, the energy of emitted ICD electrons depends sensitively on the energies and populations of the states produced by resonant Auger decay, which in turn depend on the nature of the parent, core-excited state. This offers the possibility of varying the energetic composition of the ICD spectra in a controlled manner by adjusting the energy of the initiating, high-energy photon to resonantly excite the particular parent state that will produce the desired ICD electrons. Second, the initial parent core excitation can be placed selectively not only on chemically different atoms but also on identical atoms occupying non-equivalent sites in the system. (This selectivity stems from the different chemical shifts the atoms experience in different chemical environments and is used in near edge X-ray absorption fine structure spectroscopy to study, for example, the bonding in biologically relevant organic molecules[15].) And because the resonant Auger decay tends to be local and to populate excited valence-ionized states with two holes localized predominantly on the atom bearing the initial excitation[16], the subsequent ICD will mostly ionize the environment in the vicinity of the parent core excitation (Fig. 1b). In other words, the site where the ICD electrons are produced can be selectively chosen.

We illustrate the RA–ICD cascade for ArKr, a simple system that allows for a particularly clear illustration of the processes involved. Here RA–ICD can be initiated using a photon energy of 246.51 eV selectively to populate the $2p_{1/2}^{-1}4s$ state of Ar (ref. 17), which lives for only 5.5 fs (ref. 18) and locally undergoes spectator Auger decay populating a band of excited states of Ar^+ (Methods). These excited valence-ionized states lie between 17 and 22 eV above the ground state of Ar^+ and can therefore undergo ICD with the neighbouring Kr, whose lowest ionization potential is 14 eV. The ICD rates determined from extensive *ab initio* many-body calculations confirm that the ionic states indeed further decay by ICD, with the calculated electron spectrum (Fig. 2a) exhibiting a pronounced peak between 0 and 1 eV and a weaker peak between 2 and 4 eV (see Methods for details of the computational scheme). Following the ICD, Ar^+ and Kr^+ will repel each other, resulting in a dissociation process known as a Coulomb explosion, which endows the ions with ~3.7 eV of kinetic energy.

On increasing the energy of the X-ray photon by just 0.4 eV to 246.93 eV, the $2p_{3/2}^{-1}3d$ parent state of Ar is excited. The resonant Auger decay of this core excitation populates a completely different band of excited states of Ar^+, but these can also all decay further by ICD (Methods). In this case, the spectrum of the emitted electrons (Fig. 2b) consists of one peak between 3 and 5 eV and another between 6 and 8 eV. We see that two different core excitations of the same atom result in very different

[1]Theoretische Chemie, Physikalisch-Chemisches Institut, Universität Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany. [2]Institute of Theoretical Physics, Faculty of Mathematics and Physics, Charles University in Prague, V Holešovičkách 2, 180 00 Prague, Czech Republic.

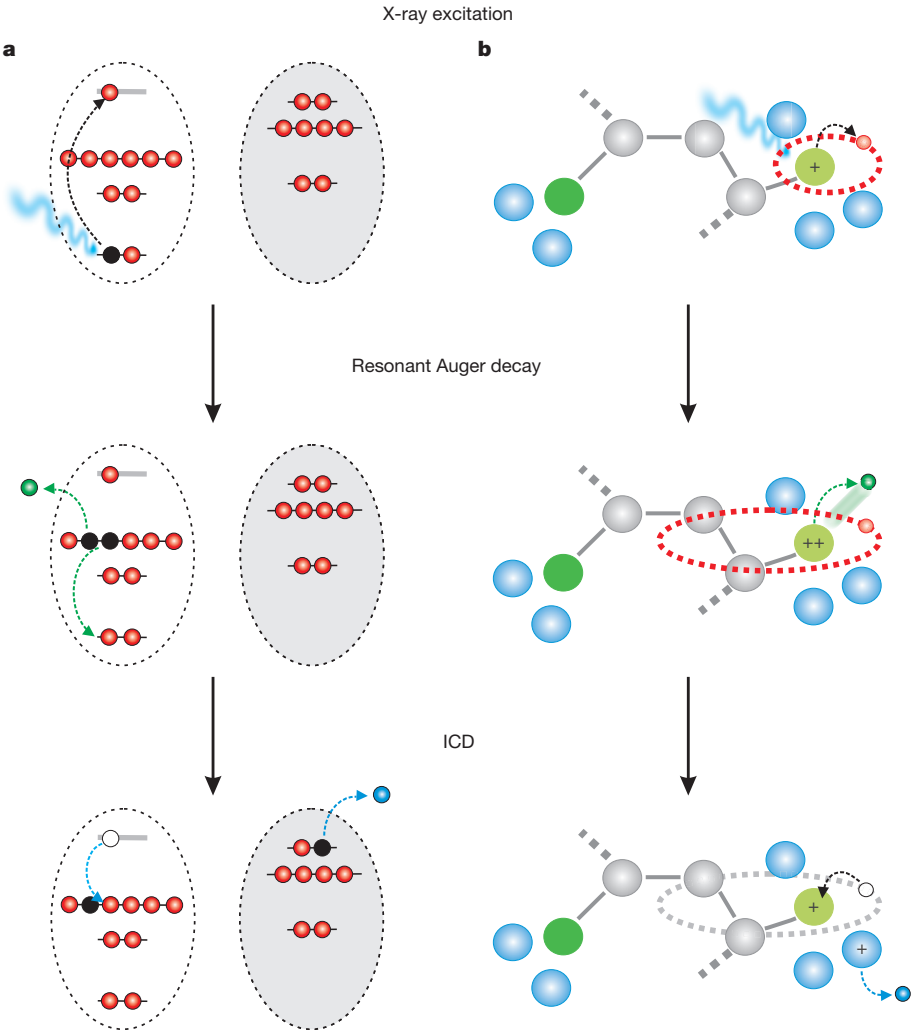


Figure 1 | Schematic illustration of the RA–ICD cascade. a, The mechanism. A parent core-excited state embedded in the environment decays locally by the spectator resonant Auger process, producing highly excited valence-ionized states. These states continue to decay by ICD, ionizing the neighbours in the environment. The two cations produced by ICD repel each other and undergo a Coulomb explosion (not shown). **b**, Selectivity property. The parent state is produced selectively in a given atom of the embedded system. The excited valence-ionized states formed in the resonant Auger process tend to be localized at the site of the initial excitation and decay by ICD, ionizing predominantly neighbours from the environment nearest to this site.

energy distributions of the emitted ICD electrons, illustrating the potential to control the energies of the ICD electrons. We note that in both excitation schemes, only a fraction of a percent of the total decay rate[19] is accounted for by the participator Auger channel, which does not result in ICD; in contrast, the spectator Auger final states, which do undergo further ICD, comprise about 75% of the total population in the case of the $2p_{1/2}^{-1}4s$ parent excitation and more than 95% in the case of the $2p_{3/2}^{-1}3d$ parent excitation. In other systems, such as the molecular dimers experimentally shown to undergo RA–ICD in the companion paper to this one[20], these and other details may of course differ; but the basic underlying mechanism of the RA–ICD cascade will be similar to what we have shown for ArKr.

The ability to control the location and energies of ICD electrons by core-exciting selected atoms to different parent states suggests that the RA–ICD cascade could serve as the foundation for a promising analytical technique. Intriguing possibilities may arise from the fact that the method combines intramolecular Auger decay, which produces Auger electrons that can be used to study the electronic structure of the molecule at the excitation site, with intermolecular, neighbour-involving ICD, which produces electrons that can be used to probe the local environment (Fig. 1b). In the latter regard, RA–ICD may seem similar to the multi-atom resonant photoemission effect suggested to be sensitive to the local chemical environment of an atom in a crystal[21]. Although that effect is also initiated by a core excitation, it involves a single interatomic de-excitation step that results in core-electron emission from a neighbouring atom. But this interatomic decay mode is strongly suppressed[22] and is thus difficult to use, owing to the strong competition of the resonant Auger process in the primary excited atom. In contrast, the final ICD step of the RA–ICD cascade in atomic systems has no competitor (except for the much slower radiative decay) and the whole process is extremely efficient. Even in molecular systems, where additional relaxation modes involving nuclear dynamics are present, the ICD process remains very effective[2,20].

In closing, we note that radiation-induced DNA damage is generally attributed to electrons with energies less than 500 eV (ref. 23) and to radical species[24]. Some radiotherapy approaches incorporate high-atomic-number elements as Auger-electron emitters[25,26] into DNA, for the targeted production of genotoxic electrons thought to arise from the local Auger cascade in the high-atomic-number element. However, a large number of interatomic decay channels will be open in a high-atomic-number Auger-electron emitter placed in an environment as complex as DNA and its solvation shell. Indeed, the probability of ICD-like processes[27,28] taking place in this system and simultaneously generating genotoxic electrons and radical cations will be very high and ought to be considered (Methods). We also note that the RA–ICD cascade is more efficient and selective in producing genotoxic species than are traditional

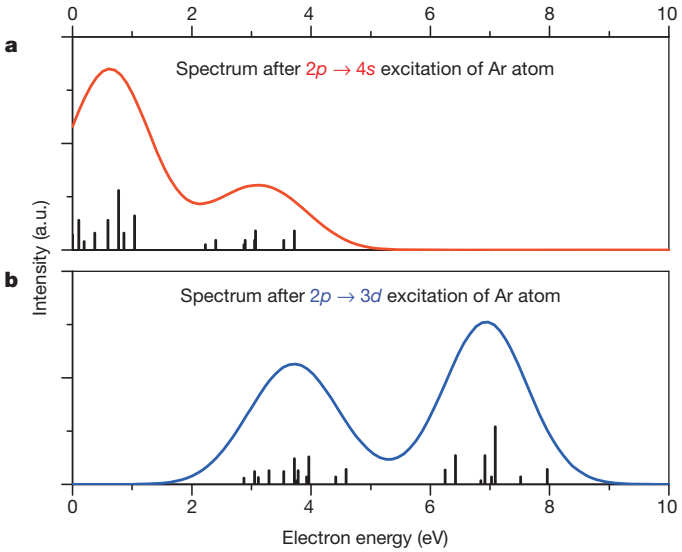


Figure 2 | Spectra of the ICD electrons emitted in the RA–ICD cascades in ArKr. ICD electron spectra for core excitation of the $Ar(2p_{1/2}^{-1}4s)$ parent state at 246.51 eV (**a**) and core excitation of the $Ar(2p_{3/2}^{-1}3d)$ parent state at 246.93 eV (**b**). The discrete lines are obtained using the frozen-nuclei approximation, whereas the continuous lines are evaluated by the convolution of each discrete line with a Gaussian with a full-width at half-maximum of 1.4 eV, qualitatively accounting for the nuclear dynamics (see Methods for computational details). The spectra illustrate that two different core excitations of the same atom lead to very different energy distributions of the ICD electrons. a.u., arbitrary units.

photon-activated techniques that initiate the Auger cascade through K-shell ionization[26]. In particular, the site- and energy-selectivity of the resonant core excitation process make it possible to tune the energies of the slow electrons generated. This feature might prove useful in optimizing radiotherapy efficiency. It has been shown[29,30] that electrons with energies between 0 and 4 eV predominantly induce single-strand breaks in DNA, and that the more damaging double-strand breaks are more efficiently produced by electrons with energies greater than 6 eV. We believe that a detailed mechanistic understanding of DNA lesions, in conjunction with the tunability of RA–ICD, could offer enough control over radiation-induced cell damage to lead to efficient cancer therapies.

METHODS SUMMARY

The ICD lifetimes of the involved states were computed using an *ab initio* many-body method. This method is based on the general Fano resonance formalism, in which the initial decaying state is represented as a bound (discrete) state embedded in the continuum of final states of the decay. The $\mathcal{L}^2$ approximations for the discrete and continuum components of the $(N-1)$-electron wavefunction are obtained within the Green's function in the algebraic diagrammatic construction (ADC) approach, and the resulting discretized spectrum is renormalized and interpolated in energy using the Stieltjes imaging technique. The potential energy curves of the initial and final ICD states were modelled using atomic data. The final ICD-electron spectra were obtained by convolving the discrete electronic transitions with an appropriate Gaussian profile.

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 7 May 2012; accepted 28 November 2013.
Published online 22 December 2013.

1. Cederbaum, L. S., Zobeley, J. & Tarantelli, F. Giant intermolecular decay and fragmentation of clusters. *Phys. Rev. Lett.* **79,** 4778–4781 (1997).
2. Jahnke, T. *et al.* Ultrafast energy transfer between water molecules. *Nature Phys.* **6,** 139–142 (2010).
3. Mucke, M. *et al.* A hitherto unrecognized source of low-energy electrons in water. *Nature Phys.* **6,** 143–146 (2010).
4. Grieves, G. A. & Orlando, T. M. Intermolecular Coulomb decay at weakly coupled heterogeneous interfaces. *Phys. Rev. Lett.* **107,** 016104 (2011).
5. Schwartz, C. P., Fatehi, S., Saykally, R. J. & Prendergast, D. Importance of electronic relaxation for inter-Coulombic decay in aqueous systems. *Phys. Rev. Lett.* **105,** 198102 (2010).
6. Stoychev, S. D., Kuleff, A. I. & Cederbaum, L. S. Intermolecular Coulombic decay in small biochemically relevant hydrogen-bonded systems. *J. Am. Chem. Soc.* **133,** 6817–6824 (2011).
7. Hergenhahn, U. Interatomic and intermolecular Coulombic decay: the early years. *J. Electron Spectrosc. Relat. Phenom.* **184,** 78–90 (2011).
8. Sisourat, N. *et al.* Ultralong-range energy transfer by interatomic Coulombic decay in an extreme quantum system. *Nature Phys.* **6,** 508–511 (2010).
9. Havermeier, T. *et al.* Interatomic Coulombic decay following photoionization of the helium dimer: observation of vibrational structure. *Phys. Rev. Lett.* **104,** 133401 (2010).
10. Kim, H. K. *et al.* Enhanced production of low energy electrons by alpha particle impact. *Proc. Natl Acad. Sci. USA* **108,** 11821–11824 (2011).
11. Santra, R. & Cederbaum, L. S. Coulombic energy transfer and triple ionization in clusters. *Phys. Rev. Lett.* **90,** 153401 (2003).
12. Morishita, Y. *et al.* Experimental evidence of interatomic Coulombic decay from the Auger final states in argon dimers. *Phys. Rev. Lett.* **96,** 243402 (2006).
13. Yamazaki, M. *et al.* Decay channel dependence of the photoelectron angular distributions in core-level ionization of Ne dimers. *Phys. Rev. Lett.* **101,** 043004 (2008).
14. Ueda, K. Core excitation and de-excitation spectroscopies of free atoms and molecules. *J. Phys. Soc. Jpn* **75,** 032001 (2006).
15. Feyer, V. *et al.* Core level study of alanine and threonine. *J. Phys. Chem. A* **112,** 7806–7815 (2008).
16. Hergenhahn, U. *et al.* The resonant Auger spectra of formic acid, acetaldehyde, acetic acid and methyl formate. *Chem. Phys.* **289,** 57–67 (2003).
17. de Gouw, J. A., van Eck, J., Peters, A. C., van der Weg, J. & Heideman, H. G. M. Resonant Auger spectra of the $2p^{-1}nl$ states of argon. *J. Phys. B* **28,** 2127–2141 (1995).
18. Fuggle, J. C. & Alvarado, S. F. Core-level lifetimes as determined by X-ray photoelectron spectroscopy measurements. *Phys. Rev. A* **22,** 1615–1624 (1980).
19. Gorczyca, T. W. & Robicheaux, F. Auger decay of the photoexcited $2p^{-1}nl$ Rydberg series in argon. *Phys. Rev. A* **60,** 1216–1225 (1999).
20. Trinter, F. *et al.* Resonant Auger decay driving intermolecular Coulombic decay in molecular dimers. *Nature* http://dx.doi.org/10.1038/nature12927 (this issue).
21. Kay, A. *et al.* Multi-atom resonant photoemission: a method for determining near-neighbor atomic identities and bonding. *Science* **281,** 679–683 (1998).
22. Carravetta, V. & Ågren, H. An ab initio method for computing multi-atom resonant photoemission. *Chem. Phys. Lett.* **354,** 100–108 (2002).
23. Pomplun, E. A new DNA target model for track structure calculations and its first application to I-125 Auger electrons. *Int. J. Radiat. Biol.* **59,** 625–642 (1991).
24. von Sonntag, C. *Free-Radical-Induced DNA Damage and Its Repair* (Springer, 2006).
25. Adelstein, J. S., Kassis, A. I., Bodei, L. & Mariani, G. Radiotoxicity of iodine-125 and other Auger-electron-emitting radionuclides: background to therapy. *Cancer Biother. Radiopharm.* **18,** 301–316 (2003).
26. Fairchild, R. G., Brill, A. B. & Ettinger, K. V. Radiation enhancement with iodinated deoxyuridine. *Invest. Radiol.* **17,** 407–416 (1982).
27. Aziz, E. F., Ottosson, N., Faubel, M., Hertel, I. V. & Winter, B. Interaction between liquid water and hydroxide revealed by core-hole de-excitation. *Nature* **455,** 89–91 (2008).
28. Pokapanich, W. *et al.* Ionic-charge dependence of the intermolecular Coulombic decay time-scale for aqueous ions probed by the core-hole clock. *J. Am. Chem. Soc.* **133,** 13430–13436 (2011).
29. Boudaïffa, B., Cloutier, P., Hunting, D., Huels, M. A. & Sanche, L. Resonant formation of DNA strand breaks by low-energy (3 to 20 eV) electrons. *Science* **287,** 1658–1660 (2000).
30. Martin, F. *et al.* DNA strand breaks induced by 0–4 eV electrons: the role of shape resonances. *Phys. Rev. Lett.* **93,** 068101 (2004).

Acknowledgements The research leading to these results received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Advanced Investigator Grant no. 227597. P.K. acknowledges the support from the Czech Science Foundation (grant no. P208/12/0521).

Author Contributions K.G., A.I.K. and L.S.C. had the idea for the cascade mechanism and its potential consequences. P.K. computed the lifetimes, and K.G. and A.I.K. evaluated the electronic spectra. K.G., A.I.K. and L.S.C. 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 K.G. (kirill.gokhberg@pci.uni-heidelberg.de) or A.I.K. (alexander.kuleff@pci.uni-heidelberg.de).

METHODS

Identifying open ICD channels. After the excited valence-ionized states of ArKr have been populated by resonant Auger decay of the $2p_{1/2}^{-1}4s$ or $2p_{3/2}^{-1}3d$ parent excitation of Ar, they can further decay by ICD that can be identified in the following simple way.

As the parent excitation and the following resonant Auger process are ultrafast (the resonant-Auger lifetime in the studied cases is only 5.5 fs (ref. 18)), the Auger transitions take place at equilibrium internuclear distance of the neutral system, $R_{eq} = 3.88$ Å. Therefore, the states that can further decay by ICD are those whose energies at around R_{eq} are larger than the energies of the final Ar^+Kr^+ states of the ICD process. To determine the relevant energies at around R_{eq}, we first make use of the fact that bound excited valence-ionized states of van der Waals dimers are known to be very shallow, having dissociation energies in the millielectronvolt range[31]. Therefore, the potential energy curves of the $Ar^+(3p^{-2}nl)$Kr states can be well approximated by horizontal lines calibrated to the corresponding $Ar^+(3p^{-2}nl)$ energies. In contrast, the Coulomb repulsion between the two ions in the final Ar^+Kr^+ states of ICD determines the $1/R$ asymptotic behaviour of the corresponding potential energy curves. As shown in our previous studies[32,33], analytical curves based on the Coulomb law and calibrated at infinite internuclear distance to the sum of the energies of the corresponding atomic fragments[34] give reliable values for the energies of these states around R_{eq}.

In Extended Data Fig. 1, we depict the energy diagram of the excited valence-ionized states of ArKr most populated in the resonant Auger decay of the $2p_{1/2}^{-1}4s$ parent state together with all possible final ICD states. From the plot, it becomes obvious that apart from the $Ar^+(3p^{-2}(^3P)4s)$Kr state, all other excited valence-ionized states depicted in the graph can decay by ICD around R_{eq}. The figure also gives the relative populations of the final Auger states[17] (in per cent).

Increasing the energy of the X-ray photon by just 0.4 eV would excite the $2p_{3/2}^{-1}3d$ parent state at 246.93 eV. The resonant Auger decay of this core excitation populates a totally different band of excited valence-ionized states. Two series of $Ar^+(3p^{-2}nl)$Kr states are populated by the resonant Auger transition[17]. The $Ar^+(3p^{-2}3d)$Kr states are produced in the strict spectator transition with the excited electron still occupying the $3d$ orbital. In addition, the resonant Auger decay strongly populates the 'shake-up satellite' $Ar^+(3p^{-2}4d)$Kr states, where the excited spectator electron is promoted to the higher lying $4d$ orbital. The energy diagram of these states together with the final Ar^+Kr^+ states populated by the ICD is shown in Extended Data Fig. 2. We see that all the excited valence-ionized states produced by the resonant Auger process can further decay by ICD around R_{eq}.

Evaluation of the ICD rates. To see whether the ICD is an operative mode of relaxation of these ionized excited states, we need to compute the ICD rates or, equivalently, the ICD lifetimes. For the evaluation of the ICD lifetimes, we used an *ab initio* many-body approach. The method is well documented in the literature[35] and is based on the general Fano resonance formalism[36], in which the initial decaying state is represented as a bound (discrete) state embedded in and interacting with the continuum of final states of the decay. The $\mathcal{L}^2$ approximations for the discrete and continuum components of the $(N-1)$-electron wavefunction are obtained within the Green's function in the ADC approach[37], and the resulting discretized spectrum is renormalized and interpolated in energy using the Stieltjes imaging technique[38]. The Green's function calculations were performed using large basis sets. Both effective core potential ECP.Dolg.6s6p3d1f.4s4p3d1f.8e-MWB (ref. 39) and standard Dunning aug-cc-pVQZ (refs 40, 41) basis sets with additional diffuse and distributed functions were used as an input data, giving very similar results for the lifetimes.

Each state produced by resonant Auger decay and decaying by ICD has been found to have its own individual lifetime. The ICD lifetimes of the various $Ar^+(3p^{-2}4s)$Kr and $Ar^+(3p^{-2}3d)$Kr decaying states were found to be between 13 and 220 fs, and the lifetimes of the shake-up $Ar^+(3p^{-2}4d)$Kr satellites are between 600 fs and 2 ps. We see that indeed the ICD is efficient and will be the primary relaxation mode of the states produced by the resonant Auger transition. This is also confirmed by the recent measurements[20] of the RA–ICD cascade in $(N_2)_2$ and $(CO)_2$ showing that the ICD takes place before the excited molecule is able to undergo dissociation, suggesting a timescale of <10 fs for the ICD process. Notably, the ICD rate strongly increases with the increase of the number of neighbours[42,43], making the ICD in larger clusters, as well as in biological media, an extremely efficient mode of relaxation.

ICD electron spectra after RA–ICD cascade in ArKr. From the generally high efficiency of the ICD process, we can estimate the ICD electron spectra assuming that the whole cascade takes place at the equilibrium distance. This assumption is strongly supported by the measurements[20] on the process in N_2 and CO dimers, which show that the whole cascade takes place at the equilibrium distance of the dimer. Within this approximation, the ICD-electron spectrum will consist of discrete lines with positions corresponding to the energy differences between the initial and final states of the ICD process, and heights reflecting the population of the given decaying state and the multiplicity of the final one. For the two studied cases, the ICD-electron spectra obtained in this way have been plotted in Fig. 2. We see that, for the $2p_{1/2}^{-1}4s$ parent state, the spectrum has two peaks: a pronounced peak between 0 and 1 eV and a weaker peak between 2 and 4 eV (Fig. 2a). For the $2p_{3/2}^{-1}3d$ parent state, the spectrum again consists of two peaks but in different energy regions: a peak between 3 and 5 eV and another between 6 and 8 eV (Fig. 2b). Comparing the spectra, we notice that a small change in core-excitation energies leads to totally different energies of electrons emitted in the ICD. In both cases, the final ICD products, Ar^+ and Kr^+, will repel each other, resulting in a Coulomb explosion. At the end of this dissociative process, the ions will acquire a kinetic energy of 3.7 eV. This energy can be directly measured in dimers, resulting in the 'kinetic energy release spectrum'[2]. In the frozen-nuclei approximation, this spectrum would consist of a single line at 3.7 eV. As for the ICD-electron distribution (see below), the nuclear motion will introduce a broadening of this line.

The spectra of discrete lines shown in Fig. 2 reflect only the electronic degrees of freedom. We can take a step further and account for the vibrational broadening that the initial distribution of the positions of the nuclei in the neutral will introduce. This initial distribution is given by a wave packet (essentially a Gaussian) that is centred at R_{eq} and has a width of about 0.4 Å. Therefore, to account for the vibrational broadening, each discrete line in Fig. 2 has to be convolved with a Gaussian with a full-width at half-maximum of 1.4 eV. This value reflects the width of the wave packet (0.4 Å) in the electronic ground state of ArKr. The results of this procedure are shown with continuous lines in the electron spectra in Fig. 2. We note that, despite its simplicity, this procedure for obtaining ICD-electron spectra usually gives reliable results. For instance, the ICD-electron spectrum of a water dimer[44,45] obtained using this procedure is in fairly good agreement with the experimental results[2].

We also note that computing highly accurate ICD electron spectra for the RA–ICD cascade in ArKr is beyond the scope of the present paper. The emphasis here is put on uncovering the potential that this cascade offers for larger systems. The example of ArKr is used only to illustrate the high degree of control with which low-energy electrons (LEE) and radical cations can be produced. Both low-energy electrons and radical cations are known to be important in radiation biology, because they induce DNA lesions.

Additional sources of ICD electrons. Relaxation of a core-excited high-Z element embedded in a biological medium will result in ICD-electron emission in the terminal step of an RA–ICD cascade, but other ICD processes can be a source of additional genotoxic electrons. Here we briefly discuss some of these possibilities, noting that all ICD processes simultaneously produce an ICD electron and a radical cation that both contribute to the DNA damage[24].

One type is core-ICD processes. For Auger cascades taking place in an environment, the emission of Auger electrons is accompanied by the emission of electrons by the core-ICD process, in which the parent excitation decays not in a local Auger cascade step but by interatomically ionizing the environment. Therefore, at each step of the corresponding cascade, the energy release accompanying the core transition on the parent species can either be used to ionize it (Auger process) or be transferred to the environment to ionize a neighbour instead (core-ICD process). The core-ICD process was experimentally observed both after core excitation[27] and after core ionization[28,46], with core ionization of solvated metallic ions shown[46] to follow decay pathways where the core-ICD/Auger branching ratio reaches values as high as 40%. More relevant for the present discussion, the core excitation of OH^-_{aq} was found[27] to relax through the core-ICD process with neighbouring water molecules, with the local Auger decay completely suppressed. This outstanding efficiency of the core-ICD process has been explained[4] by the pronounced overlap between the molecular orbitals located on the parent species and the outer valence orbitals of the water molecules in the solvation shell.

It may also happen that in the Auger cascade in the parent species some Coster–Kronig transitions are energetically forbidden and do not appear in the cascade. In an environment such transitions may, however, become allowed and proceed by ionizing the neighbours. For example, the ionization of $2s$ electron of the isolated Na^+, Mg^{2+} and Al^{3+} ions does not lead to the electronic decay of the resulting states. The energy of the $2p \rightarrow 2s$ transition is not sufficient to remove an additional electron from the ions. In aqueous solutions, new interatomic decay channels open owing to the presence of neighbours, because the energy released in the $2p \rightarrow 2s$ transition is sufficient to ionize the water molecules. By analysing the photoelectron spectra of the hydrated ions, it has been shown[47] that the lifetimes of the respective $2s$ ionized states are between 3 and 1 fs, indicating that the process is, surprisingly, highly efficient. Another relevant observation on high-Z cascades is that the rates of these core-ICD transitions tend to increase with increasing charge on the parent ion[28,47].

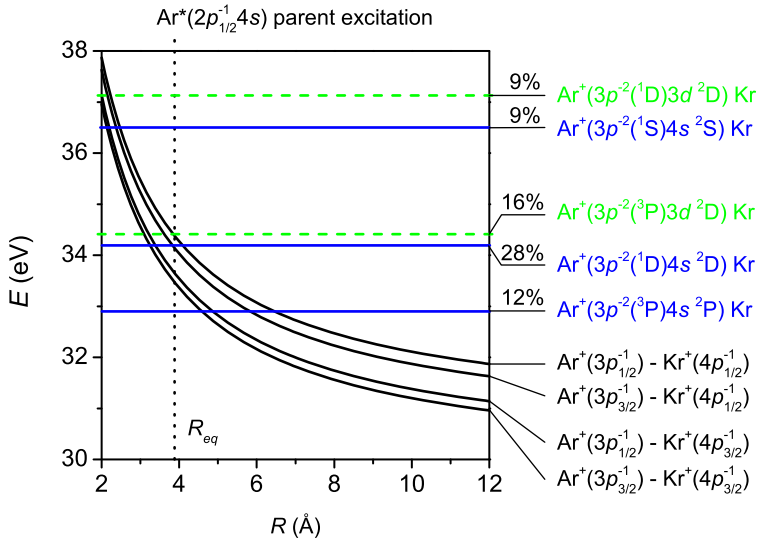
A large number of interatomic decay channels will open if a high-Z Auger electron emitter is placed in an environment as complex as DNA and its solvation shell. At each step of the cascade, the probability of a core-ICD process occurring can become considerable owing to the large number of neighbours and the high charge accumulated

on the parent high-Z element. Each Coster–Kronig-like core-ICD transition contributes an additional ICD electron. All together, the total number of ICD electrons in such a cascade can be substantial.

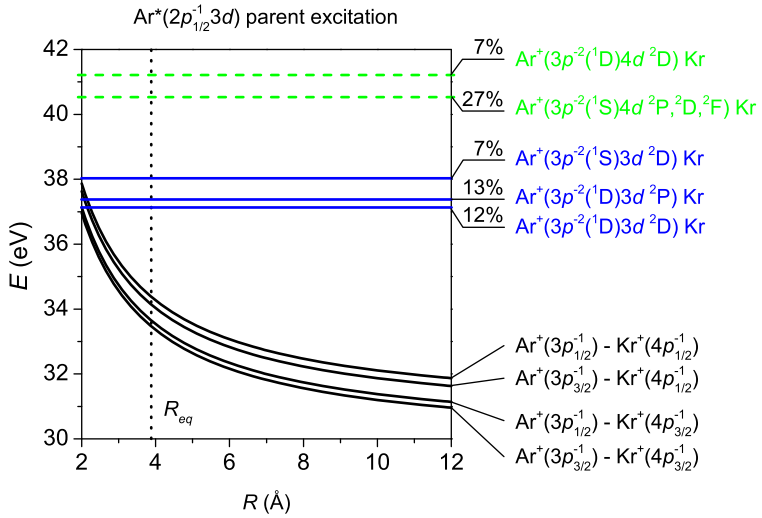
Another possible source of genotoxic electrons is ICD processes triggered by electron impact. It is generally accepted within the radiation biology community that the genotoxic electrons are those with energies below 500 eV (see, for example, ref. 23). Electrons with energies below 15 eV have been shown to generate DNA lesions by means of dissociative electron attachment[29,48], but how electrons with higher energies induce DNA strand breakage remains an open question. A plausible scenario is that these electrons further ionize the environment (water or the DNA itself) and trigger more ICD events, thus producing additional low-energy electrons. Indeed, electron scattering experiments have shown that inner-valence states can be efficiently ionized by electron impact: in the case of water, the ionization cross-section for the inner-valence $2a_1$ orbital by electrons with an incident energy of 250 eV is larger than the cross-section for ionization of all three outer valence orbitals together[49], with experiments indicating that this holds also for other molecules[50]. These inner-valence ionized (or excited) states produced by electron impact can then efficiently decay by ICD ionizing their environment[2], suggesting that the genotoxic effect of high-Z elements used as Auger emitters is at least partly due to follow-up ICD processes initiated by the Auger electrons.

We note that a similar idea has been discussed[51] to explain the effect of secondary electrons with energies above about 35 eV produced by heavy-ion impact, which can efficiently ionize the $2a_1$ orbital of a water molecule in the DNA solvation shell and thereby trigger an ICD process. The resultant simultaneous presence of three slow electrons (secondary, inner-valence ionized and ICD) in the vicinity of a DNA was suggested[51] to be very effective in inducing strand breakages. We therefore conclude that, irrespective of whether energy is deposited in the system by photons, electrons or ions, ICD will be triggered and will contribute to the DNA damage.

31. Hausamann, D. & Morgner, H. The heteronuclear rare gas ions. A simple model for the determination of the potential curves. *Mol. Phys.* **54,** 1085–1099 (1985).
32. Stoychev, S. D., Kuleff, A. I., Tarantelli, F. & Cederbaum, L. S. On the interatomic electronic processes following Auger decay in neon dimer. *J. Chem. Phys.* **129,** 074307 (2008).
33. Demekhin, Ph. V. *et al.* Interatomic Coulombic decay in NeAr following K-LL Auger transition in the Ne atom. *J. Chem. Phys.* **131,** 104303 (2009).
34. National Institute of Standards and Technology. *NIST Atomic Spectra Database* http://physics.nist.gov/asd (2012).
35. Averbukh, V. & Cederbaum, L. S. Ab initio calculation of interatomic decay rates by a combination of the Fano ansatz, Green's-function methods, and the Stieltjes imaging technique. *J. Chem. Phys.* **123,** 204107 (2005).
36. Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124,** 1866–1878 (1961).
37. Schirmer, J., Trofimov, A. B. & Stelter, G. A non-Dyson third-order approximation scheme for the electron propagator. *J. Chem. Phys.* **109,** 4734–4744 (1998).
38. Hasi, A. U. in *Electron-Molecule and Photon-Molecule Collisions* (eds Rescigno, T., McKoy, V. & Schneider, B.) 281–298 (Plenum, 1979).
39. Peterson, K. A., Figgen, D., Goll, E., Stoll, H. & Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16-18 elements. *J. Chem. Phys.* **119,** 11113–11123 (2003).
40. Woon, D. E. & Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **98,** 1358–1371 (1993).
41. Wilson, A. K., Woon, D. E., Peterson, K. A. & Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. *J. Chem. Phys.* **110,** 7667–7676 (1999).
42. Santra, R., Zobeley, J. & Cederbaum, L. S. Electronic decay of valence holes in clusters and condensed matter. *Phys. Rev. B* **64,** 245104 (2001).
43. Öhrwall, G. *et al.* Femtosecond interatomic Coulombic decay in free neon clusters: large lifetime differences between surface and bulk. *Phys. Rev. Lett.* **93,** 173401 (2004).
44. Müller, I. B. & Cederbaum, L. S. Ionization and double ionization of small water clusters. *J. Chem. Phys.* **125,** 204305 (2006).
45. Stoychev, S. D., Kuleff, A. I. & Cederbaum, L. S. On the intermolecular Coulombic decay of singly and doubly ionized states of water dimer. *J. Chem. Phys.* **133,** 154307 (2010).
46. Ottosson, N., Öhrwall, G. & Björneholm, O. Ultrafast charge delocalization dynamics in aqueous electrolytes: new insights from Auger electron spectroscopy. *Chem. Phys. Lett.* **543,** 1–11 (2012).
47. Öhrwall, G. *et al.* Charge dependence of solvent-mediated intermolecular Coster-Kronig decay dynamics of aqueous ions. *J. Phys. Chem. B* **114,** 17057–17061 (2010).
48. Huels, M. A., Boudaïffa, B., Cloutier, P., Hunting, D. & Sanche, L. Single, double, and multiple double strand breaks induced in DNA by 3-100 eV electrons. *J. Am. Chem. Soc.* **125,** 4467–4477 (2003).
49. Milne-Brownlie, D. S. *et al.* Dynamics in electron-impact ionization of H_2O. *Phys. Rev. A* **69,** 032701 (2004).
50. McCarthy, I. E. & Weigold, E. Electron momentum spectroscopy of atoms and molecules. *Rep. Prog. Phys.* **54,** 789–879 (1991).
51. Surdutovich, E. & Solov'yov, A. V. Multiscale physics of ion-beam cancer therapy. *J. Phys. Conf. Ser.* **373,** 012001 (2012).



Extended Data Figure 1 | Model potential energy curves of the initial and final ICD states of ArKr produced on excitation at 246.51 eV. The horizontal lines indicate the potential energy curves of the excited valence-ionized states produced through the resonant Auger decay of the parent state following Ar($2p_{1/2}^{-1}4s$) core excitation at 246.51 eV. The steep curves indicate the potential energy of the two-site doubly ionized final states obtained after ICD. The relative populations of the final resonant Auger states are given in per cent. Only states acquiring more than 5% of the total population are depicted. The equilibrium distance of the neutral ArKr (R_{eq} = 3.88 Å) is shown as a vertical dotted line.



Extended Data Figure 2 | Model potential energy curves of the initial and final ICD states of ArKr produced on excitation at 246.93 eV. The horizontal lines indicate the potential energy curves of the excited valence-ionized states produced through the resonant Auger decay of the parent state following $Ar(2p_{3/2}^{-1}3d)$ core excitation at 246.93 eV. The steep curves indicate the potential energy of the two-site doubly ionized final states obtained after ICD. The relative populations of the final resonant Auger states are given in per cent. Only states acquiring more than 5% of the total population are depicted. The equilibrium distance of the neutral ArKr ($R_{eq} = 3.88$ Å) is shown as a vertical dotted line.

LETTER

doi:10.1038/nature12927

Resonant Auger decay driving intermolecular Coulombic decay in molecular dimers

F. Trinter[1], M. S. Schöffler[1,2], H.-K. Kim[1], F. P. Sturm[1,2], K. Cole[1], N. Neumann[1], A. Vredenborg[1], J. Williams[3], I. Bocharova[2], R. Guillemin[4], M. Simon[4], A. Belkacem[2], A. L. Landers[3], Th. Weber[2], H. Schmidt-Böcking[1], R. Dörner[1] & T. Jahnke[1]

In 1997, it was predicted[1] that an electronically excited atom or molecule placed in a loosely bound chemical system (such as a hydrogen-bonded or van-der-Waals-bonded cluster) could efficiently decay by transferring its excess energy to a neighbouring species that would then emit a low-energy electron. This intermolecular Coulombic decay (ICD) process has since been shown to be a common phenomenon[2–12], raising questions about its role in DNA damage induced by ionizing radiation, in which low-energy electrons are known to play an important part[13,14]. It was recently suggested[15] that ICD can be triggered efficiently and site-selectively by resonantly core-exciting a target atom, which then transforms through Auger decay into an ionic species with sufficiently high excitation energy to permit ICD to occur. Here we show experimentally that resonant Auger decay can indeed trigger ICD in dimers of both molecular nitrogen and carbon monoxide. By using ion and electron momentum spectroscopy to measure simultaneously the charged species created in the resonant-Auger-driven ICD cascade, we find that ICD occurs in less time than the 20 femtoseconds it would take for individual molecules to undergo dissociation. Our experimental confirmation of this process and its efficiency may trigger renewed efforts to develop resonant X-ray excitation schemes[16,17] for more localized and targeted cancer radiation therapy.

The experiment presented here shows that resonant excitation of a K-shell electron to a bound state is followed by Auger decay to an ionic species that can then undergo ICD, as sketched in Fig. 1 and proposed in ref. 15. The initial resonant excitation of the electron occurs as in the experiments that probed resonant interatomic Coulombic decay[5,6], but the state undergoing ICD is created after partial de-excitation of the system through a local Auger decay. The Auger decay can lead to the ground state of the molecular ion through 'participator Auger decay', although in many cases the excited electron will act as just a 'spectator' to an Auger decay in which an electron from the valence or inner valence shell fills the core hole and a second electron from the valence shell is emitted. This spectator pathway produces ionic states which are high enough in excitation energy to allow ICD to occur, and in the case of carbon monoxide accounts for the decay of approximately 75% of core-excited molecules[18]. Our experiment explores the overall scenario for two simple model systems—clusters of just two carbon monoxide or two nitrogen molecules—that can be investigated in great detail. This allows us to follow the Auger decay occurring after resonant excitation of an inner-shell electron into the lowest unoccupied molecular orbital (in a Π* excitation) and the subsequent ICD:

$$h\nu + N_2/N_2 \rightarrow N_2^*(1\,s^{-1}\,\Pi^*)/N_2 \quad \text{(local, resonant excitation)} \quad (1)$$

$$N_2^*(1\,s^{-1}\,\Pi^*)/N_2 \rightarrow N_2^{+}{}^*/N_2 + e_{Auger} \quad \text{(spectator Auger decay)} \quad (2)$$

$$N_2^{+}{}^*/N_2 \rightarrow N_2^{+} + N_2^{+} + e_{ICD} \quad \text{(ICD + two-site Coulomb explosion)} \quad (3)$$

where $h\nu$ is the incident radiation, Π* is the excited molecular orbital, and e_{Auger} and e_{ICD} are the Auger- and ICD-emitted electrons ('1 s^{-1}' refers to a K-shell electron being removed during excitation). Figure 2a, b shows, for $(CO)_2$ and $(N_2)_2$, the correlation between the kinetic energy release of the two molecular ions and the kinetic energy of the electrons measured in coincidence. Unlike in similar plots for ICD in rare-gas dimers[11], no discrete structures are observed in Fig. 2. This is a direct consequence of the repulsive nature of the intermediate state populated by the resonant Auger decay and of the vibrational and rotational degrees of freedom of the ionic fragments. The resonant Auger decay onto a repulsive state of the molecule leads to a continuum

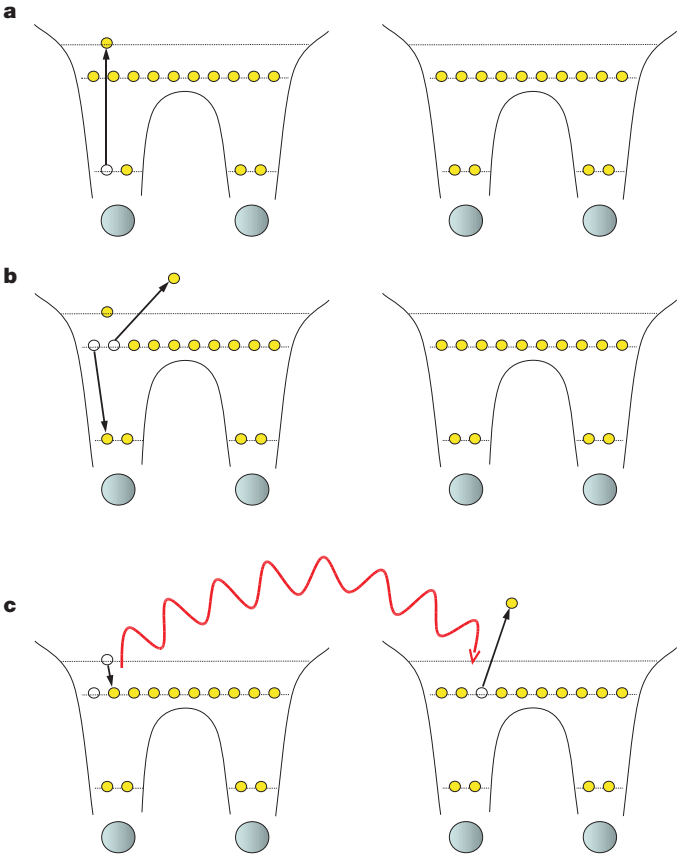


Figure 1 | The overall decay cascade mechanism. Shown is the series of events involved in resonant-Auger-driven ICD (see equations (1)–(3)). **a**, One molecule (left) of the molecular dimer is core-excited. **b**, The core-excited state decays by a spectator Auger decay to a highly excited state of the molecular ion. **c**, ICD transfers the excitation energy to the molecular neighbour (right), where a low-energy ICD electron is emitted.

[1]Institut für Kernphysik, Goethe-Universität, Max-von-Laue-Strasse 1, 60438 Frankfurt am Main, Germany. [2]Lawrence Berkeley National Laboratory, Chemical Sciences Division, Berkeley, California 94720, USA. [3]Department of Physics, Auburn University, Auburn, Alabama 36849, USA. [4]UPMC and CNRS, UMR 7614, Laboratoire de Chimie Physique Matière et Rayonnement, 75005 Paris, France.

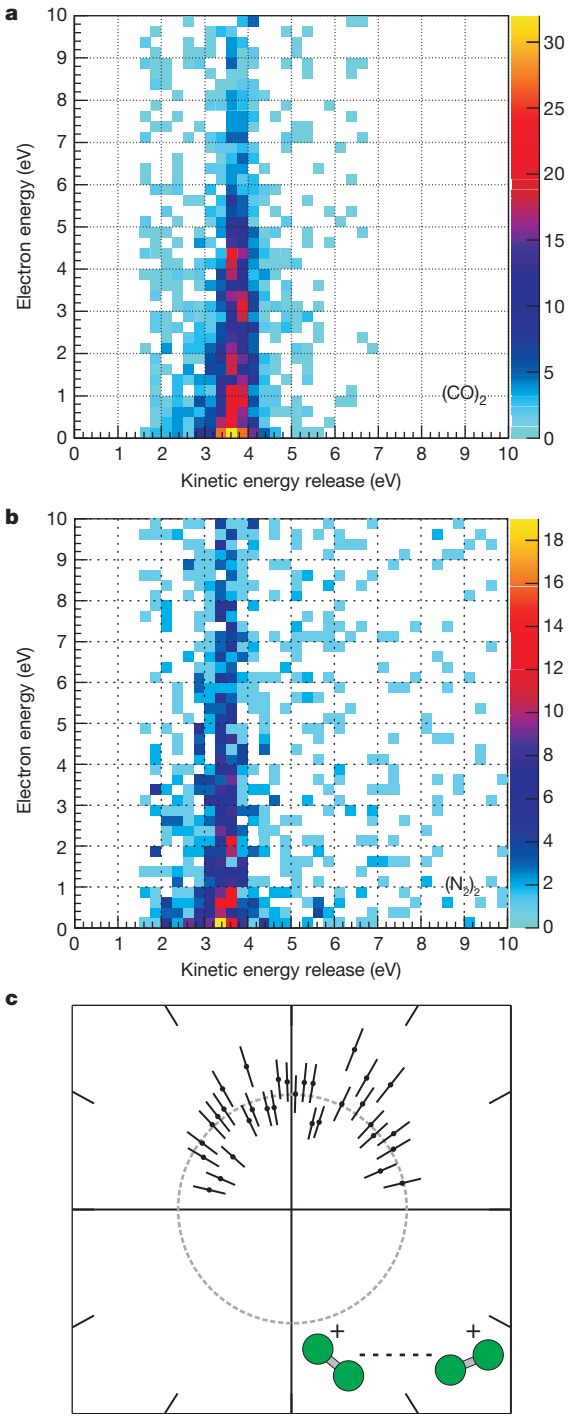


Figure 2 | Experimental results. **a**, Kinetic energy release of $(CO)_2$ versus energy of one of the two electrons created by ICD after resonant excitation and subsequent Auger decay at a photon energy of 287.4 eV (Π* excitation of CO). The colour scale shows the intensity in counts. **b**, Same plot for $(N_2)_2$ recorded at a photon energy of 401.9 eV (Π* excitation of N_2). **c**, Emission direction of the Auger electron with respect to the molecular axis of the N_2 dimer (with statistical error bars). The dimer is oriented horizontally, as depicted by the green icon. The grey circle is a line to guide the eye, corresponding to isotropic emission.

of Auger energies and hence to a continuum of excitation energies of the intermediate N_2^+*/N_2 (or CO^+*/CO) state.

The data in Fig. 2a, b are obtained from the detection of two singly charged molecular ions, revealing the Coulomb explosion of the molecular dimer as the terminal step of ICD. The actual kinetic energy release values measured strongly support the picture of ICD being the underlying process: the maximum of the observed kinetic energy release distribution is 3.4 eV for $(N_2)_2$ and 3.7 eV for $(CO)_2$, which compare fairly well with the values of 3.57 eV and 3.69 eV estimated by calculating the kinetic energy of two singly charged point particles starting a Coulomb explosion at the typical mean intermolecular distance of the CO and N_2 dimers, respectively. (This estimate assumes a simple Coulomb potential, which is a good approximation for Van-der-Waals-bound systems, so that the kinetic energy release in atomic units is given by $1/R$, with R the distance between the two charges. For R we use the dimer intermolecular distance, reported[19] to be 4.03 Å for $(N_2)_2$, whereas the mean of the values reported[20,21] for the CO dimer is 3.9 Å.) These findings support the scenario of an intermolecular decay mechanism such as ICD.

An alternative mechanism that would also lead to two singly charged molecular ions being launched at the mean intermolecular distance of the dimer in its ground state is intermolecular 'knockout', as recently observed[22] in He_2. It occurs if the fast Auger electron is emitted in the direction of the neighbouring molecule and knocks out an electron, thereby ionizing the neighbouring molecule. This process yields the two molecular ions observed: an Auger electron of reduced energy and a low-energy electron. However, this process happens only[22] if the neighbouring molecule is located in the direction of emission of the Auger electron. In our experiments, the orientation of the dimer in space at the instant of Coulomb explosion is known from the coincidence momentum measurement, whereas the direction of the fast Auger electron is measured for every event via its recoil on the centre of mass of the two Coulomb-exploding molecular ions. From this we obtain the Auger electron angular emission pattern with respect to the molecular axis of the dimer in Fig. 2c, which is nearly isotropic and thus eliminates the intermolecular knockout scenario (which would lead to an emission pattern strongly directed along the molecular axis).

Our experimental data thus reveal ICD to be a prominent decay channel for excited dimers of N_2 and CO. ICD might have a substantial (yet poorly studied) effect on the fate of an excited molecular ion: excited molecular ions are rarely stable and dissociate if they are present as isolated species, but might survive as stable entities in solution or in any other chemical environment where ICD can occur. The competition between the ICD (equation (3)) and energy relaxation through dissociation without release of an electron is crucial:

$$N_2^+*/N_2 \rightarrow N^+ + N + N_2 \qquad \text{(one-site dissociation)} \qquad (4)$$

ICD can thus effectively suppress dissociation if it occurs quickly enough. Alternatively, the inverse might also be true: ICD in a molecular system (equation (3)) might become a rare phenomenon if one-site dissociation (equation (4)) is a very fast competing channel. The CO^+* potential energy curves above the single ionization potential of CO are all steeply repulsive[23], but the fact that we observe ICD and a breakup of the molecular dimer into CO^+/CO^+ shows that ICD nonetheless outpaces dissociation. This allows us to use the molecular dissociation as a clock to obtain an estimate of the timescale on which ICD occurs in the present case. The typical slopes of the potential energy curves involved are 10 eV $Å^{-1}$. From this and the fact that we observe only ICD events where the CO^+ does not fragment, we can estimate the maximum time which could have elapsed before the molecular ion has to relax via ICD. For a repulsive state the potential energy of the system decreases as the molecule dissociates. Therefore, for a given potential energy surface we can calculate how long it will take before an internuclear distance has been reached at which the potential energy has dropped below the threshold for ICD. In the case of the molecules and states populated here, this time is less than 20 fs. Accordingly, ICD must occur on a timescale shorter than 20 fs.

ICD has been discussed[15] in the context of radiation biology and also cancer radiotherapy, which still usually uses broadband irradiation of biological tissue to destroy cancerous cells, with considerable adverse

side effects. Tagging cancerous cells with molecular markers containing at least one atom of a high-Z element for resonant excitation by energetically well-defined X-rays, in order to localize the radiation damage to the required site inside a biological system while leaving surrounding tissue unaffected[16], is thus an attractive prospect. In principle, because a resonant excitation of the kind described here is typically ten times stronger than the non-resonant ionization used by broadband irradiation, the overall radiation dose can be minimized using monochromatized X-rays tuned to a suitable resonance. ICD offers the added advantage of directly generating low-energy electrons that are known to be genotoxic and thus are effective mediators of the anticancer effects of radiotherapy. The present decay cascade, with ICD occurring efficiently after resonant excitation of a selected atom and its subsequent Auger decay, is particularly attractive, because it is possible to target a specific site in a larger system at which ICD and the emission of genotoxic low-energy electrons should take place. We expect that our experimental validation of this process, and other studies published during review of this contribution that also confirm its existence and even the tunability of the ICD electron energy in rare-gas clusters[24–26], will stimulate further exploration of Auger-electron-driven cancer therapy.

METHODS SUMMARY

We use cold target recoil ion momentum spectroscopy (COLTRIMS)[27,28] to measure in coincidence all charged particles created in a single reaction, using beamline 11.0.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory. The N_2 and CO dimers were produced by expanding the gas through a 30-μm nozzle at a stagnation pressure of 10 bar. The nozzle was cooled to approximately 140 K to enhance dimer production. The supersonic beam was collimated by a set of two skimmers and then crossed with the photon beam inside a COLTRIMS spectrometer[29]. An electric field of 7.4 V cm^{-1} and a parallel magnetic field of 7.0 gauss guided electrons and ions to two position-sensitive micro-channel plate detectors with delay line readout (see http://www.roentdek.com for details on the detectors). The fields were adjusted such that electrons of up to 15 eV kinetic energy could be collected with a 4π solid angle. Owing to the long ion drift arm the spectrometer accepted only N_2^+ or CO^+ (in case their kinetic energies were higher than 5 eV), which were emitted within 10° with respect to the spectrometer axis. The light polarization was circular in the case of N_2. The data were recorded in list mode. For each ionization event we recorded the positions of impact and times of flight of all registered particles. This allowed us to extract the very weak dimer signal from our data, because the dimer fraction in our beam was only 0.1% to 1%. Thus most recorded ions and electrons resulted from ionization of the monomer. We can identify the ICD channel by selecting only events in which two N_2^+ (or two CO^+) ions with equal and opposite momentum occur. This back-to-back emission is a unique signature of the final step of Coulomb explosion following ICD (see equation (3)).

Received 8 May 2012; accepted 28 November 2013.
Published online 22 December 2013.

1. Cederbaum, L. S., Zobeley, J. & Tarantelli, F. Giant intermolecular decay and fragmentation of clusters. *Phys. Rev. Lett.* **79,** 4778–4781 (1997).
2. Jahnke, T. *et al.* Ultrafast energy transfer between water molecules. *Nature Phys.* **6,** 139–142 (2010).
3. Mucke, M. *et al.* A hitherto unrecognized source of low-energy electrons in water. *Nature Phys.* **6,** 143–146 (2010).
4. Kim, H.-K. *et al.* Enhanced production of low energy electrons by alpha particle impact. *Proc. Natl Acad. Sci. USA* **108,** 11821–11824 (2011).
5. Barth, S. *et al.* Observation of resonant interatomic Coulombic decay in Ne clusters. *J. Chem. Phys.* **122,** 241102 (2005).
6. Aoto, T. *et al.* Properties of resonant interatomic Coulombic decay in Ne dimers. *Phys. Rev. Lett.* **97,** 243401 (2006).
7. Jahnke, T. *et al.* Experimental separation of virtual photon exchange and electron transfer in interatomic Coulombic decay of neon dimers. *Phys. Rev. Lett.* **99,** 153401 (2007).
8. Ueda, K. *et al.* Interatomic Coulombic decay following the Auger decay: experimental evidence in rare-gas dimers. *J. Electron Spectrosc. Relat. Phenom.* **166–167,** 3–10 (2008).
9. Grieves, G. A. & Orlando, T. M. Intermolecular Coulomb decay at weakly coupled heterogeneous interfaces. *Phys. Rev. Lett.* **107,** 016104 (2011).
10. Marburger, S., Kugeler, O., Hergenhahn, U. & Möller, T. Experimental evidence for interatomic Coulombic decay in Ne clusters. *Phys. Rev. Lett.* **90,** 203401 (2003).
11. Jahnke, T. *et al.* Experimental observation of interatomic Coulombic decay in neon dimers. *Phys. Rev. Lett.* **93,** 163401 (2004).
12. Öhrwall, G. *et al.* Femtosecond interatomic Coulombic decay in free neon clusters: large lifetime differences between surface and bulk. *Phys. Rev. Lett.* **93,** 173401 (2004).
13. Boudaïffa, B., Cloutier, P., Hunting, D., Huels, M. A. & Sanche, L. Resonant formation of DNA strand breaks by low-energy (3 to 20 eV) electrons. *Science* **287,** 1658–1660 (2000).
14. Hanel, G. *et al.* Electron attachment to uracil: effective destruction at subexcitation energies. *Phys. Rev. Lett.* **90,** 188104 (2003).
15. Gokhberg, K., Kolorenč, P., Kuleff, A. I. & Cederbaum, L. S. Site- and energy-selective slow-electron production through intermolecular Coulombic decay. *Nature* http://dx.doi.org/10.1038/nature12936 (this issue).
16. Pradhan, A. K. *et al.* Resonant X-ray enhancement of the Auger effect in high-Z atoms, molecules, and nanoparticles: potential biomedical applications. *J. Phys. Chem. A* **113,** 12356–12363 (2009).
17. Howell, R. W. Auger processes in the 21st century. *Int. J. Radiat. Biol.* **84,** 959–975 (2008).
18. Botting, S. K. & Lucchese, R. R. Auger decay of the C 1s -> 2π* excitation of CO. *Phys. Rev. A* **56,** 3666–3674 (1997).
19. Aquilanti, V., Bartolomei, M., Cappelletti, D., Carmona-Novillo, E. & Pirani, F. The N_2–N_2 system: an experimental potential energy surface and calculated rotovibrational levels of the molecular nitrogen dimer. *J. Chem. Phys.* **117,** 615 (2002).
20. Havenith, M., Petri, M., Lubina, C., Hilpert, G. & Urban, W. J. IR spectroscopy of $(CO)_2$ using concentration-frequency double modulation in a supersonic jet expansion. *J. Mol. Spectrosc.* **167,** 248–261 (1994).
21. Meredith, A. W. & Stone, A. J. An ab initio and diffusion Monte Carlo study of the potential energy surface of the CO dimers. *J. Phys. Chem. A* **102,** 434–445 (1998).
22. Havermeier, T. *et al.* Single photon double ionization of the helium dimer. *Phys. Rev. Lett.* **104,** 153401 (2010).
23. Baltzer, P. *et al.* Inner-valence states of CO^+ between 22 eV and 46 eV studied by high resolution photoelectron spectroscopy and ab initio CI calculations. *J. Phys. At. Mol. Opt. Phys.* **27,** 4915 (1994).
24. Kimura, M. *et al.* Efficient site-specific low-energy electron production via interatomic Coulombic decay following resonant Auger decay in argon dimers. *Phys. Rev. A* **87,** 043414 (2013).
25. O'Keeffe, P. *et al.* The role of the partner atom and resonant excitation energy in interatomic Coulombic decay in rare gas dimers. *J. Phys. Chem. Lett.* **4,** 1797–1801 (2013).
26. Kimura, K. *et al.* Controlling low-energy electron emission via resonant-auger-induced interatomic Coulombic decay. *J. Phys. Chem. Lett.* **4,** 1838–1842 (2013).
27. Dörner, R. *et al.* Cold target ion momentum spectrocopy: a 'momentum microscope' to view atomic collisions dynamics. *Phys. Rep.* **330,** 95–192 (2000).
28. Ullrich, J. *et al.* Recoil-ion and electron momentum spectroscopy: reaction-microscopes. *Rep. Prog. Phys.* **66,** 1463 (2003).
29. Schöffler, M. S. *et al.* Matter wave optics perspective at molecular photoionization: K-shell photoionization and Auger decay of N_2. *New J. Phys.* **13,** 095013 (2011).

Acknowledgements This work was supported by the Deutsche Forschungsgemeinschaft and the Deutscher Akademischer Austauschdienst. We thank the staff of the Advanced Light Source for excellent support during the beam time. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, and by the Division of Chemical Sciences, Geosciences, and Biosciences of the US Department of Energy at the Lawrence Berkeley National Laboratory under contract number DE-AC02-05CH11231. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract number DE-AC02-05CH11231. We thank K. Gokhberg and L. Cederbaum for suggesting this experiment and for many discussions. M.S.S. thanks the Alexander von Humboldt foundation for financial support.

Author Contributions All authors contributed to the experiment. F.T. and T.J. performed the data analysis. All authors contributed to 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 T.J. (jahnke@atom.uni-frankfurt.de) or F.T. (trinter@atom.uni-frankfurt.de).

LETTER

doi:10.1038/nature12882

Australian tropical cyclone activity lower than at any time over the past 550–1,500 years

Jordahna Haig[1], Jonathan Nott[1] & Gert-Jan Reichart[2,3]

The assessment of changes in tropical cyclone activity within the context of anthropogenically influenced climate change has been limited by the short temporal resolution of the instrumental tropical cyclone record[1,2] (less than 50 years). Furthermore, controversy exists regarding the robustness of the observational record, especially before 1990[3–5]. Here we show, on the basis of a new tropical cyclone activity index (CAI), that the present low levels of storm activity on the mid west and northeast coasts of Australia are unprecedented over the past 550 to 1,500 years. The CAI allows for a direct comparison between the modern instrumental record and long-term palaeotempest (prehistoric tropical cyclone) records derived from the $^{18}O/^{16}O$ ratio of seasonally accreting carbonate layers of actively growing stalagmites. Our results reveal a repeated multicentennial cycle of tropical cyclone activity, the most recent of which commenced around AD 1700. The present cycle includes a sharp decrease in activity after 1960 in Western Australia. This is in contrast to the increasing frequency and destructiveness of Northern Hemisphere tropical cyclones since 1970 in the Atlantic Ocean[6–8] and the western North Pacific Ocean[6,7]. Other studies project a decrease in the frequency of tropical cyclones towards the end of the twenty-first century in the southwest Pacific[7,9], southern Indian[9,10] and Australian[11] regions. Our results, although based on a limited record, suggest that this may be occurring much earlier than expected.

Trend analysis of the instrumental tropical cyclone record has proven difficult owing to errors associated with changes in observational techniques (leading to inaccurate intensity estimates and storm counts in the recent past), detection issues, data homogeneity issues[3,6,12] and inconsistent procedures between and within agencies[1,3–5]. As a result, differentiating natural variability from anthropogenically induced change is complicated; this may also explain to a certain extent the disparity between current trend estimates[6,13].

In an effort to remedy this we have developed a new technique, which calibrates high-resolution, long-term palaeorecords of tropical cyclone activity against the instrumental tropical cyclone record. This scale allows for a direct comparison between the past and present, and enables an examination of tropical cyclone climatology at higher temporal resolution and on annual, decadal or millennial scales simultaneously, without the need to interpolate or extrapolate to account for missing data. Our index, CAI, is based on tropical cyclone activity indices developed by the National Oceanic and Atmospheric Administration and others, which describe the severity of a season in terms of the number of storms, their intensity (V_{max}), their size (R_{max}) and their longevity. These indices include the accumulated cyclone energy index[14], the revised accumulated cyclone energy index[15], the power dissipation index[6] and the hurricane intensity index[16] (Methods). CAI is the average accumulated energy expended over the tropical cyclone season within range of the site, accounting for the number of days since genesis and the intensity and size of the storm relative to its distance from the site at each point along its track (Fig. 1):

$$\mathrm{CAI} = \frac{1}{N}\sum_{n=1}^{N} K_{n,t}$$

where $K_{n,t} = (K_t + K_{t-1})_n$, $K_t = V_{max}^3(t)R_{max}(t)/d(t)$, N is the number of storms within the season, n enumerates the individual storms, t denotes time along the storm track (recorded at 6-h intervals), $d(t)$ is the distance from the site in kilometres at time t, $V_{max}(t)$ is the maximum 10-min-mean wind speed in metres per second at time t, and $R_{max}(t)$ is the radius of maximum wind in kilometres at time t.

Tropical cyclones produce precipitation that is depleted in the heavier oxygen isotope (^{18}O) by >6‰ relative to average monsoonal precipitation, owing to the recycling of water within the system, high condensation efficiency and large size and longevity of such cyclones as intense convective systems[17]. The resulting $\delta^{18}O$ content (expressed as $\delta^{18}O = [(^{18}O/^{16}O)_{sample}/(^{18}O/^{16}O)_{standard} - 1)] \times 1{,}000‰$) of tropical cyclone precipitation at a site is influenced by a number of factors, including the number of days since genesis (that is, rainout) and the intensity of the storm, its source region[18] and the distance of its centre from the sampling path. Because tropical stalagmites are archives of monsoonal $\delta^{18}O$, signatures of past tropical cyclones are also recorded within the $\delta^{18}O$ of their carbonate layers, typically within 400 km of the storm centre[19,20].

Two cylindrical stalagmites were collected from regions in Queensland and Western Australia prone to tropical cyclones (Chillagoe in Queensland, stalagmite CH-1; Cape Range in Western Australia, stalagmite CR-1). Both show a continuous, uninterrupted record of distinct seasonal growth banding composed of alternating layers of dark and light calcite corresponding to wet- and dry-season deposition. No visible hiatuses were present. The first 1,500 wet-season (dark-calcite) layers were analysed for their carbonate $\delta^{18}O$. Observed differences between maxima and minima in $\delta^{18}O$ over the time period are 4.38‰ (CH-1) and 5.81‰ (CR-1). In both locations, $\delta^{18}O$ is highly variable between wet seasons: yearly differences range from −1.6‰ to 2.08‰ in CH-1 and from −2.5‰ to 2.2‰ in CR-1, which are too large to be explained by a cave temperature effect because this would imply a shift in annual temperatures of 6–8 °C (ref. 21). Neither CR-1 nor CH-1 exhibits a significant relationship between $\delta^{18}O$ and the seasonal rainfall total, the annual rainfall total or the number of rain days at the corresponding site ($\rho < 0.07$ (Spearman's rho), $P > 0.5$ for CR-1; $\rho < -0.08$, $P > 0.2$ for CH-1). In the absence of cave temperature or rainfall 'amount effects' we conclude that rainfall composition rather than cave temperature and rainfall amount or frequency, or both, influences the resulting $\delta^{18}O$. However, periods of non-tropical cyclone rainfall and changes in the strength of the Australian–Indonesian monsoon are expected to dilute the cave reservoir. Stalagmite monsoon records from latitudes below 8° S (which are therefore less influenced by tropical cyclone activity) show variations of up to 0.7‰–1.2‰ (ref. 22) over a 1,500-year period. These values are considerably less than the 4‰–6‰ variation between the maxima and minima and the 1.6‰–2.5‰ seasonal variation within the stalagmite $\delta^{18}O$ presented here. Nevertheless, we account for the monsoonal contribution to $\delta^{18}O$ using empirical methods for determining the average value of precipitation $\delta^{18}O$ VSMOW (that is, $\delta^{18}O$ where the standard is Vienna standard mean ocean water) at both sites, and we account for centennial scale changes

[1]Earth and Environmental Sciences, James Cook University, Cairns, Queensland 4870, Australia. [2]Department of Geochemistry, Utrecht University, Utrecht 3508 TA, The Netherlands. [3]Geology Department, Royal Netherlands Institute for Sea Research, Den Hoorn (Texel) 1797 SZ, The Netherlands.

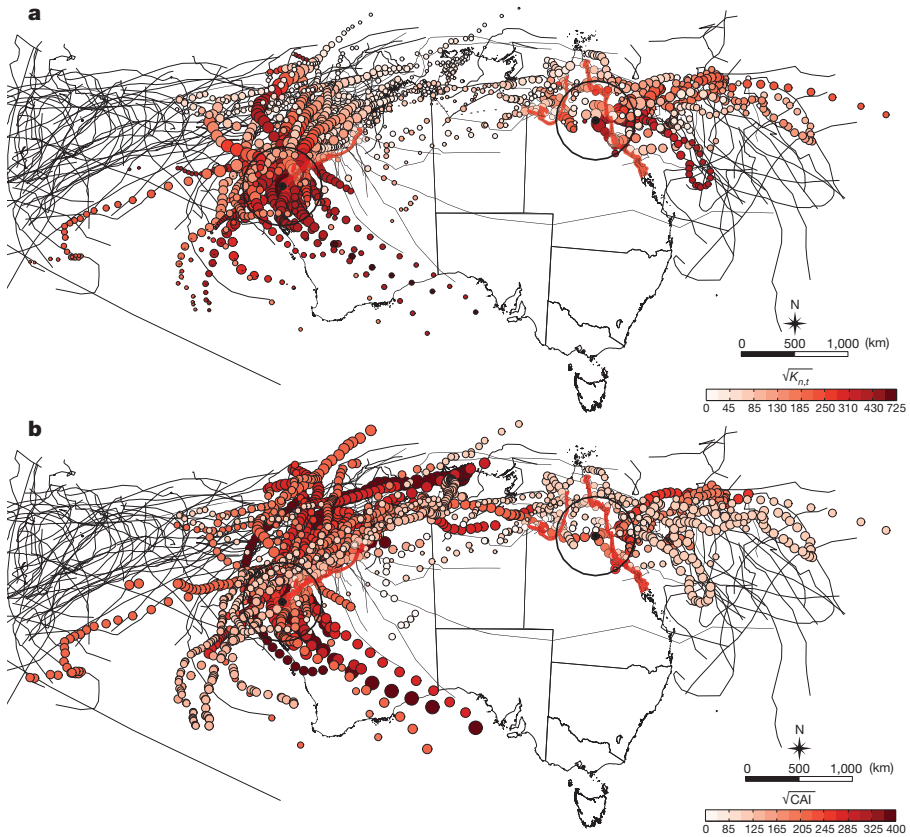


Figure 1 | Site map showing the four-stage calculation of CAI. Chillagoe and Cape Range (black points) are shown with the 400-km radius around each study site. Tropical cyclones whose tracks did not lie within the study area during the training period in Queensland and Western Australia are shown in black. Red shading indicates the coastlines most prone to tropical cyclones in both states. **a**, Tropical cyclones from the 1990–2010 training period and their corresponding K_t value (point size), showing the influence of V_{max}, R_{max} and distance; cumulative $K_{n,t}$ values are shown in colour. **b**, Point size indicates K_n (individual storm averages) calculated from **a** and subsequent seasonal CAI values (gradated colour).

in monsoonal activity using published Australian–Indonesian monsoon records[22] (Methods).

Figure 2 shows the relationship between CAI, calculated from 'best tracks' in the recently updated tropical cyclone database for the region[23], and the de-trended $\delta^{18}O$ (that is, $\delta^{18}O_A$) for the corresponding period 1990–2010. The model predicts CAI well in 63% of cases ($P < 0.001$) within the $\delta^{18}O_A$ range of −6.37‰ to −1.03‰. That being the case, larger negative excursions in $\delta^{18}O$ correlate with higher CAI values. Although this range is representative of the data obtained from the whole series (2,276 measurements in total), $\delta^{18}O_A$ values that fall outside the model range may not be calculated effectively. However, $\delta^{18}O_A$ values exceeded or fell below the range in only 28 or 88 cases, respectively. Of these, only four were more than 1 s.d. outside the range. Each series was standardized before statistical analysis. No patterns are discernable within the residuals and an even spread of error is indicated. The relationship is expressed as follows (where the per mille value of $\delta^{18}O_A$ is meant):

$$CAI = (-40.27\delta^{18}O_A + 43.12)^2$$

Because our CAI–$\delta^{18}O_A$ relationship was developed using best-track records for 1990–2010, our model is not likely to be subject to the degree of intensity bias generated by changes in observational techniques. Nonetheless, when the period of investigation is extended to include 1970–2010, the relationship between CAI and $\delta^{18}O_A$ still holds ($r = -0.5$ (Pearson's correlation coefficient), $P = 0.0001$, $n = 60$) and the average difference in model estimates of CAI is 10^3 that is, less than half the r.m.s.e. of the model. In addition, CAI after 1990 is modelled from $\delta^{18}O$ and is therefore not subject to the same errors noted previously in the pre-1990 instrumental and historical data sets.

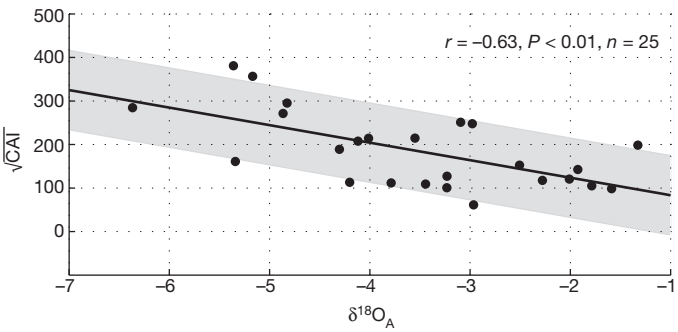


Figure 2 | Calculated CAI versus the de-trended carbonate values from CR-1 and CH-1 ($\delta^{18}O_A$). Grey region indicates the root mean squared error (r.m.s.e.) of the model (difference between actual and modelled CAI values for 1970–2010). $r = 0.63$, $P \leq 0.01$, $n = 25$.

Figure 3a and Fig. 3c give the calculated CAI values over the past 1,500 and 700 years at Cape Range and Chillagoe, respectively. Although it is clear from the analysis of instrumental records that the west coast of Australia is more prone to tropical cyclones than the east coast[24,25], our data indicate that this is not a recent phenomenon. Tropical cyclone activity on the mid west coast of Australia is on average three times higher than on the northeast coast, with CAI values ranging from 10^4 to 1.3×10^5 at Cape Range compared with CAI values of 0.25×10^4 to 1.2×10^5 at Chillagoe. Analysis of CAI indicates that tropical cyclone activity has been highly variable over the past 1,500 years and that tropical cyclone activity in the past was higher than it is today. There has been significantly less tropical cyclone activity at Chillagoe in the past century than in the previous 550 years ($Z = 24.73$ (Z-test statistic),

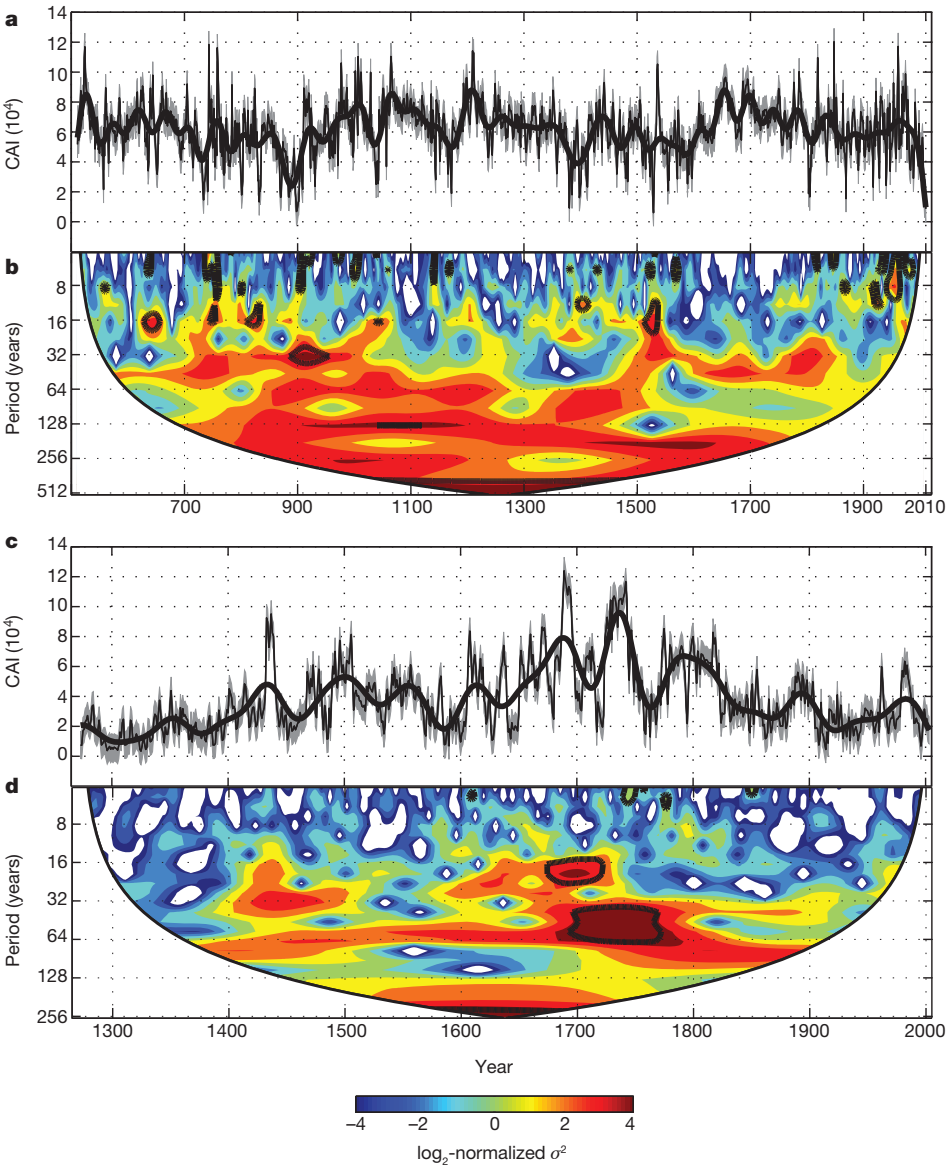


Figure 3 | CAI over the last 1,500 and 700 years. **a**, **c**, Cape Range (**a**) and Chillagoe (**c**); black line indicates smoothing of the series using ref. 31 (smoothed data were not used in the statistical analysis). Grey shading indicates the r.m.s.e. of the model. Four values, which were more than 1 s.d. outside the $\delta^{18}O_A$ range specified in Fig. 1, were removed from the series. **b**, **d**, Wavelet power spectra (Morlet wavelet) of Cape Range (**b**) and Chillagoe (**d**). Power increases from blue to red, black contours indicate regions above the 95% confidence level, and the white areas are regions subject to edge effects. The spectra have lag-1 autocorrelation coefficients of 0.75 (Cape Range) and 0.78 (Chillagoe). Software provided by C. Torrence and G. Compo (http://atoc.colorado.edu/research/wavelets/).

$P < 0.001$). At Cape Range, tropical cyclone activity since 1970 has been significantly lower than it was during the 1,460 years prior ($Z = 22.42$, $P < 0.001$). Wavelet analysis of the time series (Fig. 3b) indicates a reduction in the variance of CAI between the mid 1800s and today at Cape Range within the 16–128-year band. It also highlights an increase in variability within the 4–8-year band before 1960 (although at relatively low power). Significance testing indicates that the majority of the oscillations occur within the 4–32-year frequency band, although the emergence of a 128-year oscillation is indicated between ~AD 1100 and ~1200 and again between ~1400 and ~1600 (however, there is less than 95% confidence in the evidence for the latter). Figure 3c indicates that, relative to the rest of the time series, the variability at Chillagoe was limited during the period before 1400 and after 1900. Significant variations in power are evident between 1700 and 1800 within the 16–64-year band, during which time CAI at Chillagoe was highest (Fig. 3d).

We assessed the rate of the decline in activity over the past few centuries at both sites by conducting a Mann–Kendall test in conjunction with a Theil–Sen estimator. Serial correlation was accounted for by removing the lag-1 autoregressive process after computing the lag-1 serial correlation coefficients for both data sets. Chillagoe shows a significant decline in activity towards the present day since AD 1743 ($\tau = -0.4$ (Kendall's tau), $P < 0.001$, $n = 262$); similarly, an overall decline in activity is seen at Cape Range since 1650 ($\tau = -0.4$, $P < 0.001$, $n = 360$). A more abrupt decline at Cape Range since 1960 is evident in Fig. 3. We assessed this period in relation to the rest of the Cape Range record using a sliding window of 50 years with a 1-year step. The significant downward trend since 1960 is unprecedented in the Cape Range record ($\tau = -0.5$, $P < 0.001$, $n = 50$). These factors in conjunction suggest that the modern instrumental era (1970–2010) provides a poor reflection of the true natural variability of tropical cyclone activity in both regions.

Trend analysis of instrumental data globally has shown a reduction in frequency in all basins[26] (excluding the Atlantic[6,27]) but in many cases an increase in the number and proportion of severe tropical cyclones[26]. Within the Australian region, the downward trend in tropical cyclone

activity over the past 30 years on the east[28] and west coasts[25] is in contrast to reports of no trend[4]. It has been suggested that the downward trend noted in the former studies is probably due to an improvement in the ability to differentiate tropical cyclones from severe tropical lows and to the greater number of El Niño events since 1970[29].

We performed a Mann–Kendall trend test within both the Chillagoe and the Cape Range data sets on the timescales used in the studies referred to above. Our results agree with those of ref. 4 in that no significant trends in tropical cyclone activity in central Western Australia are indicated within the period 1980–2007 ($\tau = -0.2$, $P = 0.103$); however, a significant decrease in activity is evident when the period of investigation is extended beyond the past 30 years (as previously noted). Similarly, our results are also in agreement with those of an analysis of the Eastern Australian region from 1870–2010, showing a distinct downward trend in tropical cyclone activity in northeast Queensland[28] ($\tau = -0.4$, $P < 0.001$).

The Australian region seems to be experiencing the most pronounced phase of tropical cyclone inactivity for the past 550–1,500 years. The dramatic reductions in activity since the industrial revolution suggest that climate change cannot be ruled out as a causative factor. This reduction is also in line with present projections for the late twenty-first century from global climate models, yet our results suggest that this is occurring much sooner than expected. However, we cannot say whether this downward trend in activity will be sustained.

We anticipate that CAI will be a starting point for more sophisticated analysis of other palaeotempest records from around the globe as potential inputs for regional or global climate models and long-range statistical or dynamical forecast models. Deriving a scale of tropical cyclone activity from established high-resolution climate palaeorecords such as stalagmites makes it possible to examine tropical cyclone activity on multiple temporal scales in conjunction with other climate indices, such as temperature, atmospheric CO_2 concentrations, El Niño/Southern Oscillation, the Madden–Julian Oscillation and the Dipole Mode Index. Therefore, CAI provides one seamless index allowing for the incorporation of much longer tropical cyclone records into climate or forecasting models. CAI could be calculated from other stalagmite records and potentially other palaeotempest records from other basins when verified against the local instrumental tropical cyclone record using the method presented here. This provides the means to examine not only how tropical cyclone activity has varied as a result of industrialization but also potentially to forecast future trends in tropical cyclone activity under changing climate conditions, given that it is now possible to discern natural variability from anthropogenically induced change.

METHODS SUMMARY

The most recent 1,500 dark-calcite layers representing wet-season deposition were subsampled using a video-controlled micromill. Oxygen and carbon isotope analyses were performed using a Kiel III carbonate device coupled to a Finnigan MAT 253 IRMS. Each calcite sample was reacted with three drops of H_3PO_4 at 70 °C. Replicate analysis of the standard NBS-19 resulted in a standard deviation of 0.04‰ for $\delta^{13}C$ and 0.06‰ for $\delta^{18}O$. All measurements are reported relative to Vienna PeeDee Belemnite (VPDB). To ensure that the isotopes within the calcite had been deposited in equilibrium with the cave drip water, we conducted a Hendy test for equilibrium deposition at 4.09 and 16.2 cm from the apices of CH-1 and CR-1, respectively. Four or five subsamples were milled for each test at 2–5-mm intervals along the growth horizon from the centre of the layer toward the flanks. Both stalagmites pass Hendy's first test for equilibrium[30] because the maximum variation in $\delta^{18}O$ across the layer is less than 0.8‰ (specifically 0.27‰ for CH-1 and 0.61‰ for CR-1) and neither stalagmite exhibits progressive enrichment in either $\delta^{18}O$ or $\delta^{13}C$ towards the flanks.

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 19 June; accepted 14 November 2013.

1. Knutson, T. R. *et al.* Tropical cyclones and climate change. *Nature Geosci.* **3,** 157–163 (2010).
2. Nott, J. & Hayne, M. High frequency of 'super-cyclones' along the Great Barrier Reef over the past 5,000 years. *Nature* **413,** 508–512 (2001).
3. Landsea, C. W., Harper, B. A., Hoarau, K. & Knaff, J. A. Climate change: can we detect trends in extreme tropical cyclones? *Science* **313,** 452–454 (2006).
4. Harper, B. A., Stroud, S. A., McCormack, M. & West, S. A. A review of historical tropical cyclone intensity in north-western Australia and implications for climate change trend analysis. *Aust. Meteorol. Mag.* **57,** 121–141 (2008).
5. Kamahori, H., Yamazaki, N., Mannoji, N. & Takahashi, K. Variability in intense tropical cyclone days in the western North Pacific. *SOLA* **2,** 104–107 (2006).
6. Emanuel, K. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436,** 686–688 (2005).
7. Emanuel, K., Sundararajan, R. & Williams, J. Hurricanes and global warming: results from downscaling IPCC AR4 simulations. *Bull. Am. Meteorol. Soc.* **89,** 347–367 (2008).
8. Elsner, J. B., Kossin, J. P. & Jagger, T. H. The increasing intensity of the strongest tropical cyclones. *Nature* **455,** 92–95 (2008).
9. Oouchi, K. *et al.* Tropical cyclone climatology in a global-warming climate as simulated in a 20 km-mesh global atmospheric model: frequency and wind intensity analyses. *J. Meteorol. Soc. Jpn* **84,** 259–276 (2006).
10. Zhao, M., Held, I. M., Lin, S. J. & Vecchi, G. A. Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Clim.* **22,** 6653–6678 (2009).
11. Abbs, D. *The Impact of Climate Change on the Climatology of Tropical Cyclones in the Australian Region*. CAF Working Paper 11 (CSIRO, 2010).
12. Nicholls, N., Landsea, C. & Gill, J. Recent trends in Australian region tropical cyclone activity. *Meteorol. Atmos. Phys.* **65,** 197–205 (1998).
13. Kossin, J. P. *et al.* A globally consistent reanalysis of hurricane variability and trends. *Geophys. Res. Lett.* **34,** L04815 (2007).
14. Bell, G. D. *et al.* Climate Assessment for 1999. *Bull. Am. Meteorol. Soc.* **81,** 1328 (2000).
15. Yu, J. Y., Chou, C. & Chiu, P. G. A revised accumulated cyclone energy index. *Geophys. Res. Lett.* **36,** L14710 (2009).
16. Kantha, L. Time to replace the Saffir-Simpson hurricane scale? *Eos* **87,** 3–6 (2006).
17. Gedzelman, S., Lawrence, J., Gamache, J. & Black, M. Probing hurricanes with stable isotopes of rain and water vapor. *Mon. Weath. Rev.* **131,** 1112–1127 (2003).
18. Lawrence, J. R., Gedzelman, S. D., Zhang, X. & Arnold, R. Stable isotope ratios of rain and vapor in 1995 hurricanes. *J. Geophys. Res.* **103,** 11381–11400 (1998).
19. Frappier, A. B. *et al.* Stalagmite stable isotope record of recent tropical cyclone events. *Geology* **35,** 111–114 (2007).
20. Nott, J., Haig, J., Neil, H. & Gillieson, D. Greater frequency variability of landfalling tropical cyclones at centennial compared to seasonal and decadal scales. *Earth Planet. Sci. Lett.* **255,** 367–372 (2007).
21. Mühlinghaus, C., Scholz, D. & Mangini, A. Modelling fractionation of stable isotopes in stalagmites. *Geochim. Cosmochim. Acta* **73,** 7275–7289 (2009).
22. Griffiths, M. L. *et al.* Increasing Australian-Indonesian monsoon rainfall linked to early Holocene sea-level rise. *Nature Geosci.* **2,** 636–639 (2009).
23. Bureau of Meteorology. Australian Tropical Cyclone Database. *Previous Tropical Cyclones* (12 March 2012).
24. Liu, K. S. & Chan, J. C. L. Interannual variation of Southern Hemisphere tropical cyclone activity and seasonal forecast of tropical cyclone number in the Australian region. *Int. J. Climatol.* **32,** 190–202 (2012).
25. Hassim, M. E. E. & Walsh, K. J. E. Tropical cyclone trends in the Australian region. *Geochem. Geophys. Geosyst.* **9,** Q07V07 (2008).
26. Webster, P. J., Holland, G. J., Curry, J. A. & Chang, H. R. Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science* **309,** 1844–1846 (2005).
27. Mann, M. E. & Emanuel, K. A. Atlantic hurricane trends linked to climate change. *Eos Trans. AGU* **87,** 233–241 (2006).
28. Callaghan, J. & Power, S. Variability and decline in the number of severe tropical cyclones making land-fall over eastern Australia since the late nineteenth century. *Clim. Dyn.* **37,** 647–662 (2011).
29. Braganza, K. *et al.* in *Climate Science Update: A Report to the 2011 Garnaut Review* (eds Keenan, T. D. & Cleugh, H. A.) 3–16 (Centre for Australian Weather and Climate Research, 2011).
30. Hendy, C. H. The isotopic geochemistry of speleothems–I. The calculation of the effects of different modes of formation on the isotopic composition of speleothems and their applicability as palaeoclimatic indicators. *Geochim. Cosmochim. Acta* **35,** 801–824 (1971).
31. Mann, M. E. Smoothing of climate time series revisited. *Geophys. Res. Lett.* **35,** L16708 (2008).

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

Acknowledgements Funding for this project was provided by the Australian Research Council (ARC) Discovery Projects Scheme (reference number DP0772691) and financial assistance given as part of the Queensland Smart State Initiative. Analytical procedures were conducted at James Cook University and the University of Utrecht. Special thanks are extended to C. Zwart for his input and advice regarding the development of CAI and his editing of this manuscript. Additional thanks are extended to A. van Dijk and L. Lourens for technical help with IRMS analyses.

Author Contributions J.N. collected the two stalagmites, provided funding for the project through the ARC grant and participated in critical discussions on the

development of CAI. G.-J.R. provided access to facilities and advice on the data processing. J.H. helped collect one stalagmite, processed both stalagmites, analysed the data, developed and tested CAI, and wrote the manuscript. 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.H. (jordahna.haig@jcu.edu.au).

METHODS

Analytical procedures. The most recent 1,500 dark-calcite layers representing wet-season deposition were subsampled using a video-controlled micromill. Oxygen and carbon isotope analyses were performed using a Kiel III carbonate device coupled to a Finnigan MAT 253 IRMS. Each calcite sample was reacted with three drops of H_3PO_4 at 70 °C. Replicate analysis of the standard NBS-19 resulted in a standard deviation of 0.04‰ for $\delta^{13}C$ and 0.06‰ for $\delta^{18}O$. All measurements are reported relative to Vienna PeeDee Belemnite (VPDB). To ensure that the isotopes within the calcite had been deposited in equilibrium with the cave drip water, we conducted a Hendy test for equilibrium deposition at 4.09 and 16.2 cm from the apices of CH-1 and CR-1, respectively. Four or five subsamples were milled for each test at 2–5-mm intervals along the growth horizon from the centre of the layer toward the flanks. Both stalagmites pass Hendy's first test for equilibrium[30] because the maximum variation in $\delta^{18}O$ across the layer is less than 0.8‰ (specifically 0.27‰ for CH-1 and 0.61‰ for CR-1) and neither stalagmite exhibits progressive enrichment in either $\delta^{18}O$ or $\delta^{13}C$ towards the flanks.

CAI formulation. We calculate $K_{n,t}$ for each tropical cyclone that passes within 400 km of either of the two sites, at each observation point along its path since genesis. $K_{n,t}$ is cumulative, and so reflects not only the condition of the system at time t but also its history up until that point:

$$K_{n,t} = (K_t + K_{t-1})_n$$

$$K_t = \frac{V_{\max}^3(t) R_{\max}(t)}{d(t)}$$

Here n enumerates the individual storms, t denotes time along the storm track (recorded at 6-h intervals), d denotes the distance from the site in kilometres at time t, $V_{\max}$ is the maximum 10-min-mean wind speed in metres per second at time t, and $R_{\max}$ is the radius of maximum wind in kilometres at time t.

The resulting $\delta^{18}O$ of the stalagmite carbonate is an average of the collective precipitation events over a season, and we therefore average rather than sum the resulting $K_{n,t}$ values. Thus, CAI is the average accumulated energy expended over the tropical cyclone season within range of the site, accounting for the number of days since genesis of the storm and the intensity and size of the storm relative to its distance from the site at each point in time:

$$\mathrm{CAI} = \frac{1}{N}\sum_{n=1}^{N} K_n$$

Here N is the number of storms within the season.

CAI differs from the accumulated cyclone energy index, the revised accumulated cyclone energy index, the power dissipation index and the hurricane intensity index in that it is tailored to reflect the effects of tropical cyclone activity on the resulting $\delta^{18}O$ of the carbonate layers. CAI is location specific (that is, it accounts for the distance between the site and the centre of the storm track) and gives an average of these tropical cyclone events rather than the sum of the total energy expelled within a season. Because tropical cyclone $\delta^{18}O$ precipitation values are radially asymmetrical within a storm (Extended Data Fig. 1), the inclusion of distance in the calculation of K_t has a dampening effect on the resulting K_n value of that storm with increasing distance. As such, a tropical cyclone located 400 km from the study site at $K_{t=1}$, is weighted less than when located 200 km from the site at $K_{t=2}$, given the same $V_{\max}$ and $R_{\max}$. $K_{n,t}$ does not, however, take into account the angle of approach (for example, the parameter d does not take into account the orientation of the system relative to the study site and does not distinguish between approach or retreat of the system).

***K_t* versus $\delta^{18}O$ VSMOW.** In the absence of tropical cyclone rainfall measurements in Australia, to test how well K_t is reflected in the $\delta^{18}O$ of tropical cyclone precipitation we calculated the corresponding K_t values for Hurricane Olivia (a 1994 eastern North Pacific hurricane) using the NHC's updated HURDAT Best Track Database[32]. These were compared against $\delta^{18}O$ VSMOW measurements[33] made at 30-min intervals between 24 and 26 September 1994. The results are plotted in Extended Data Fig 1. We find that $\delta^{18}O$ depletion increases with increasing K_t ($\rho = -0.5$, $P = 0.02$, $n = 25$), supporting our derivation of K_t and, thus, CAI. Within the eyewall (the ring or belt of thunderstorms surrounding the central eye within the radius of maximum wind), $R_{\max}$ is statistically not significant, K_t in Extended Data Fig. 1b is therefore calculated as a function of $V_{\max}$ and distance alone.

De-trending monsoon. Because tropical cyclone rainfall accounts for only 20.05% and, respectively, 17% of the total rainfall at Chillagoe and Cape Range, it is necessary to exclude the average monsoonal component of the stalagmite carbonate ($\delta^{18}O_M$). We estimated $\delta^{18}O_M$ at Chillagoe and Cape Range using.

$$\delta^{18}O\,(‰) = -0.005 \times \mathrm{Longitude}\,(^\circ) - 0.034 \times \mathrm{Latitude}\,(^\circ) - 0.003 \times \mathrm{Altitude}\,(\mathrm{m}) - 4.753 \quad (1)$$

(adjusted $R^2 = 0.79$) and

$$\delta^{18}O = (6.67 \times 10^{-6})P^2 - 0.009P + 0.015 \times \mathrm{Eva} + 0.007 \times \mathrm{Rad} - 9.670 \quad (2)$$

(adjusted $R^2 = 0.645$), which are empirical equations for geographical (equation (1)) and local (equation (2)) meteorological controls on $\delta^{18}O$ derived from an analysis of Global Network of Isotopes in Precipitation (GNIP) and instrumental meteorological data[34]. Here P is total monthly precipitation, Eva is average monthly evaporation and Rad is average monthly radiation. Rainfall events 3 days on either side of a tropical cyclone event within 400 km of our sites were excluded from our calculations. It is important to note that Liu's geographical model does not take into account factors such as the source region, transport and condensation history of the air masses. Precipitation at Chillagoe is derived from sources in the Coral Sea and the Gulf of Carpentaria. These may have originally been part of a larger air mass, which has travelled north from cooler waters or south from warmer waters. In contrast, precipitation at Cape Range is largely derived from oceanic air masses from the Indian Ocean. However, at this stage there are no other longitudinal, in-depth analyses of $\delta^{18}O$ in precipitation from the east or west coast of Australia excepting the model used here from ref. 34. Using this relationship and local historical climate data from the two sites, we calculated the average seasonal $\delta^{18}O_M$ VSMOW from 1990 to 2010. $\delta^{18}O_M$ was then normalized and used to de-trend the modern $\delta^{18}O$ to remove the modern monsoonal trend. $\delta^{18}O_M$ beyond the instrumental record was de-trended from the $\delta^{18}O$ data using a spline-interpolated, normalized data set generated from an established Australian–Indonesian monsoonal proxy record[22]. This record has a resolution of ~10 years and extends from 7 years BP to 12,000 years BP, the region experiences a relatively low tropical cyclone frequency (on average, 0.24 tropical cyclones per year pass within 400 km of the site[23]), and the record is comparable with other established monsoonal records from the region[22].

CAI calculation from the Australian tropical cyclone database. From the period 1990–2010, CAI values for Chillagoe and Cape Range were calculated within 400 km of each site using the data available within the Australian Bureau of Meteorology's tropical cyclone database[23]. Of the tropical cyclones recorded within the database, 32 and 35 passed within 400 km of Chillagoe and Cape Range, respectively. Of the 2,114 observation points within the combined data sets, 225 do not contain wind speed measurements. Given the limited number of environmental pressure measurements available, $V_{\max}$ was estimated using the Atkinson/Holliday wind–pressure relationship[35]:

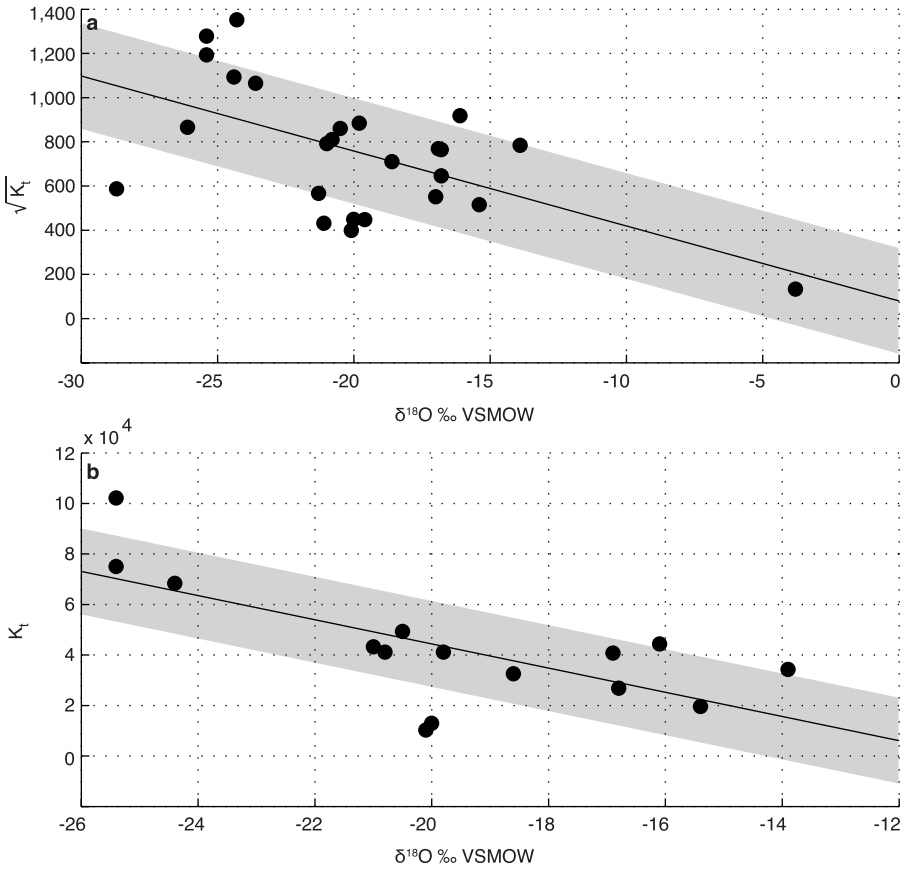
$$V_{\max} = 0.514[6.7(1{,}010 - P_c)^{0.644}] \quad (3)$$

Here P_c is the central pressure in millibars and $V_{\max}$ is the maximum 10-min-mean wind speed in metres per second. There was an average discrepancy of 3 m s^{-1} between the value of $V_{\max}$ estimated using equation (3) and the recorded $V_{\max}$ (Dvorak technique[36]) within the remaining 1,889 observations. Missing $R_{\max}$ estimates from 1,702 observations were calculated using[37]

$$R_{\max} = 46.4 \exp(-0.0155 V_{\max} + 0.0169\varphi)$$

An average discrepancy of 17.5 km was found between the measured and estimated values.

32. NOAA. HURDAT Best Track Data. *NOAA Hurricane Research Division of AOML* http://www.aoml.noaa.gov/hrd/hurdat/tracks1949to2011_epa.html (13 March 2013).
33. Lawrence, J. R., Gedzelman, S. D., Gamache, J. & Black, M. Stable isotope ratios: Hurricane Olivia. *J. Atmos. Chem.* **41,** 67–82 (2002).
34. Liu, J. *et al.* Stable isotopic compositions in Australian precipitation. *J. Geophys. Res.* **115,** D23307 (2010).
35. Atkinson, G. D. & Holliday, C. R. Tropical cyclone minimum sea level pressure/maximum sustained wind relationship for the western North Pacific. *Mon. Weath. Rev.* **105,** 421–427 (1977).
36. Velden, C, *et al.* The Dvorak tropical cyclone intensity estimation technique: a satellite-based method that has endured for over 30 years. *Bull. Am. Meteorol. Soc.* **87,** 1195–1210 (2006).
37. Willoughby, H. E. & Rahn, M. E. Parametric representation of the primary hurricane vortex. Part I: Observations and evaluation of the Holland (1980) model. *Mon. Weath. Rev.* **132,** 3033–3048 (2004).



Extended Data Figure 1 | δ^{18}O VSMOW measured from Hurricane Olivia (1995) versus the calculated K_t values for the corresponding measurement interval. **a**, δ^{18}O versus K_t for all rain types in Hurricane Olivia. $r = -0.58$, $P = 0.02$, $n = 25$. **b**, δ^{18}O versus K_t within the eye wall. Shaded area indicates the r.m.s.e. $r = -0.70$, $P = 0.02$, $n = 15$.

LETTER

doi:10.1038/nature12949

Within-group male relatedness reduces harm to females in *Drosophila*

Pau Carazo[1]*, Cedric K. W. Tan[1]*, Felicity Allen[1], Stuart Wigby[1] & Tommaso Pizzari[1]

To resolve the mechanisms that switch competition to cooperation is key to understanding biological organization[1]. This is particularly relevant for intrasexual competition, which often leads to males harming females[2]. Recent theory proposes that kin selection may modulate female harm by relaxing competition among male relatives[3–5]. Here we experimentally manipulate the relatedness of groups of male *Drosophila melanogaster* competing over females to demonstrate that, as expected, within-group relatedness inhibits male competition and female harm. Females exposed to groups of three brothers unrelated to the female had higher lifetime reproductive success and slower reproductive ageing compared to females exposed to groups of three males unrelated to each other. Triplets of brothers also fought less with each other, courted females less intensively and lived longer than triplets of unrelated males. However, associations among brothers may be vulnerable to invasion by minorities of unrelated males: when two brothers were matched with an unrelated male, the unrelated male sired on average twice as many offspring as either brother. These results demonstrate that relatedness can profoundly affect fitness through its modulation of intrasexual competition, as flies plastically adjust sexual behaviour in a manner consistent with kin-selection theory.

We first tested the effect of relatedness of males within a group on female fitness, by quantifying different aspects of fitness and life history (experiment 1) in females exposed to male triplets. Males were unrelated to the female and either full-sibling brothers of each other (AAA) or unrelated to each other (ABC), and were replaced weekly until female death. Consistent with expectations[3–5], we found that females exposed to AAA males had significantly higher lifetime reproductive success than females exposed to ABC males (Fig. 1a). This was due to the fact that whereas total female lifespan did not differ on average between treatments ($F_{1,\,119} = 1.66$, $P = 0.2$), females exposed to AAA males had significantly longer reproductive lifespan (from eclosion to last egg-laying day[6], Fig. 1b), and female reproductive lifespan was positively correlated with female lifetime reproductive success ($F_{1,\,117} = 484.59$, $P < 0.001$). Two non-mutually exclusive mechanisms might cause this. First, high-fecundity females may die faster when exposed to ABC males, leading to an average higher productivity of AAA replicates ('selective death'). Second, individual females might suffer a steeper rate of age-dependent decline in reproductive output when exposed to ABC rather than AAA males ('reproductive ageing'). We found no evidence of 'selective death': across both treatments (AAA and ABC) females characterized by a relatively low (rather than high) initial oviposition rate died significantly faster than high-fecundity females ($F_{1,\,117} = 11.038$, $P = 0.0012$; treatment–oviposition rate interaction, $F_{1,\,117} = 0.224$, $P = 0.64$), which does not support the prediction that high-fecundity females die faster in ABC compared to AAA trials. In contrast, we found robust support for 'reproductive ageing': the rate of offspring production declined with age significantly faster for females exposed to ABC males than for females exposed to AAA males (Fig. 1c). This was partly due to the fact that offspring egg-to-adult viability declined significantly faster as females aged in the ABC than the AAA treatment (Fig. 1d). We explored the generality of these results by estimating rate-sensitive female fitness costs under different intrinsic rates of population growth[6], and confirmed that exposure to ABC males resulted in relative fitness costs, both for individual females and entire female cohorts, that were particularly pronounced in contracting or stable populations (Extended Data Fig. 1). Experiment 1 therefore indicates that relatedness within male groups promotes female lifetime reproductive success largely by delaying reproductive ageing.

We then investigated the signature of within-group relatedness on male competition. Relatedness can influence the way in which males compete over access to mating opportunities (pre-copulatory competition) and/or the way in which their ejaculates compete over fertilization (post-copulatory competition)[4]. For example, when females mate then disperse to mate again elsewhere, pre-copulatory competition occurs locally and post-copulatory competition occurs globally. We tested the effect of male relatedness within a group on male pre-copulatory competition (experiment 2), by measuring how males respond to changes in within-group male relatedness. We assembled male triplets that consisted of three full-sibling brothers (AAA treatment), two full-sibling brothers and an unrelated male (AAB), or three males unrelated to each other (ABC), and exposed each triplet to a single female unrelated to the males, without replacing males throughout the trial. We detected no difference in mating rates across treatments ($\chi^2_2 = 0.071$, $P = 0.965$; mating rate (number of matings per 100 scans) estimate ± s.e.m.: AAA = 0.70 ± 0.158, AAB = 0.76 ± 0.214, ABC = 0.83 ± 0.260). However, consistent with expectations, fighting was more common in triplets of unrelated males (ABC) than in AAA and AAB triplets (Fig. 2a). ABC males also courted the female more intensely than AAA triplets (Fig. 2b). We confirmed the effect of within-group male relatedness on male behaviour using the first axis of a principal component analysis, summarizing different aspects of male fighting and courting (see online Methods). Within-group relatedness was also associated with variation in male longevity. First, AAA males lived on average longer than ABC males (Fig. 2c). Second, survival analysis by means of a Cox proportional hazards model detected significant overall treatment effects in male mortality risk across treatments (Fig. 2d). Although this experiment was not designed to test treatment effects on female fitness because males were allowed to co-age with females, and we found no significant differences in female lifespan or reproductive success between females exposed to AAA and ABC males, the trends for females exposed to ABC males to suffer shorter reproductive lifespan and lower lifetime reproductive success were in line with the findings of experiment 1 (Extended Data Table 1). We next tested whether within-group relatedness also influences the intensity of male post-copulatory competition. For example, competing with relatives may inhibit male allocation of seminal fluid products such as the *Drosophila* sex peptide, which boosts female egg-laying rates and inhibits female re-mating, hence delaying sperm competition[7,8], but can also contribute to female harm and reproductive ageing under certain conditions[9,10]. We tested this idea (experiment 3) by monitoring mating duration with the first male, latency to re-mate with a new male, and egg-laying rates in females, which were

[1]Edward Grey Institute, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.
*These authors contributed equally to this work.

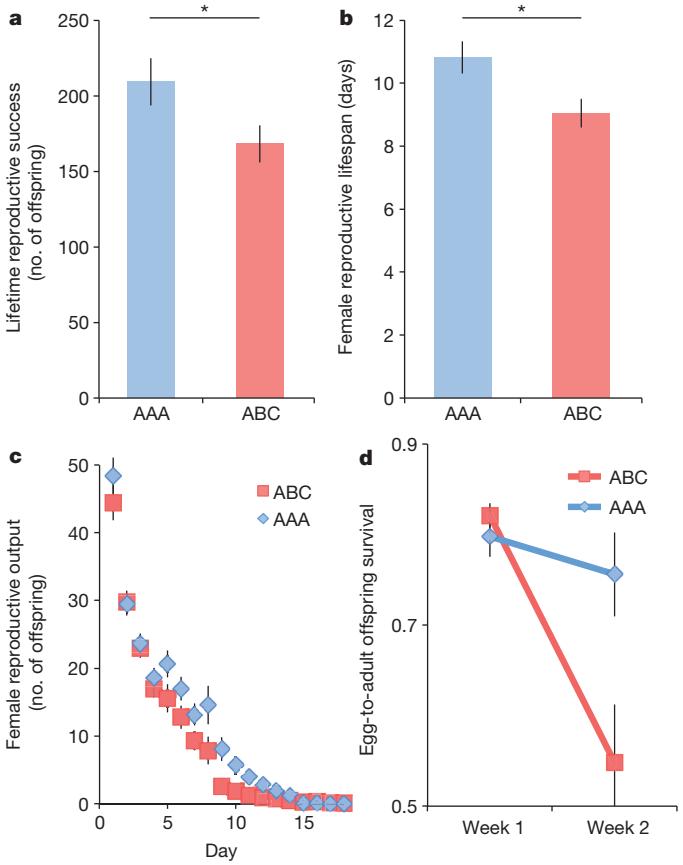


Figure 1 | The effect of male–male relatedness on female fitness. **a**, Female lifetime reproductive success was higher in the high male-relatedness treatment (AAA) than in the low male-relatedness treatment (ABC; $F_{1,\ 119} = 4.11$, $P = 0.045$). This difference was highly significant when we included female reproductive lifespan and its interaction with treatment as factors in the analysis ($F_{1,117} = 20.83$, $P < 0.001$). **b**, Female reproductive lifespan was longer in the high-male relatedness treatment (AAA) than in the low-male relatedness treatment (ABC; $F_{1,119} = 6.55$, $P = 0.012$) and the probability to cease reproducing at any given time was lower ($\chi^2_2 = 3.95$, $P = 0.047$; $n_{AAA} = 63$, $n_{ABC} = 62$). **c**, Female reproductive rates declined more sharply in individual females exposed to ABC rather than to AAA males (average number of offspring produced by AAA and ABC females over successive days of their life: treatment, $\chi^2_1 = 4.11$, $P = 0.043$; day, $\chi^2_1 = 1570.8$, $P < 0.001$; treatment–day interaction, $\chi^2_1 = 7.55$, $P = 0.006$). **d**, Offspring viability (egg-to-adult survival) declined more sharply over time in females exposed to ABC rather than AAA males (treatment–week interaction: $\chi^2_1 = 9.23$, $P = 0.002$, estimated difference in viability drop AAA–ABC, mean ± s.e.m.: estimate = −0.231 ± 0.075). Error bars represent mean ± s.e.m.; *$P < 0.05$; $n_{AAA} = 61$, $n_{ABC} = 60$ unless stated otherwise.

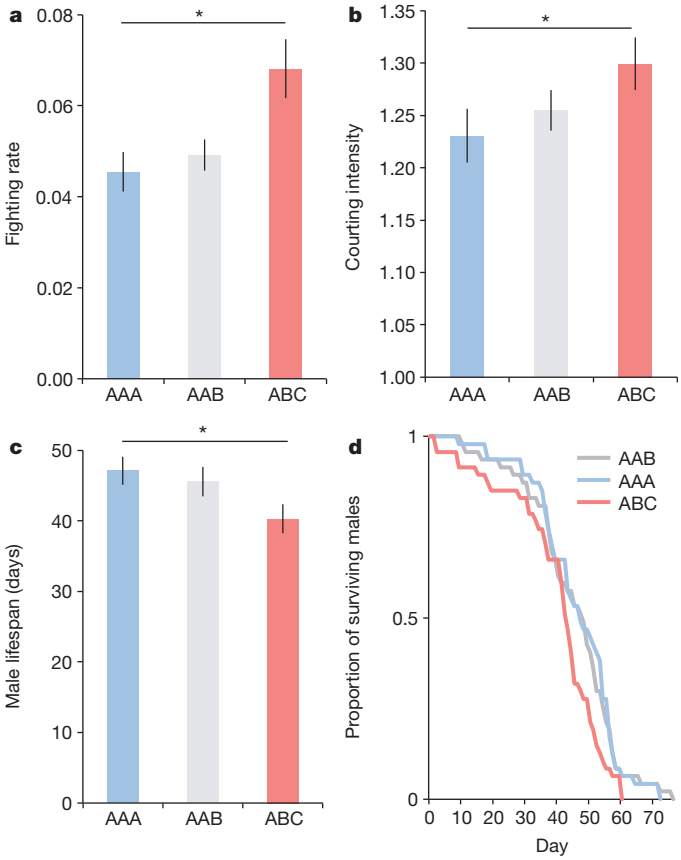


Figure 2 | The effect of male–male relatedness on male sexual behaviour and longevity. **a**, Triplets of unrelated males (ABC) had a significantly higher frequency of male–male fighting than triplets of brothers (AAA) (proportion of focal scans in which male–male fighting was observed, $\chi^2_2 = 14.46$, $P < 0.001$; Tukey, ABC–AAA, $z = 3.73$, $P < 0.001$, ABC–AAB, $z = 2.92$, $P = 0.01$, $n_{AAA} = 47$, $n_{AAB} = 47$, $n_{ABC} = 45$). **b**, Compared to triplets of brothers (AAA), triplets of unrelated males (ABC) were characterized by higher courting intensity (that is, number of courting males when courting was observed, $\chi^2_2 = 5.01$, $P = 0.081$; Tukey ABC–AAA: $z = 2.38$, $P = 0.045$; $n_{AAA} = 47$, $n_{AAB} = 47$, $n_{ABC} = 45$). **c**, Male longevity was significantly lower in unrelated triplets (ABC) than among full-sibling brothers (AAA; $F_{2,\ 128} = 3.77$, $P = 0.026$; estimated differential lifespan for ABC, mean ± s.e.m.: −5.62 ± 2.63, $t = -2.139$, $P = 0.034$; $n_{AAA} = 43$, $n_{AAB} = 44$, $n_{ABC} = 45$). **d**, We found significant differences in male mortality risk across treatments ($\chi^2_2 = 10.47$, $P = 0.005$), and post-hoc direct comparisons between the treatments indicated that this effect was due to males in unrelated triplets (ABC) being more likely to die than in AAA triplets ($\chi^2_2 = 9.55$, $P = 0.002$) and AAB triplets ($\chi^2_2 = 6.66$, $P = 0.010$; $n_{AAA} = n_{AAB} = n_{ABC} = 47$). Error bars represent mean ± s.e.m.; asterisks represent significant post-hoc comparisons. *$P < 0.05$.

first mated to a male from the AAA treatment, a male from the ABC treatment or a control male kept in isolation. We found no difference in the mating duration, re-mating latency or egg-laying rate of the females first mated to AAA versus ABC males (Extended Data Table 2). These results suggest that within-group relatedness is associated with longer male lifespan and relaxes the key aspects of pre- (rather than post-) copulatory competition in this species: courtship and fighting.

To study how groups of relatives interact with unrelated competitors, we assembled (experiment 4) triplets comprising two brothers and one male unrelated to them (that is, AAB), replicated across three different genetic stocks (wild-type, and two homozygous recessive mutants—*sepia* (*se*)[11] and *sparkling poliert* (*spa*, an allele of the *shaven* (*sv*) gene)[12]—each backcrossed into the wild-type Dahomey population[9,13,14]) and exposed to a single female double homozygous recessive for both *se* and *spa*. This design enabled us to test whether males behaved differentially towards related (A) or unrelated (B) competitors, and to assign offspring paternity to A or B males in each trial. We found no evidence of differential behavioural interactions (Extended Data Table 3). An A male was just as likely to fight with his brother than with the unrelated B male (mean ± s.e.m. proportion of all fights that were direct to the B male = 0.51 ± 0.07; effect of relatedness: $z = 0.20$, $P = 0.84$). Similarly, the unrelated of the three males (B) did not court (0.34 ± 0.03, difference from expected 0.33: $z = 0.20$, $P = 0.84$) or mate with the female more frequently than each of the two brothers (0.38 ± 0.07, difference from expected 0.33: $z = 0.63$, $P = 0.53$). However, the unrelated B male sired on average twice as many offspring as either A male (Fig. 3, Extended Data Tables 3 and 4), suggesting that a minority of unrelated competitors may gain a disproportionate share of reproductive success.

Sexual selection favours males that outcompete each other over access to females or their ova to a point that often harms female fitness[2], with pronounced repercussions for the population as a whole, reducing productivity and even leading to local extinctions[15,16], a process akin to the tragedy of the commons[17]. However, in structured populations, in which

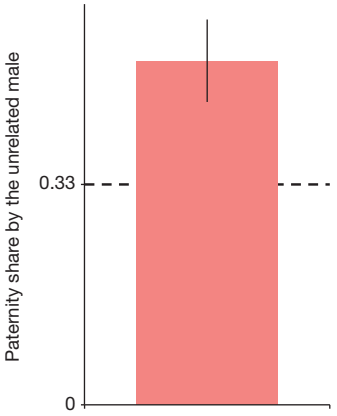


Figure 3 | Unrelated males outcompete brothers. Proportion of offspring sired by the unrelated male (B) in male triplets in which two brothers were matched with an unrelated male (AAB, $n = 54$). The B male sired on average half of the offspring produced by the female, with the two brothers siring the other half between them. This distribution of paternity deviated significantly from an equalitarian distribution of paternity across the three males (that is, 0.33; $z = 3.99$, $P < 0.001$), and was independent of male stock (that is, *se*, *spa*). Error bar represents mean ± s.e.m.

local rivals can be more genetically related to each other than the population average, harming females impacts the inclusive fitness of a male by reducing the reproductive success of his male relatives, and kin selection should discourage female harm by relaxing competition among related males[3–5]. Our study provides experimental support for these expectations in *D. melanogaster*. A proximate explanation is that elevated rates of harassment and male–male fighting, induced by low within-group male relatedness, impose cumulative costs on females and accelerate their reproductive ageing[13]. By mating with genetically different (that is, unrelated) males, females could also incur higher immunological costs[18]. We found little evidence that differential female harm is mediated by male adaptations to post-copulatory sexual selection, suggesting that post-copulatory male competition may occur on a more global scale than pre-copulatory competition[4]. It would therefore appear that in the evolutionary past, the structure of natural *D. melanogaster* populations generated sufficient opportunity for the evolution of kin-selected sexual behaviours. Natural fly populations display limited dispersal and a tendency for local aggregations[19,20], and although the extent to which different laboratory-adapted populations have retained kin-biased sexual behaviour is unclear, evidence of differential sexual responses based on kinship have been shown in some fly laboratory populations, including our own study population[14].

Although insects have inspired a large body of literature documenting how relatedness among group members structures social interactions, this work has largely focused on the particular case of eusociality[1,21,22]. However, the influence of relatedness transcends eusociality and can modulate fundamental aspects of social behaviour more broadly. Sexual cooperation among related males has been observed in different animal societies[23–27], but the fitness consequences for females have previously received little attention. Although the idea that sexual selection results in males harming females is well established[2], we currently lack a framework to understand the high variability in female harm observed across and within taxa[5]. Our study indicates that variation in relatedness and conditional behavioural responses to kin are potentially key factors underpinning such diversity. Although the genetic make-up of social groups was proposed as a modulator of female harm[28,29], it was only recently that kin selection was explicitly applied to sexually selected female harm[3–5]. This process is reminiscent of the way in which kin selection modulates virulence in pathogens[30]. In both female harm and virulence, selfishness leads to a tragedy of the commons, which is inhibited by the relatedness of local competitors[5,30]. As in other cooperative systems[1], we found that minorities of selfish unrelated rivals may be able to invade and persist in groups of male relatives. This may be due to a number of mechanisms, including an imperfect kin recognition system[1]; for example, males might respond to the average relatedness of the group because they are unable to recognize their relatedness to individual group members. Although it is difficult to extrapolate these experimental findings to the complexities of natural populations (for example, variable patterns of relatedness among the offspring of polyandrous females), these results indicate that the benefits of relaxed competition among relatives may be dynamic, diminishing rapidly as populations become less viscous, a result consistent with our finding that the benefits of within-group male relatedness are higher in contracting populations. In conclusion, we present an experimental demonstration that genetic relatedness of social groups modulates the intensity of intrasexual competition and female harm. Future work should investigate the generality of these results and further resolve underpinning proximate mechanisms and evolutionary dynamics.

METHODS SUMMARY

Across experiments, male triplets were set up by collecting recently eclosed (virgin) adult males from controlled 24-h pairings of 1-week-old (virgin) pairs of flies. Families were brought up in the same vials. Triplets consisted of three full-sibling males (AAA), two full-sibling males and one unrelated male (AAB), or three unrelated males (ABC). Male triplets were set up between 48 and 72 h before the beginning of a trial, which began by introducing a 48–72-h-old virgin female (unrelated to any of the males in the triplet) into a vial with a male triplet.

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 September; accepted 11 December 2013.
Published online 22 January 2014.

1. Bourke, A. F. G. *Principles of Social Evolution* (Oxford Univ. Press, Oxford, 2011).
2. Parker, G. A. Sexual conflict over mating and fertilization: an overview. *Phil. Trans. R. Soc. Lond. B* **361,** 235–259 (2006).
3. Rankin, D. J. Kin selection and the evolution of sexual conflict. *J. Evol. Biol.* **24,** 71–81 (2011).
4. Wild, G., Pizzari, T. & West, S. A. Sexual conflict in viscous populations: the effect of the timing of dispersal. *Theor. Popul. Biol.* **80,** 298–316 (2011).
5. Pizzari, T. & Gardner, A. The sociobiology of sex: inclusive fitness consequences of inter-sexual interactions. *Phil. Trans. R. Soc. B* **367,** 2314–2323 (2012).
6. Edward, D. A., Fricke, C., Gerrard, D. T. & Chapman, T. Quantifying the life-history response to male exposure in female *Drosophila melanogaster*. *Evolution* **65,** 564–573 (2011).
7. Liu, H. & Kubli, E. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100,** 9929–9933 (2003).
8. Chapman, T. *et al.* The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. *Proc. Natl Acad. Sci. USA* **100,** 9923–9928 (2003).
9. Wigby, S. & Chapman, T. Sex peptide causes mating costs in female *Drosophila melanogaster*. *Curr. Biol.* **15,** 316–321 (2005).
10. Fricke, C., Bretman, A. & Chapman, T. Female nutritional status determines the magnitude and sign of responses to a male ejaculate signal in *Drosophila melanogaster*. *J. Evol. Biol.* **23,** 157–165 (2010).
11. Anxolabéhère, D. Heterosis overdominance and frequency-dependent selection in *Drosophila melanogaster* at the *sepia* locus. *Evolution* **30,** 523–534 (1976).
12. Fu, W. & Noll, M. The *Pax2* homolog *sparkling* is required for development of cone and pigment cells in the *Drosophila* eye. *Genes Dev.* **11,** 2066–2078 (1997).
13. Partridge, L. & Fowler, K. Non-mating costs of exposure to males in female *Drosophila melanogaster*. *J. Insect Physiol.* **36,** 419–425 (1990).
14. Tan, C. K. W. *et al.* Sex-specific responses to sexual familiarity, and the role of olfaction in *Drosophila*. *Proc. R. Soc. Lond. B* **280,** 20131691 (2013).
15. Le Galliard, J. F. *et al.* Sex ratio bias, male aggression, and population collapse in lizards. *Proc. Natl Acad. Sci. USA* **102,** 18231–18236 (2005).
16. Rankin, D. J. & Kokko, H. Sex, death and tragedy. *Trends Ecol. Evol.* **21,** 225–226 (2006).
17. Rankin, D. J., Dieckmann, U. & Kokko, H. Sexual conflict and the tragedy of the commons. *Am. Nat.* **177,** 780–791 (2011).
18. Fedorka, K. M. & Zuk, M. Sexual conflict and female immune suppression in the cricket, *Allonemobious socius*. *J. Evol. Biol.* **18,** 1515–1522 (2005).
19. McInnis, D. O., Schaffer, H. E. & Mettler, L. E. Field dispersal and population sizes of native *Drosophila* from North Carolina. *Am. Nat.* **119,** 319–330 (1982).
20. Robinson, S. P., Kennington, W. J. & Simmons, L. W. Preference for related mates in the fruit fly, *Drosophila melanogaster*. *Anim. Behav.* **84,** 1169–1176 (2012).
21. Hamilton, W. D. Altruism and related phenomena, mainly in social insects. *Annu. Rev. Ecol. Syst.* **3,** 193–232 (1972).

22. Bourke, A. F. G. & Frank, N. R. *Social Evolution in Ants* (Princeton Univ. Press, 1995).
23. Packer, C. & Pusey, A. E. Cooperation and competition within coalitions of male lions: kin selection or game theory. *Nature* **296,** 740–742 (1982).
24. McDonald, D. B. & Potts, W. K. Cooperative display and relatedness among males in a lek-mating bird. *Science* **266,** 1030–1032 (1994).
25. Petrie, M., Krupa, A. & Burke, T. Peacocks lek with relatives even in the absence of social and environmental cues. *Nature* **401,** 155–157 (1999).
26. Krakauer, A. H. Kin selection and cooperative courtship in wild turkeys. *Nature* **434,** 69–72 (2005).
27. Den Boer, S., Baer, B. & Boomsma, J. J. Seminal fluid mediates ejaculate competition in social insects. *Science* **327,** 1506–1509 (2010).
28. Moore, A. J. & Pizzari, T. Quantitative genetic models of sexual conflict based on interacting phenotypes. *Am. Nat.* **165,** S88–S97 (2005).
29. Eldakar, O. T., Dlugos, M., Pepper, J. W. & Wilson, D. S. Population structure mediates sexual conflict in water striders. *Science* **326,** 816 (2009).
30. Buckling, A. & Brockhurst, M. A. Kin selection and the evolution of virulence. *Heredity* **100,** 484–488 (2008).

Acknowledgements We thank the following funding agencies: Marie Curie fellowship (PIEF-GA-2010-273010 to P.C.), the Wellcome Trust VIP award and NERC fellowship (to S.W.), NERC research grant and the Leverhulme Trust (to T.P.). We thank C. Garroway, J. Perry and S. Michaelides for technical help; and M. Bonsall, A. Buckling, G. McDonald, D. Noble, J. Perry, P. Pizzari, R. Snook and S. West for helpful discussions.

Author Contributions Experiment 1 was designed by P.C., S.W. and T.P., conducted by P.C. and F.A., and analysed by P.C. Experiment 2 was designed by P.C., C.K.W.T., S.W. and T.P., and conducted and analysed by P.C. Experiment 3 was designed and conducted by S.W. and P.C., and analysed by P.C. Experiment 4 was designed by C.K.W.T., T.P. and S.W., and conducted and analysed by C.K.W.T. The article was conceived and written by T.P. with input from P.C., C.K.W.T. and S.W.

Author Information Data have been deposited in the Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.9c7bq. 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 T.P. (tommaso.pizzari@zoo.ox.ac.uk).

METHODS

Across experiments, male triplets were set up by collecting recently eclosed (virgin) adult males from controlled 24-h pairings of 1-week-old (virgin) pairs of flies. Families were brought up in the same vials. Triplets consisted of three full-sibling males (AAA), two full-sibling males and one unrelated male (AAB), or three unrelated males (ABC). Male triplets were set up between 48 and 72 h before the beginning of a trial, which began by introducing a 48–72-h-old virgin female (unrelated to any of the males in the triplet) into a vial with a male triplet.

Experiments 1–3 used a laboratory-adapted, wild-type Dahomey stock of *D. melanogaster*, maintained outbred since 1970 (ref. 31). Experiment 4 used males from three different stocks: wild-type, and two homozygous recessive mutants, sepia (*se*) and *sparkling poliert* (*spa*), each backcrossed into the wild-type Dahomey population for at least five generations. Females for experiment 4 were from the same stocks and were double homozygous recessive for *se* and *spa*. Flies were maintained at 25 °C with overlapping generations to minimize selection on replication rate and life span. Across experiments, families were set up from eggs raised at a standard density (~100 flies per bottle)[31]. Virgins were aged for 1 week before pairing for 24 h to produce experimental flies, which were all aged 48–72 h post eclosion at the beginning of trials. Families developed in the same vials. Triplets consisted of three full-sibling males (AAA), two full-sibling males and one unrelated male (AAB), or three unrelated males (ABC). Male triplets were set up between 48 and 72 h before the beginning of a trial, which began by introducing a 48–72-h-old virgin female (unrelated to any of the males in the triplet) into a vial with a male triplet. Sample sizes were estimated from prior experiments, flies were haphazardly allocated to experimental groups in all experiments, behavioural observations were conducted by an observer who was blind to vial treatments, and animals were only excluded from analyses if they escaped during manipulation (see below) or due to missing data. We checked that data met all necessary assumptions before running tests, including evidence for over- or under-dispersion. The potential influence of extreme outliers ($\alpha = 0.01$–0.05) was explored by substituting extreme outliers for the next non-outlier value[32], however this did not affect the qualitative outcome (direction and significance) of statistical tests. All reported *P* values are two-tailed.

Experiment 1. Experiment 1 was designed to quantify the impact of within-group male relatedness on female fitness. We placed a single virgin female with three virgin males under two different social treatments: all three males were full-siblings (AAA), or all three males were from different families (ABC) ($n_{AAA} = 63$, $n_{ABC} = 62$; 1 ABC vial was excluded because one male in the triplet died before introducing the experimental female). To avoid male co-ageing, we replaced male triplets with fresh young triplets (48–72-h old) every 7 days. For each female, all new triplets were always constructed from the same families used to construct previous triplets. To achieve this, parental pairs were crossed 16 days before introducing each batch of triplets; to minimize ageing, parental flies were isolated in vials containing standard sugar-yeast medium (but no live yeast) and maintained in a chamber at 20 °C. Each parental family contributed males to only one male triplet (that is, 3 males to an AAA triplet or 1 male to an ABC triplet; 252 parental families were used in total). To avoid sampling biases, we only used males from families that produced at least three males following each cross. Experimental foursomes (that is, male triplet plus experimental female) were changed to a fresh vial with live yeast 24 h after triplets were introduced, which enabled us to estimate fecundity and egg-to-adult viability during the first 24 h after having exposed experimental females to a set of novel triplet of males. Apart from that, foursomes were changed to a new fresh vial with live yeast every 3 days, and collected eggs were incubated at standard conditions for 12–15 days after oviposition, at which time we counted emerging offspring. Offspring were collected in 3 batches per week in which the first batch consisted of offspring from day 1, the second of offspring from days 2–4, and the third of offspring from days 5–7. Vials were checked daily for female mortality until female death, at which time males were discarded. Vials in which the date of death of one of the individuals is unknown due to unexpected contingencies (for example, they escaped during a change of vial) were eliminated from linear lifespan models but were included in the demographic survival analysis as 'right-censored individuals' up until the date the individual disappeared[33]. We quantified female lifespan (to the nearest day), the number of offspring each female produced per batch, egg-to-adult viability (only for offspring collected on day one each week; that is, 24 h after the introduction of each new male triplet) and lifetime reproductive success (total number of offspring). We also calculated the fitness index ω at the population (w_{pop}) and individual (w_{ind}) level[6] as rate-sensitive fitness measures (see below). To generate daily offspring counts, offspring emerging from days 2–4 and 5–7 each week were assumed to follow a linear pattern of increase or decrease in number from the known count in day 1 of that week to the known count of day 1 of the next week[6]. We used linear models to test for differences in female lifespan, reproductive lifespan and lifetime reproductive fitness, for which analyses we excluded two AAA and two ABC females (right-censored, see above; final sample size: $n_{AAA} = 61$, $n_{ABC} = 60$). We also ran a Cox proportional hazards survival model (that included right-censored females) to look at differences in mortality risk and in the risk of ceasing to reproduce. To test for 'selective death', we examined whether early fecundity (that is, fecundity during the first 24 h), treatment, and the interaction between the two explained standardized female lifespan or standardized female reproductive lifespan. To examine 'reproductive ageing', we tested for an interaction effect between treatment and time (day) on variation in reproductive rate (that is, offspring produced per day) with a generalized linear mixed model (GLMM) in which we included female reproductive lifespan, treatment, day and treatment–day interaction as fixed factors, and female identity as a random factor. We also tested for a treatment–week interaction in our egg-to-adult viability estimates of week one and week two (most flies had died by week three so we only included these two time points in the analysis). Values of w_{pop} and w_{ind} were calculated from a fitness index developed previously[34]. Values of r were taken in the range of -0.4 to 0.4 as suggested for laboratory populations of *D. melanogaster*[35]. Values of w_{pop} were used to determine the relative costs (C_r) of decreasing within-group male relatedness for different values of r defined as: $C_r = w_{pop\ ABC}/w_{pop\ AAA}$ (ref. 6). To facilitate comparisons with other studies, offspring counts were halved to take into account each female's genetic contribution[36].

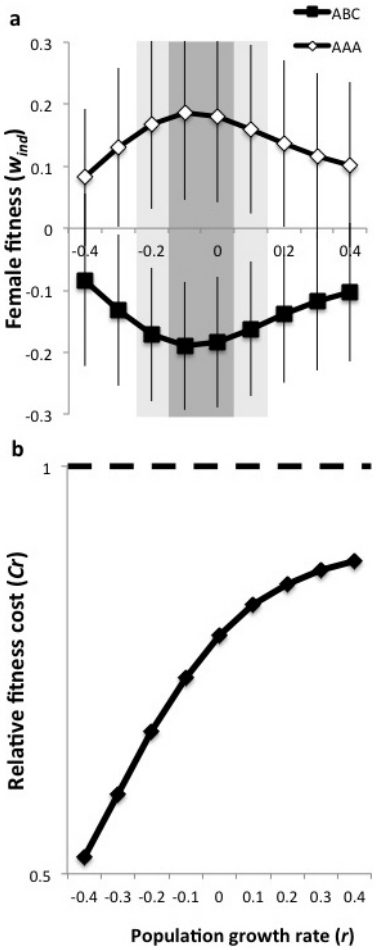
Experiment 2. In experiment 2 we followed the same focal male triplet along with its associated experimental female until the first male in the vial died (see below). For this experiment, we added a third treatment with two full siblings and one unrelated male (AAB). The underlying rationale was to include a treatment with both related and unrelated males as behavioural responses might vary in this treatment (for example, related males may cooperate or be more aggressive against the unrelated male). In this experiment, the design was paired: all the A males belonged to the same family and therefore each family of males were represented three times (AAA, AAB and ABC; one set). We set up 47 sets of male AAA–AAB–ABC triplets by balancing the order in which triplets reflecting different within-group male relatedness treatments were set up. Systematic behavioural observations began 24 h after the start of the experiment, and were conducted every day for the first 5 days and then every second day for the next 5 days (that is, days 2–6, 8 and 10). Observations started after lights on and lasted for a total of 3 h, during which vials were scanned approximately every 10 min by a single observer who was blind to the treatment of each vial. We quantified matings, courtship events directed at the female[37], and the frequency of male–male aggressive events[38], which were operationally defined as either a charging or boxing event as previously described[39,40]. We used these behavioural data to estimate: mating rate (proportion of scans where mating was observed), probability of mating (whether a female mated or not during the 3h observation period), courtship rate (proportion of scans where courtship was observed), courtship intensity (number of courting males when courting was observed) and aggression rate (proportion of scans where male aggression was observed). We excluded two ABC triplets from this analysis because in one triplet one male died before the end of the first observation period, and the other triplet was lost during manipulation. In contrast to experiment 1, experimental vials were not supplemented with live yeast to maximize female survival during the first 10 days of behavioural observations. Flies were transferred to a new fresh vial after the end of behavioural observations every day for the first 2 weeks of the experiment, and every second day thereafter. Vials were kept and checked daily for mortality until the first male in the vial died. In most vials, females died before the first male, in which case we discarded the female and retained the males until one of them died. We tested for treatment differences in male lifespan (that is, first male to die in each vial) by fitting a linear model with treatment and the days males outlived the female as fixed factors. The latter variable was included to control for the fact that males that coexist with females that die soon may experience a more benign environment. We excluded four AAA, three AAB and two ABC males from this analysis because they were lost during manipulations (for example, while moving them to fresh vials). We also fitted a Cox proportional hazards survival model (with 'days outlived' as covariate) to test for differences in mortality risk across treatments, including the males lost during manipulations as 'right-censored' individuals (that is, individuals that are taken into account for demographic analysis until the day they disappear[33]). Differences in reproductive behaviours across treatments were analysed using a time-explicit analysis by fitting five separate GLMMs with treatment, day and treatment–day interaction as fixed factors and female identity as a random factor; we used Gaussian error distributions for all the variables except for 'mated', which was modelled with a binomial error distribution. Given that there were no treatment differences in the variation of behavioural rates with time, we complemented this analysis by pooling behavioural data across days and testing for treatment effects on the averaged values of courtship rate, courtship intensity and fighting rate, and on the total number of matings. We fitted generalized linear models (GLMs) with Gaussian error distributions for courtship rate and fighting rate, with Poisson error distribution for total number of matings (which allowed us to test for over- or under-dispersion of data), and with Gamma error distribution for courtship intensity

(data positively skewed due to positive extreme outliers). These analyses confirmed results from the time-explicit analysis (total number of matings: $F_{2,\ 136} = 0.026$, $P = 0.974$, residual deviance divided by residual degrees of freedom = 1.21; courtship rate, $F_{2,\ 136} = 1.136$, $P = 0.324$; courtship intensity, $F_{2,\ 136} = 5.583$, $P = 0.005$, ABC estimate ± s.e.m. = 0.04 ± 0.02, $t = 2.05$, $P = 0.042$; fighting rate, $F_{2,\ 136} = 6.872$, $P = 0.001$, ABC estimate ± s.e.m. = 0.02 ± 0.006, $t = 3.50$, $P < 0.001$). For courtship intensity, substituting extreme outliers for the next non-outlier value[32] ($\alpha = 0.1$) was effective in transforming positively skewed courtship intensity data to a normal distribution, and a GLM with Gaussian error distribution on this data also showed a significant treatment effect (courtship intensity, $F_{2,\ 136} = 3.056$, $P = 0.05$, ABC estimate ± s.e.m. = 0.06 ± 0.02, $t = 2.45$, $P = 0.015$). Finally, because male fighting rate and courtship intensity were positively correlated across triplets ($F_{1,133} = 25.250$, $P < 0.001$), and because the strength of such correlation was greater in ABC triplets (treatment–courtship intensity interaction term, $F_{2,133} = 4.071$, $P = 0.019$; ABC–courtship-intensity interaction, estimate ± s.e.m. = 0.083 ± 0.038, $t = 2.22$, $P = 0.028$; relationship between fighting rate and courtship intensity simple effects for: AAA, $F_{1,45} = 4.206$, $P = 0.046$, $F_{adj} = 0.065$; AAB, $F_{1,45} = 0.463$, $P = 0.45$, $F_{adj} = -0.012$; ABC, $F_{1,43} = 15.54$, $P < 0.001$, $F_{adj} = 0.248$), we performed a principal component analysis (PCA) on averaged data of male fighting and both measures of male courting (that is, courtship rate and courtship intensity). Given that there were no treatment differences in the variation of behavioural rates with time, we used data averaged across days to look at correlations between behavioural measures, and to run the PCA. The first axis (PC1) explained over 62% of the variance and captured a concordant proportion of variation in courting rate, courting intensity and fighting intensity (loadings = 0.582, 0.598 and 0.550, respectively), so we retained this variable as a combined measure of male–male competition. We confirmed that PC1 significantly varied with within-group male relatedness ($\chi^2_2 = 6.675$, $P = 0.036$), which was driven by higher values of PC1 in ABC than in AAA triplets (Tukey's test, $z = 2.539$, $P = 0.033$).

Experiment 3. To test for potential differences in ejaculate allocation between AAA and ABC males, we conducted an experiment in which we examined how mating with males kept under different relatedness treatments influenced the key ejaculate-mediated female post-mating responses (receptivity and egg-laying rate). We set up 300 male vials ($n = 100$ each) containing: three full-siblings (AAA), three unrelated males (ABC) or a single male (control). All males were isolated as virgins upon emergence and were kept in treatment vials for 72–96 h before the beginning of the experiment (day 1). On day 1, after lights on, we randomly selected one male in each vial and aspirated it into a fresh vial containing a young (3–4-day-old) unrelated virgin female. Pairs were left together to mate and vials in which matings did not occur within 120 min were discarded (discarded $n_{AAA} = 15$, $n_{ABC} = 11$, $n_{control} = 26$). In vials in which mating did occur, we measured mating duration. At the end of matings, we discarded the male and left the female to lay eggs until the following day. On day 2, after lights on, we aspirated females into a fresh vial with a young (6–7-day-old), unrelated virgin male, and monitored them for 8 h or until re-mating was observed. We retained 'old' vials to count the eggs laid by the female and calculated egg-laying rate as total eggs laid/total egg-laying time (that is, time from end of mating on day 1 until transfer into fresh vial on day 2). We discarded from the analysis 6 AAA, 8 ABC and 9 control females that did not lay eggs (final sample size: $n_{AAA} = 79$, $n_{ABC} = 81$, $n_{control} = 65$). We used three separate GLMs to test for: differences in mating duration across treatments (that is, AAA, ABC and control); the effect of within-group male relatedness on female probability to re-mate, with re-mating (that is, re-mated or not) as a binomial response variable and mating duration, treatment and their interaction as fixed effects; and to look at whether within-group male relatedness affected early egg-laying rate (that is, during the first 24 h of experiment), with egg-laying rate as response and treatment, mating duration and their interaction as fixed effects.

Experiment 4. We set up AAB triplets ($n = 54$ each) using males from three different stocks: wild-type, and two homozygous recessive mutants, sepia (*se*)[11] and sparkling (*spa*)[12], each backcrossed into the wild-type for five generations. Females were double homozygous recessive experimental females (*se spa*). Families used in one set were not used for another. Males from different families also possessed different eye colour to facilitate calculation of paternity estimates (see below). We adopted a randomized balanced design: 54 vials of triplets were set up, comprising 18 vials of wild-type males designated as 'A', 18 vials of *se* males designated as 'A', and 18 vials of *spa* males designated as 'A'. Males were marked with red, yellow or green acrylic paint[41] in a randomized balanced design to enable identification and detailed observations of inter- and intrasexual interactions. We quantified the courtship rate, aggression rate and mating rate in 2-min spot-checks. This was done for 3 hours after lights on, on the first 3 days of the experiment. To quantify paternity in treatment AAB, we counted the number of offspring with different eye colour. We analysed the effect of male relatedness on courtship, male–male aggression, mating and paternity share, using binomial GLMs and beta-binomial GLMs whenever we detected evidence of over- or under-dispersion[42] (see Extended Data Table 2). We tested the effect of male relatedness on courtship in three ways. First, we conducted a GLM with beta-binomial error distribution with the proportion of courtship achieved by the B male as the response variable and the genotypes of A and B males as covariates, and tested whether the parameter estimate of proportion of courtship was different from the null expectation of 0.33 with a *z*-test. This analysis showed that there was no effect of genotype on the proportion achieved by the B male (Extended Data Table 3). Second, we then conducted another beta-binomial GLM with three-alternative forced choices (3-AFC)[43] to verify that the proportion of courtship attained by the B male differed significantly from the null expectation of 0.33. Finally, we tested whether the mean of the distribution of the mean courtships for each of the six genotypic combinations differed from the null mean of 0.33 with a one-sample *t*-test. We tested for the effect of male relatedness on male-male aggression in a similar way: one of the two A males was haphazardly chosen as the focal male and the proportion of all aggression counts that he directed towards the B male was tested against the null expectation of 0.5 with a *z*-test using the parameter estimate obtained from a beta-binomial GLM with the genotype of A male and genotype of B male as covariates; a beta-binomial GLM with two-alternative forced choices (2-AFC)[43]; and with a one sample *t*-test comparing the mean of the distribution of mean proportion of aggressive counts across the six genotypic combinations against the null expectation of a mean of 0.5. We tested whether the proportion of mating by the B male differed from 0.33 using a binomial GLM and *z*-test, and a one-sample *t*-test comparing the mean of the distribution of mean proportion of mating across the six genotypic combinations against the null expectation of a mean of 0.33. Finally, we tested whether the share in paternity of the B-males deviated from the null expectation of 0.33 using: a *z*-test comparing the parameter estimate of paternity share obtained from a beta-binomial GLM with the genotype of A male and genotype of B-male as covariates, against the null expectation of 0.33; a beta-binomial GLM with 3-AFC; and a one sample *t*-test comparing the mean of the distribution of mean paternity share across the six genotypic combinations against the null expectation of a mean of 0.33.

31. Clancy, D. J. & Kennington, J. A simple method to achieve consistent larval density in bottle cultures. *Drosoph. Inf. Serv.* **84,** 168–169 (2001).
32. Quinn, G. P. & Keough, M. J. *Experimental Design and Data Analysis for Biologists.* (Cambridge Univ. Press, 2002).
33. Kleinbaum, D. G. & Klein, M. *Survival analysis: A self-learning text.* 3rd edn (Springer, 2012).
34. Charlesworth, B. Selection in populations with overlapping generations. I. The use of Malthusian parameters in population genetics. *Theor. Popul. Biol.* **1,** 352–370 (1970).
35. Prout, T. & McChesney, F. Competition among immatures affects their adult fertility: population-dynamics. *Am. Nat.* **126,** 521–558 (1985).
36. McGraw, J. B. & Caswell, H. Estimation of individual fitness from life-history data. *Am. Nat.* **147,** 47–64 (1996).
37. Bastock, M. & Manning, A. The courtship behaviour of *Drosophila melanogaster. Behaviour* **8,** 85–111 (1955).
38. Chen, S., Lee, A. Y., Bowens, N. M., Huber, R. & Kravitz, E. A. Fighting fruit flies: A model system for the study of aggression. *Proc. Natl Acad. Sci. USA* **99,** 5664–5668 (2002).
39. Jacobs, M. E. Influence of light on mating of *Drosophila melanogaster. Ecology* **41,** 182–188 (1960).
40. Dierick, H. A. A method for quantifying aggression in male *Drosophila melanogaster. Nature Protocols* **2,** 2712–2718 (2007).
41. Nilsen, S. P., Chan, Y.-B., Huber, R. & Kravitz, E. A. Gender-selective patterns of aggressive behaviour in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA* **101,** 12342–12347 (2004).
42. McCullagh, P. & Nelder, J. A. *Generalized Linear Models.* 2nd edn (Chapman & Hall, 1989).
43. Liggett, R. E. & Delwiche, J. F. The beta-binomial model: variability in overdispersion across methods and over time. *J. Sens. Stud.* **20,** 48–61 (2005).



Extended Data Figure 1 | **a**, Rate-sensitive estimates of individual female fitness (w_{ind}) over a gradient in population growth rates (r). Female fitness was estimated to be higher under high within-group male relatedness for values of r ranging from -0.1 to 0 (dark shaded area), a similar non-significant ($0.05 < P < 0.08$) pattern was extended for $r = -0.2$ and $r = 0.1$ (light shaded area). **b**, The effect of within-group male relatedness on population fitness. The relative fitness cost of reducing within-group male relatedness at different population growth rates (r). The dashed line identifies relative fitness of 1, where reduction in within-group male relatedness has no fitness cost. Reducing within-group male relatedness is always costly over the range of population growth rates explored, but particularly so with smaller growth rates.

Extended Data Table 1 | Female rate-insensitive fitness measures in experiment 2

Variable	Treatment	Mean±se	Statistic	*P*-value
Female reproductive lifespan	AAA	21.92±3.51	$F_{2, 114} = 0.657$	0.520
	AAB	21.98±3.47		
	ABC	19.29±3.13		
Female lifespan	AAA	26.72±4.29	$F_{2, 114} = 0.782$	0.460
	AAB	26.58±4.20		
	ABC	23.18±3.76		
Lifetime reproductive success	AAA	175.90±28.17	$F_{2, 114} = 0.179$	0.836
	AAB	179.15±28.33		
	ABC	167.97±27.25		

Due to the co-ageing of males in each experimental vial and to potential Coolidge effects, experiment 2 was not adequate to detect the effect of within-group male relatedness on female fitness, and we found no significant treatment effects in rate-insensitive measures of female fitness. However, fitness measures follow the same trends observed in experiment 1. Furthermore, the analysis of survival curves in experiment 2 suggests a relatively higher initial mortality in ABC compared to AAA vials at day 8, which is when male triplets were replaced by fresh males in experiment 1 (survival, mean ± s.e.m.: AAA = 0.98 ± 0.02; AAB = 0.92 ± 0.04; ABC = 0.87 ± 0.05).

Extended Data Table 2 | Female post-mating responses in experiment 3

Female response	Treatment	Mean ± se
Mating duration (minutes)	AAA	22.23±0.54
	ABC	21.12±0.46
	Control	17.91±0.53
Re-mating propensity (re-mated/total)	AAA	28/79
	ABC	24/81
	Control	34/65
Egg-laying rate (eggs/hour)	AAA	2.17±0.10
	ABC	2.31±0.10
	Control	2.26±0.11

We did not find any evidence of differences in female receptivity or egg-laying rate between females mated to AAA versus ABC males ($n_{AAA} = 79$, $n_{ABC} = 81$, $n_{control} = 65$). We found a significant treatment effect on mating duration ($F_{2,\ 222} = 17.98$, $P < 0.001$) but this was due to both AAA and ABC males mating for longer than control males (Tukey, control–AAA, $t = -5.839$, $P < 0.001$; control–ABC, $t = -3.975$, $P < 0.001$; ABC–AAA, $t = -1.023$, $P = 0.251$). Similarly, we found a significant treatment effect on female re-mating propensity (treatment effect, deviance = 10.448, $P = 0.005$; interaction term, deviance = 1.208, $P = 0.547$), but this was again due to females mated with AAA and ABC males having a significantly lower probability of re-mating than females mated to control males (Tukey, control–AAA, $t = -0.923$, $P = 0.038$; control–ABC, $t = -1.133$, $P = 0.006$; ABC–AAA, $t = -0.210$, $P = 0.813$). Finally, we did not find significant treatment differences in egg-laying rate (treatment effect, deviance = 5.540, $P = 0.063$; interaction term, deviance = 0.476, $P = 0.788$; Tukey, ABC–AAA, $z = 1.532$, $P = 0.275$; control–AAA, $z = 2.296$, $P = 0.056$; control–ABC, $z = 0.976$, $P = 0.591$).

Extended Data Table 3 | Summary of statistical tests in experiment 4

Variable	Test	Estimate ± se	Statistic	*P*-value
Paternity share by B-male (test against 0.33)*	β-binomial, genotypes as covariates	0.579 ± 0.067	$z = 3.66$	< 0.001
	β-binomial, with 3-AFC	0.563 ± 0.058	$z = 3.99$	< 0.001
	t-test	0.515 ± 0.063	$t_5 = 2.89$	0.034
Proportion courtship by B-male (test against 0.33)*	β-binomial, genotypes as covariates	0.363 ± 0.065	$z = 0.46$	0.645
	β-binomial, with 3-AFC	0.339 ± 0.029	$z = 0.20$	0.842
	t-test	0.351 ± 0.049	$t_5 = 0.35$	0.726
Proportion mating by B-male (test against 0.33)	Binomial	0.379 ± 0.072	$z = 0.63$	0.529
	t-test	0.277 ± 0.041	$t_5 = 1.38$	0.174
Proportion of aggressions toward B-male (test against 0.5)†	β-binomial, genotypes as covariates	0.460 ± 0.068	$z = 0.59$	0.555
	β-binomial, with 2-AFC	0.513 ± 0.068	$z = 0.20$	0.842
	t-test	0.513 ± 0.046	$t_5 = 0.27$	0.787

** original binomial model over-dispersed, † original binomial model under-dispersed*

Paternity share by the B male was significantly different from 0.33. The proportion of courtship and mating by B males did not differ from 0.33 and the proportion of all aggressive events performed by one haphazardly-selected of the two A males towards the B male did not differ from 0.5.

Extended Data Table 4 | Effect of genotype of A male and genotype of B male on the response variable

Variable	Covariate	χ^2	df	*P*-value
Paternity share by B-male	Genotype A	0.38	2	0.829
	Genotype B	4.76	2	0.093
Proportion courtship by B male	Genotype A	3.53	2	0.171
	Genotype B	3.20	2	0.202
Proportion mating by B male	Genotype A	0.21	2	0.898
	Genotype B	1.01	2	0.601
Proportion aggression by B male	Genotype A	3.19	2	0.179
	Genotype B	1.47	2	0.414

There was no effect of the genotype of either A or B males on any of the paternity or behavioural responses measured.

LETTER

doi:10.1038/nature12969

Bidirectional developmental potential in reprogrammed cells with acquired pluripotency

Haruko Obokata[1,2,3], Yoshiki Sasai[4], Hitoshi Niwa[5], Mitsutaka Kadota[6], Munazah Andrabi[6], Nozomu Takata[4], Mikiko Tokoro[2], Yukari Terashita[1,2], Shigenobu Yonemura[7], Charles A. Vacanti[3] & Teruhiko Wakayama[2,8]

We recently discovered an unexpected phenomenon of somatic cell reprogramming into pluripotent cells by exposure to sublethal stimuli, which we call stimulus-triggered acquisition of pluripotency (STAP)[1]. This reprogramming does not require nuclear transfer[2,3] or genetic manipulation[4]. Here we report that reprogrammed STAP cells, unlike embryonic stem (ES) cells, can contribute to both embryonic and placental tissues, as seen in a blastocyst injection assay. Mouse STAP cells lose the ability to contribute to the placenta as well as trophoblast marker expression on converting into ES-like stem cells by treatment with adrenocorticotropic hormone (ACTH) and leukaemia inhibitory factor (LIF). In contrast, when cultured with Fgf4, STAP cells give rise to proliferative stem cells with enhanced trophoblastic characteristics. Notably, unlike conventional trophoblast stem cells, the Fgf4-induced stem cells from STAP cells contribute to both embryonic and placental tissues *in vivo* and transform into ES-like cells when cultured with LIF-containing medium. Taken together, the developmental potential of STAP cells, shown by chimaera formation and *in vitro* cell conversion, indicates that they represent a unique state of pluripotency.

We recently discovered an intriguing phenomenon of cellular fate conversion: somatic cells regain pluripotency after experiencing sublethal stimuli such as a low-pH exposure[1]. When splenic CD45$^+$ lymphocytes are exposed to pH 5.7 for 30 min and subsequently cultured in the presence of LIF, a substantial portion of surviving cells start to express the pluripotent cell marker Oct4 (also called Pou5f1) at day 2. By day 7, pluripotent cell clusters form with a bona fide pluripotency marker profile and acquire the competence for three-germ-layer differentiation as shown by teratoma formation. These STAP cells can also efficiently contribute to chimaeric mice and undergo germline transmission using a blastocyst injection assay[1]. Although these characteristics resemble those of ES cells, STAP cells seem to differ from ES cells in their limited capacity for self-renewal (typically, for only a few

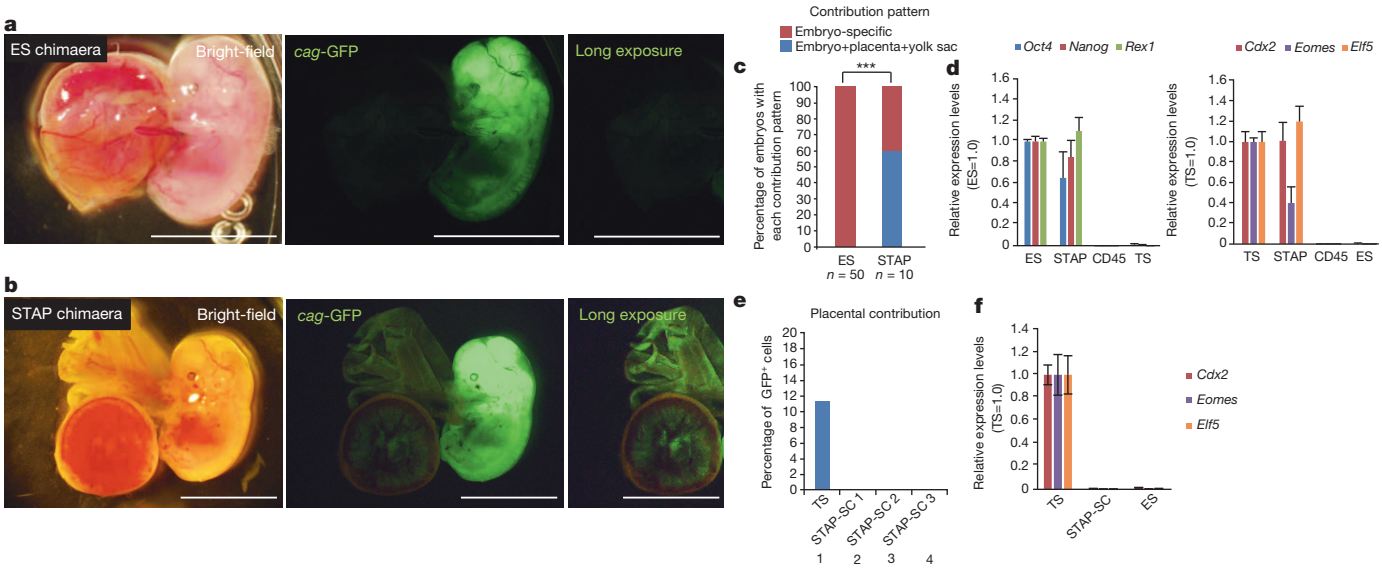


Figure 1 | STAP cells contribute to both embryonic and placental tissues *in vivo*. **a**, **b**, E12.5 embryos from blastocysts injected with ES cells (**a**) and STAP cells (**b**). Both cells are genetically labelled with GFP driven by a constitutive promoter. Progeny of STAP cells also contributed to placental tissues and fetal membranes (**b**), whereas ES-cell-derived cells were not found in these tissues (**a**). Scale bar, 5.0 mm. **c**, Percentages of fetuses in which injected cells contributed only to the embryonic portion (red) or also to placental and yolk sac tissues (blue). ***$P < 0.001$ with Fisher's exact test. **d**, qPCR analysis of FACS-sorted *Oct4*-GFP-strong STAP cells for pluripotent marker genes (left) and trophoblast marker genes (right). Values are shown as ratio to the expression level in ES cells. Error bars represent s.d. **e**, Contribution to placental tissues. Unlike parental STAP cells and trophoblast stem (TS) cells, STAP stem cells (STAP-SCs) did not retain the ability for placental contributions. Three independent lines were tested and all showed substantial contributions to the embryonic portions. **f**, qPCR analysis of trophoblast marker gene expression in STAP stem cells. Error bars represent s.d.

[1]Laboratory for Cellular Reprogramming, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. [2]Laboratory for Genomic Reprogramming, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. [3]Laboratory for Tissue Engineering and Regenerative Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Laboratory for Organogenesis and Neurogenesis, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. [5]Laboratory for Pluripotent Stem Cell Studies, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. [6]Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. [7]Electron Microscopy Laboratory, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. [8]Faculty of Life and Environmental Sciences, University of Yamanashi, Yamanashi 400-8510, Japan.

passages) and in their vulnerability to dissociation[1]. However, when cultured in the presence of ACTH and LIF for 7 days, STAP cells, at a moderate frequency, further convert into pluripotent 'stem' cells that robustly proliferate (STAP stem cells).

Here we have investigated the unique nature of STAP cells, focusing on their differentiation potential into the two major categories (embryonic and placental lineages) of cells in the blastocyst[5–8]. We became particularly interested in this question after a blastocyst injection assay revealed an unexpected finding. In general, progeny of injected ES cells are found in the embryonic portion of the chimaera, but rarely in the placental portion[5,7] (Fig. 1a; shown with *Rosa26*-GFP). Surprisingly, injected STAP cells contributed not only to the embryo but also to the placenta and fetal membranes (Fig. 1b and Extended Data Fig. 1a–c) in 60% of the chimaeric embryos (Fig. 1c).

In quantitative polymerase chain reaction (qPCR) analysis, STAP cells (sorted for strong *Oct4*-GFP signals) expressed not only pluripotency marker genes but also trophoblast marker genes such as *Cdx2* (Fig. 1d and Supplementary Table 1 for primers), unlike ES cells. Therefore, the blastocyst injection result is not easily explained by the idea that STAP cells are composed of a simple mixture of pluripotent cells ($Oct4^{+}Cdx2^{-}$) and trophoblast-stem-like cells ($Oct4^{-}Cdx2^{+}$).

In contrast to STAP cells, STAP stem cells did not show the ability to contribute to placental tissues (Fig. 1e, lanes 2–4), indicating that the derivation of STAP stem cells from STAP cells involves the loss of competence to differentiate into placental lineages. Consistent with this idea, STAP stem cells show little expression of trophoblast marker genes (Fig. 1f).

We next examined whether an alteration in culture conditions could induce *in vitro* conversion of STAP cells into cells similar to trophoblast stem cells[8,9], which can be derived from blastocysts during prolonged adhesion culture in the presence of Fgf4. When we cultured STAP cell clusters under similar conditions (Fig. 2a; one cluster per well in a 96-well plate), flat cell colonies grew out by days 7–10 (Fig. 2b, left; typically in ~30% of wells). The Fgf4-induced cells strongly expressed the trophoblast marker proteins[9–12] integrin α7 (Itga7) and eomesodermin (Eomes) (Fig. 2c, d) and marker genes (for example, *Cdx2*; Fig. 2e).

These Fgf4-induced cells with trophoblast marker expression could be expanded efficiently in the presence of Fgf4 by passaging for more than 30 passages with trypsin digestion every third day. Hereafter, these proliferative cells induced from STAP cells by Fgf4 treatment are referred to as Fgf4-induced stem cells. This type of derivation into trophoblast-stem-like cells is not common with ES cells (unless genetically manipulated)[13] or STAP stem cells.

In the blastocyst injection assay, unlike STAP stem cells, the placental contribution of Fgf4-induced stem cells (*cag*-GFP-labelled) was observed with 53% of embryos (Fig. 2f, g; $n = 60$). In the chimaeric

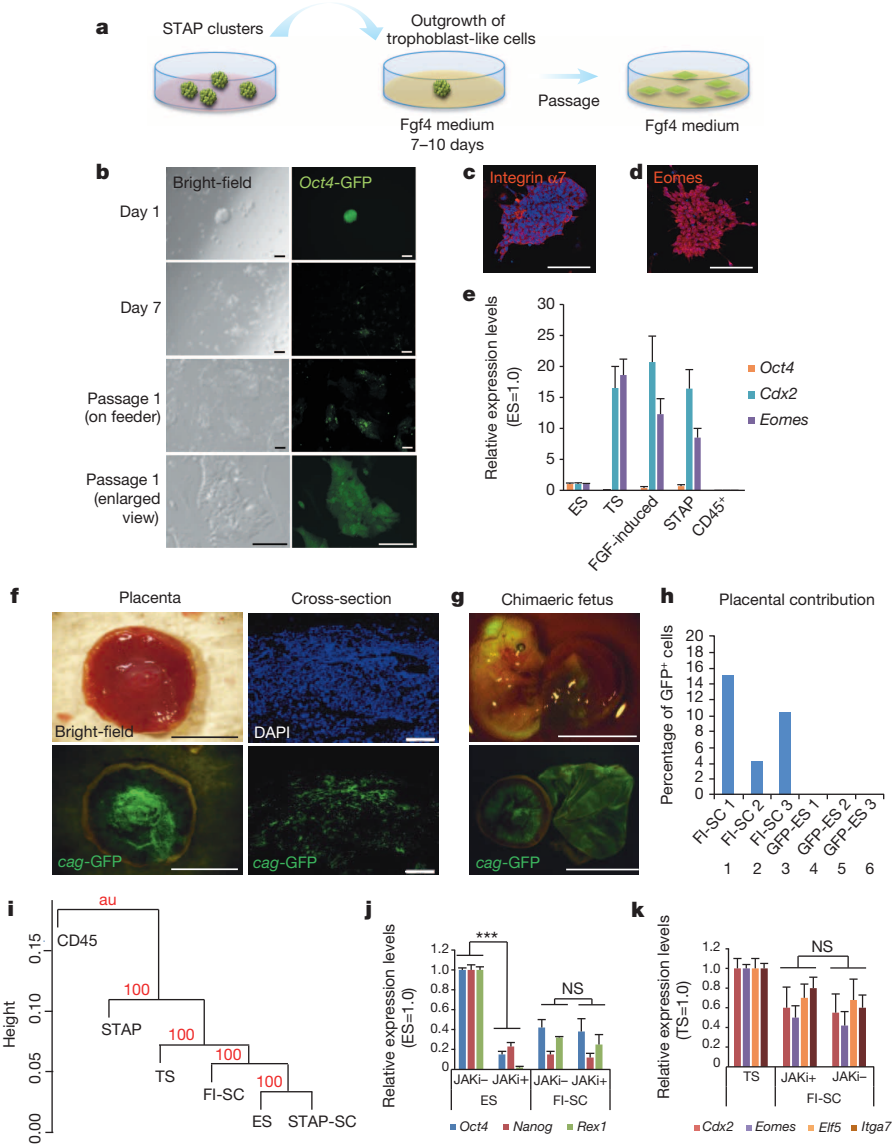


Figure 2 | Fgf4 treatment induces some trophoblast-lineage character in STAP cells. **a**, Schematic of Fgf4 treatment to induce Fgf4-induced stem cells from STAP cells. **b**, Fgf4 treatment promoted the generation of flat cell clusters that expressed *Oct4*-GFP at moderate levels (right). Top and middle: days 1 and 7 of culture with Fgf4, respectively. Bottom: culture after the first passage. Scale bar, 50 μm. **c**, **d**, Immunostaining of Fgf4-induced cells with the trophoblast stem cell markers integrin α7 (**c**) and eomesodermin (**d**). Scale bar, 50 μm. **e**, qPCR analysis of marker expression. **f**, **g**, Placental contribution of Fgf4-induced stem cells (FI-SCs) (genetically labelled with constitutive GFP expression). Scale bars: 5.0 mm (**f** (left panel) and **g**); 50 μm (**f**, right panel). In addition to placental contribution, Fgf4-induced stem cells contributed to the embryonic portion at a moderate level (**g**). **h**, Quantification of placental contribution by FACS analysis. Unlike Fgf4-induced cells, ES cells did not contribute to placental tissues at a detectable level. **i**, Cluster tree diagram from hierarchical clustering of global expression profiles. Red, approximately unbiased *P* values. **j**, qPCR analysis of Fgf4-induced cells (cultured under feeder-free conditions) with or without JAK inhibitor (JAKi) treatment for pluripotent marker genes. **k**, qPCR analysis of FI-SCs with or without JAK inhibitor (JAKi) treatment for trophoblast marker genes. Values are shown as ratio to the expression level in ES cells (**j**) or trophoblast stem cells (**k**). $***P < 0.001$; NS, not significant; *t*-test for each gene between groups with and without JAK inhibitor treatment. $n = 3$. Statistical significance was all the same with three pluripotency markers. None of the trophoblast marker genes showed statistical significance. Error bars represent s.d.

placentae, Fgf4-induced stem cells typically contributed to ~10% of total placental cells (Fig. 2h and Extended Data Fig. 2a, b).

Despite their similarities, we noted that Fgf4-induced stem cells also possessed some critical differences compared with blastocyst-derived trophoblast stem cells. First, Fgf4-induced stem cells exhibited moderate GFP signals and expressed a moderate level of *Oct4* (Fig. 2b; moderate and low levels of immunostaining signals were also seen for Oct4 and Nanog proteins, respectively; Extended Data Fig. 2c), unlike conventional trophoblast stem cells[9] that have little *Oct4* expression (Fig. 2e). Second, unlike trophoblast stem cells, blastocyst-injected Fgf4-induced stem cells also contributed to embryonic tissues (in all cases that involved chimaeric placentae; $n = 32$), although the extent of contribution was generally modest (Fig. 2g). Third, immunostaining revealed that the level of Cdx2 protein accumulation in the nuclei of Fgf4-induced stem cells was marginal as compared to the cytoplasmic level, although the transcript expression level was substantial (Fig. 2e). This may suggest complex and dynamic post-transcriptional regulations for this key transcription factor in Fgf4-induced stem cells (a similar situation was seen for STAP cells, in which clear nuclear localization was not observed for either Cdx2 or Eomes, despite substantial expression of their transcripts). Fourth, in the absence of Fgf4, Fgf4-induced stem cells gradually died in 7–10 days and did not differentiate into large and multi-nuclear cells, unlike trophoblast stem cells (Extended Data Fig. 2d).

To investigate the relationship among STAP cells, STAP stem cells, Fgf4-induced stem cells, ES cells and trophoblast stem cells, we performed genome-wide RNA-sequencing analysis (Fig. 2i for dendrogram; Extended Data Figs 3 and 4 for expression analyses of representative genes[14,15]; Supplementary Tables 2 and 3 for analysis conditions). Whereas STAP cells formed a cluster with STAP stem cells, Fgf4-induced stem cells, ES cells and trophoblast stem cells and not with the parental $CD45^+$ cells, STAP cells were an outlier to the rest of the cell types in the cluster. In contrast, STAP stem cells were closely clustered with ES cells. Fgf4-induced stem cells formed a cluster with a sub-cluster of ES cells and STAP stem cells, whereas trophoblast stem cells comprised an outlier to this cluster, indicating a close relationship of Fgf4-induced stem cells with these pluripotent cells.

However, as Fgf4-induced stem cells lay between STAP stem cells and trophoblast stem cells in the dendrogram, the possibility of contamination of STAP stem cells in the Fgf4-induced stem-cell population cannot be ruled out. Previous studies have indicated that inner cell mass (ICM)-type pluripotent cells can be removed from culture by treating the culture with a JAK inhibitor[16] (Extended Data Fig. 5a, b). In contrast, the JAK inhibitor treatment had no substantial effect on *Oct4*-GFP expression in Fgf4-induced stem-cell culture (Extended Data Fig. 5c, d; see Extended Data Fig. 5e, f for control). Expression of neither pluripotency markers (Fig. 2j) nor trophoblast markers (Fig. 2k) was substantially affected, indicating that pluripotency marker expression is unlikely to reflect contaminating STAP stem cells (ICM-type). Consistent with this idea, Fgf4-induced stem cells that were strongly positive for the trophoblast marker Itga7 (a surface marker for trophoblasts but not ES cells) also expressed high levels of *Oct4*-GFP (Extended Data Fig. 5g).

Notably, when cultured in LIF+FBS-containing medium for 4 days, Fgf4-induced stem cells underwent substantial changes in morphology and started to form ES-cell-like compact colonies with strong GFP signals (Fig. 3a). These cells showed expression of pluripotency makers, but not trophoblast markers (Fig. 3b and Extended Data Fig. 6a), and formed teratomas in mice (Extended Data Fig. 6b). These ES-like cells were generated from Fgf4-induced stem cells sorted for strong expression of the trophoblast marker Itga7, but rarely from Itga7-dim cells (Fig. 3c, d).

To confirm further that Fgf4-induced stem cells with a trophoblast-like nature were converted into ES-like cells, rather than just selecting ES-like cells pre-existing in the Fgf4-induced stem cell culture, we examined the effect of the MEK inhibitor PD0325901 on the ES-like cell generation from Fgf4-induced stem cells. Like trophoblast stem cells, Fgf4-induced stem-cell survival is dependent on FGF–MEK signals, and

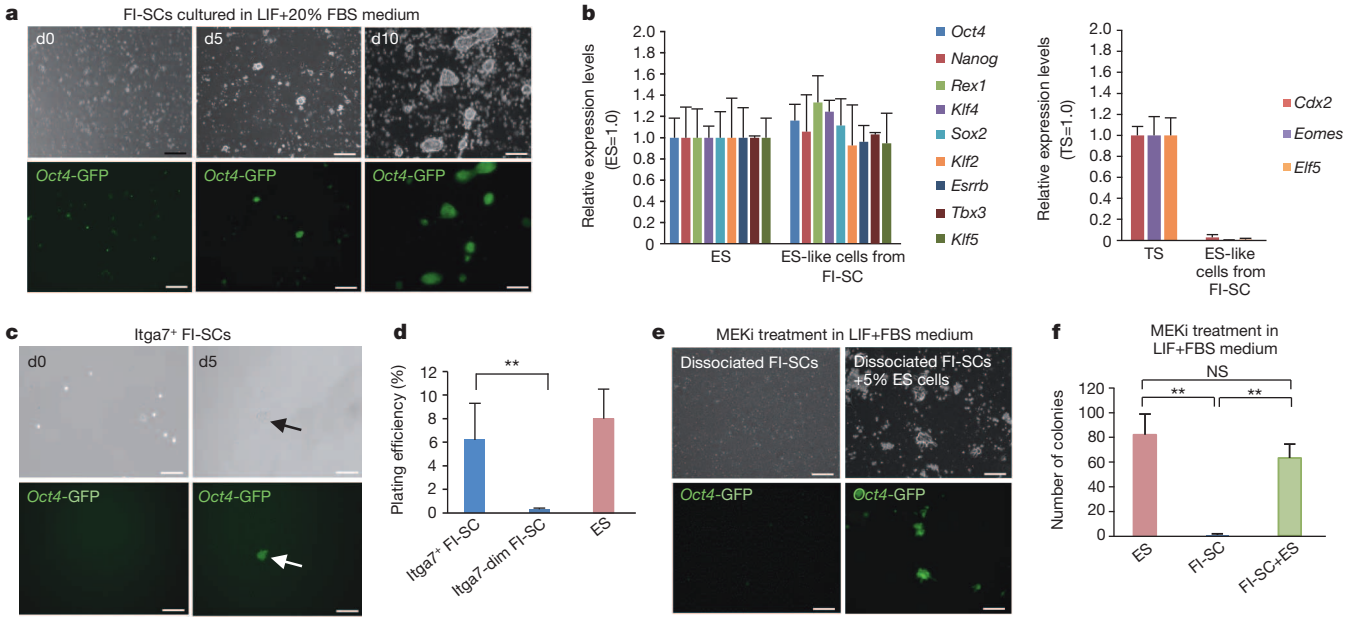


Figure 3 | Fgf4 treatment induces some trophoblast-lineage character in STAP cells. **a**, Culture of *Oct4*-GFP Fgf4-induced cells in LIF + 20% FBS medium. **b**, qPCR analysis of ES-like cells derived from Fgf4-induced cells for pluripotent marker genes (left) and trophoblast marker genes (right). Values are shown as ratio to the expression level in ES cells (left) or trophoblast stem (TS) cells (right). **c**, **d**, Culture of *Oct4*-GFP Fgf4-induced cells sorted by FACS for strong integrin α7 (Itga7) expression in LIF + 20% FBS medium. **d**, Formation frequency (shown by percentage) of *Oct4*-GFP^+ colonies from cells plated on gelatin-coated dishes at a clonal density. $^{**}P < 0.01$; *t*-test; $n = 3$. **e**, **f**, Culture of *Oct4*-GFP Fgf4-induced cells (dissociated) in LIF + 20% FBS medium with MEK inhibitor. $^{**}P < 0.01$; NS, not significant; Tukey's test; $n = 3$. **e**, No substantial formation of *Oct4*-GFP^+ colonies was seen from Fgf4-induced cells in the presence of MEK inhibitor (left), whereas colonies frequently formed when cells were co-plated with *Oct4*-GFP ES cells (right; plated cells were 1/20 of Fgf4-induced cells). **f**, Quantification of colony formation per plated cells (1×10^3 Fgf4-induced cells and/or 1×10^3 ES cells). Unlike Fgf4-induced cells, ES cells formed colonies (regardless of co-plating with FI-SCs) in the presence of MEK inhibitor. Bars and error bars represent mean values and s.d., respectively (**b**, **d**, **f**). Scale bars: 100 μm (**a**, **c**, **e**).

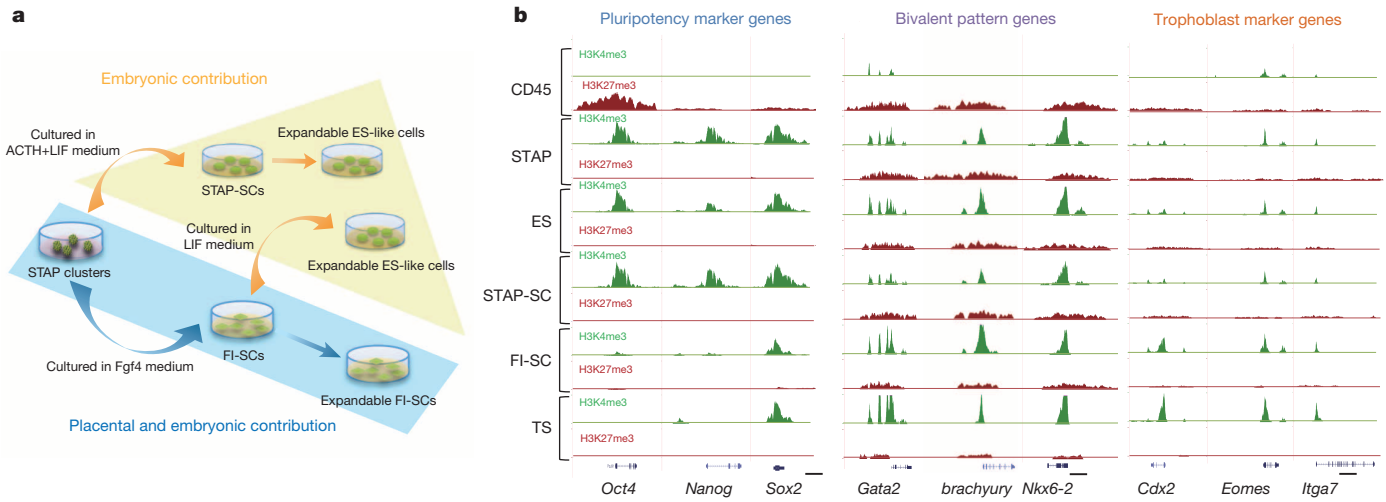


Figure 4 | Differentiation potential and epigenetic state of STAP and STAP-derived stem cells. **a**, Schematic diagram of stem-cell conversion cultures from STAP cells under different conditions. **b**, ChIP-sequencing results of histone H3K4 (green) and H3K27 (red) trimethylation at the loci of pluripotent marker genes (left), bivalent pattern genes (middle) and trophoblast marker genes (right). Scale bars indicate 10 kb for pluripotency marker genes and trophoblast marker genes, and 20 kb for bivalent pattern genes.

the inhibition of MEK activity caused massive cell death (Extended Data Fig. 6c). However, PD0325901 is also known to be a main effector in 2i medium[17] and to promote ES cell maintenance. Addition of PD0325901 to LIF+FBS-containing medium strongly inhibited the formation of ES-like colonies from Fgf4-induced stem cells (Fig. 3e, left, and Fig. 3f). This inhibition was unlikely to be due to secondary toxic effects from massive cell death of Fgf4-induced stem cells, as colonies formed in the presence of PD0325901 when ES cells were co-plated in the same culture with Fgf4-induced stem cells (Fig. 3e, right, and Fig. 3f).

Collectively, these findings demonstrate that STAP-derived Fgf4-induced stem cells not only express both pluripotency markers and trophoblast genes but also have the potential to convert into ES-like cells when cultured in LIF+FBS-containing medium (Fig. 4a).

Here we demonstrate that STAP cells, which have a limited self-renewal ability, can be induced to generate two distinct types of robustly self-renewing stem cells—STAP stem cells and Fgf4-induced stem cells—under different culture conditions. Chromatin immunoprecipitation (ChIP) sequencing analysis showed distinct accumulation patterns of modified histone H3 in the two types of STAP-cell-derived stem cells (Fig. 4b). STAP stem cells (as well as STAP cells) had accumulation patterns of H3K4 and H3K27 trimethylation that resembled those of ES cells at the loci of pluripotency marker genes (*Oct4*, *Nanog*, *Sox2*), bivalent pattern genes[18] (*Gata2*, *brachyury*, *Nkx6-2*) and trophoblast marker genes (*Cdx2*, *Eomes*, *Itga7*). In contrast, the accumulation patterns in Fgf4-induced stem cells at these loci matched more closely those of trophoblast stem cells, except that low levels of accumulation of H3K4 trimethylation in *Oct4* and *Nanog* and of H3K27 trimethylation in the trophoblast marker genes were observed in Fgf4-induced stem cells but not trophoblast stem cells.

Recent studies have also begun to reveal dynamic regulations in multiple cellular states related to pluripotency. These include reports of co-expression of Oct4 and Cdx2 in rat ES cells maintained in the presence of a GSK-3β inhibitor[19,20] and of Oct4 expression in rat extra-embryonic precursors[21]. Another recent study has indicated that conventional ES cell culture also contains a very minor population of Oct4^{-} cells with features resembling those of very early-stage embryos[22], including bidirectional potential. However, these cells are dissimilar to STAP cells as they are Oct4^{-}, unlike STAP cells and Fgf4-induced stem cells. Our preliminary genome-wide RNA-sequencing analysis indicated that both morulae and blastocysts are outliers to the cluster of STAP and ES cells (Extended Data Fig. 6d–f and Supplementary Tables 4 and 5).

A key conclusion drawn from this study is that the reprogramming in STAP conversion goes beyond the pluripotent state of ES cells and involves the acquisition of a wider developmental potential related to both ICM- and trophoectoderm-like states. Because of the inability to clone STAP cells from single cells, we must await future technical advancement to examine whether their dual-directional differentiation potential at the population level may reflect one totipotent state at the single-cell level or two different states of STAP cells coexisting (or fluctuating between them) in culture. As for STAP-cell-derived Fgf4-induced stem cells, which can also contribute to both embryonic and placental tissues, our *in vitro* conversion study combined with inhibitor treatments clearly indicate that the bidirectional potential of Fgf4-induced stem cells is unlikely to reflect the co-presence of separate subpopulations of ES-like and trophoblast-stem-like cells in the culture. Collectively, our study indicates that STAP-based conversion can reprogram somatic cells to acquire not only pluripotency but also the ability of trophoblast differentiation.

METHODS SUMMARY

Cell culture. STAP cells were generated from mouse splenic CD45^{+} cells by a transient exposure to low-pH solution, followed by culture in B27+LIF medium[1]. For establishment of the Fgf4-induced stem-cell line, STAP cell clusters were transferred to Fgf4 (25 ng μl)-containing trophoblast stem-cell medium[9] on MEF feeder cells in 96-well plates. The cells were subjected to the first passage during days 7–10 using a conventional trypsin method. For inducing conversion from Fgf4-induced stem cells into ES-like cells, Fgf4-induced stem cells were trypsinized, and suspended cells were plated in ES maintenance medium containing LIF and 20% FBS. For the establishment of STAP stem-cell lines, STAP spheres were transferred to ACTH-containing medium[15] on a MEF feeder or gelatin-coated dish. Four to seven days later, the cells were subjected to the first passage using a conventional trypsin method, and suspended cells were plated in ES maintain medium containing 5% FBS and 1% KSR.

Chimaeric mice generation and analyses. For injection of STAP stem cells, Fgf4-induced stem cells and ES cells, a conventional blastocyst injection method was used. For STAP cell injection, STAP cell clusters were injected en bloc, because trypsin treatment caused low chimaerism. STAP spherical colonies were cut into small pieces using a microknife under microscopy, then small clusters of STAP colony were injected into day-4.5 blastocysts by large pipette. The next day, the chimaeric blastocysts were transferred into day-2.5 pseudopregnant females.

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 10 March; accepted 20 December 2013.

1. Obokata, H. *et al.* Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* **505,** 641–647 (2014).

2. Gurdon, J. B. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* **10,** 622–640 (1962).
3. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394,** 369–374 (1998).
4. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
5. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90,** 8424–8428 (1993).
6. Niwa, H. How is pluripotency determined and maintained? *Development* **134,** 635–646 (2007).
7. Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105,** 733–737 (1989).
8. Quinn, J., Kunath, T. & Rossant, J. Mouse trophoblast stem cells. *Methods Mol. Med.* **121,** 125–148 (2006).
9. Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282,** 2072–2075 (1998).
10. Tanaka, T. S. *et al.* Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. *Genome Res.* **12,** 1921–1928 (2002).
11. Klaffky, E. *et al.* Trophoblast-specific expression and function of the integrin alpha 7 subunit in the peri-implantation mouse embryo. *Dev. Biol.* **239,** 161–175 (2001).
12. Russ, A. P. *et al.* Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404,** 95–99 (2000).
13. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123,** 917–929 (2005).
14. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
15. van Oosten, A. L., Costa, Y., Smith, A. & Silva, J. C. JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naive pluripotency. *Nature Commun.* **3,** 817 (2012).
16. Yang, J. *et al.* Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell* **7,** 319–328 (2010).
17. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
18. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
19. Meek, S. *et al.* Tuning of β-catenin activity is required to stabilize self-renewal of rat embryonic stem cells. *Stem Cells* **31,** 2104–2115 (2013).
20. Chen, Y., Blair, K. & Smith, A. Robust self-renewal of rat embryonic stem cells requires fine-tuning of glycogen synthase kinase-3 inhibition. *Stem Cell Rep.* **1,** 209–217 (2013).
21. Debeb, B. G. *et al.* Isolation of Oct4-expressing extraembryonic endoderm precursor cell lines. *PLoS ONE* **4,** e7216 (2009).
22. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).

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

Acknowledgements We thank S. Nishikawa and N. Love for discussion and M. Ohgushi, S. Kuraku, M. Eiraku, S. Ohtsuka and K. Kakiguchi for help with experimental planning, material preparation and analyses. Financial support for this research was provided by Intramural RIKEN Research Budget (H.O., T.W. and Y.S.), a Scientific Research in Priority Areas (20062015) to T.W., the Network Project for Realization of Regenerative Medicine to Y.S., and Department of Anesthesiology, Perioperative and Pain Medicine at Brigham and Women's Hospital to C.A.V.

Author Contributions H.O. and Y.S. wrote the manuscript. H.O., Y.S., M.K., M.A., N.T., S.Y. and T.W. performed experiments, and M.T. and Y.T. assisted with H.O.'s experiments. H.O., Y.S., H.N., C.A.V. and T.W. designed the project.

Author Information RNA-seq and ChIP-seq files have been submitted to the NCBI BioSample databases under accessions SAMN02393426, SAMN02393427, SAMN02393428, SAMN02393429, SAMN02393430, SAMN02393431, SAMN02393432, SAMN02393433, SAMN02393434 and SAMN02393435. 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 H.O. (obokata@cdb.riken.jp), T.W. (teru@cdb.riken.jp) or Y.S. (yoshikisasai@cdb.riken.jp).

LETTER

doi:10.1038/nature12864

Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*

Binny M. Mony[1]*, Paula MacGregor[1]*, Alasdair Ivens[1], Federico Rojas[1], Andrew Cowton[1], Julie Young[1], David Horn[2] & Keith Matthews[1]

The protozoan parasites *Trypanosoma brucei* spp. cause important human and livestock diseases in sub-Saharan Africa. In mammalian blood, two developmental forms of the parasite exist: proliferative 'slender' forms and arrested 'stumpy' forms that are responsible for transmission to tsetse flies. The slender to stumpy differentiation is a density-dependent response that resembles quorum sensing in microbial systems and is crucial for the parasite life cycle, ensuring both infection chronicity and disease transmission[1]. This response is triggered by an elusive 'stumpy induction factor' (SIF) whose intracellular signalling pathway is also uncharacterized. Laboratory-adapted (monomorphic) trypanosome strains respond inefficiently to SIF but can generate forms with stumpy characteristics when exposed to cell-permeable cAMP and AMP analogues. Exploiting this, we have used a genome-wide RNA interference library screen to identify the signalling components driving stumpy formation. In separate screens, monomorphic parasites were exposed to 8-(4-chlorophenylthio)-cAMP (pCPT-cAMP) or 8-pCPT-2′-*O*-methyl-5′-AMP to select cells that were unresponsive to these signals and hence remained proliferative. Genome-wide Ion Torrent based RNAi target sequencing identified cohorts of genes implicated in each step of the signalling pathway, from purine metabolism, through signal transducers (kinases, phosphatases) to gene expression regulators. Genes at each step were independently validated in cells naturally capable of stumpy formation, confirming their role in density sensing *in vivo*. The putative RNA-binding protein, RBP7, was required for normal quorum sensing and promoted cell-cycle arrest and transmission competence when overexpressed. This study reveals that quorum sensing signalling in trypanosomes shares similarities to fundamental quiescence pathways in eukaryotic cells, its components providing targets for quorum-sensing interference-based therapeutics.

Protozoan parasites undergo developmental responses to adapt to the different environments encountered within their mammalian host, or during passage through their arthropod vectors[2–4]. As a preparation for transmission, specialized developmental forms are often generated to promote survival when ingested by a biting insect[1,5]. The abundance of these transmission stages can fluctuate during the course of a blood parasitaemia, as can the abundance of the proliferative forms that sustain the infection. The balance of these different cell types determines the within-host dynamics of a parasite, ensuring that the population can maximize its longevity within a host, but also optimize its capacity for spread to new hosts[6–8].

African trypanosomes, *Trypanosoma brucei* spp., are extracellular parasites responsible for human African trypanosomiasis (HAT) and the livestock disease 'nagana'[9]. In the bloodstream, trypanosomes proliferate as morphologically 'slender' forms that evade host immunity by antigenic variation, generating characteristic waves of infection. As each wave of parasitaemia ascends, slender forms stop proliferating and undergo morphological and molecular transformation to stumpy forms, the parasite's transmission stage[10,11]. This differentiation is parasite density dependent[12], resembling quorum-sensing systems common in microbial communities[13]. However, the differentiation-inducing factor SIF ('stumpy induction factor') is unidentified and, although some inhibitors of development have been identified[14–16], the signal-response pathway that promotes stumpy formation is uncharacterized. Moreover, density sensing is reduced in laboratory-adapted 'monomorphic' parasite strains, although they can undergo cell-cycle arrest and the limited expression of some stumpy-specific genes when exposed to cell-permeable analogues of cAMP or AMP[16–18]. This is distinct from cAMP-based signalling because only hydrolysable cAMP drives development, which is metabolized to AMP in the parasite[17,19].

The availability of monomorphic parasite RNAi libraries capable of tetracycline-inducible gene silencing on a genome-wide scale[20] and their ability to respond to hydrolysable-cAMP and AMP analogues allowed us to investigate genes that regulate stumpy formation. Thus, an RNAi library population of 2.5×10^7 cells (maintained with ~fivefold genome coverage) was selected with 100 μM pCPT-cAMP or 10 μM 8-pCPT-2′-*O*-Me-5′-AMP[17] in several replicate flasks, RNAi being induced, or not, with tetracycline (Fig. 1a). Uninduced populations underwent division arrest and eventual death over 5 days (Fig. 1b), whereas three pCPT-cAMP-selected and five 8-pCPT-2′-*O*-Me-5′-AMP-selected populations outgrew in the RNAi-induced populations, these being subject to DNA isolation and RNAi insert amplification (Fig. 1b and Extended Data Fig. 1a). The resulting amplicon profiles varied in intensity but there was remarkable similarity between independently selected populations under each regimen (Extended Data Fig. 1a). To analyse the amplicon complexity in depth, populations from each screen were subjected to Ion Torrent sequencing[21]. Reads were aligned to the *T. brucei* TREU 927/4 reference genome (http://www.genedb.org), identifying 43 genes potentially targeted in either screen (Fig. 1c, Supplementary Data Set 1 and Supplementary Table 1). Twelve genes were common to both screens, five were 8-pCPT-2′-*O*-Me-5′-AMP-specific and 26 were pCPT-cAMP-specific, probably reflecting the observed complexity in each amplicon population (Supplementary Table 1). Analysing the reads for genome alignment and for the presence of the appropriate RNAi library primer flanks refined the list to 27–30 distinct gene targets (Supplementary Table 2, Supplementary Table 3, Extended Data Fig. 1b and Supplementary Data Set 2).

As expected, genes encoding enzymes involved in cAMP/AMP-analogue processing and cellular purine balance were identified, with six selected RNAi targets predicted to alter intracellular AMP levels (Extended Data Fig. 2a). For example, 8-pCPT-2′-*O*-Me-5′-AMP is converted to 8-pCPT-2′-*O*-Me-5′-adenosine in culture medium[17], such that RNAi against adenosine kinase would prevent the conversion of the transported pCPT-adenosine analogue to its AMP equivalent[22] (Extended Data Fig. 2a). Similarly, depletion of adenylosuccinate synthetase (*ADSS*) and adenylosuccinate lyase (*ADSL*) reduce the conversion of inosine monophosphate (IMP) to AMP, potentially counteracting the effect of the membrane-permeable cAMP or AMP analogues. The identified

[1]Centre for Immunity, Infection and Evolution, Institute for Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK. [2]Biological Chemistry & Drug Discovery, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK.
*These authors contributed equally to this work.

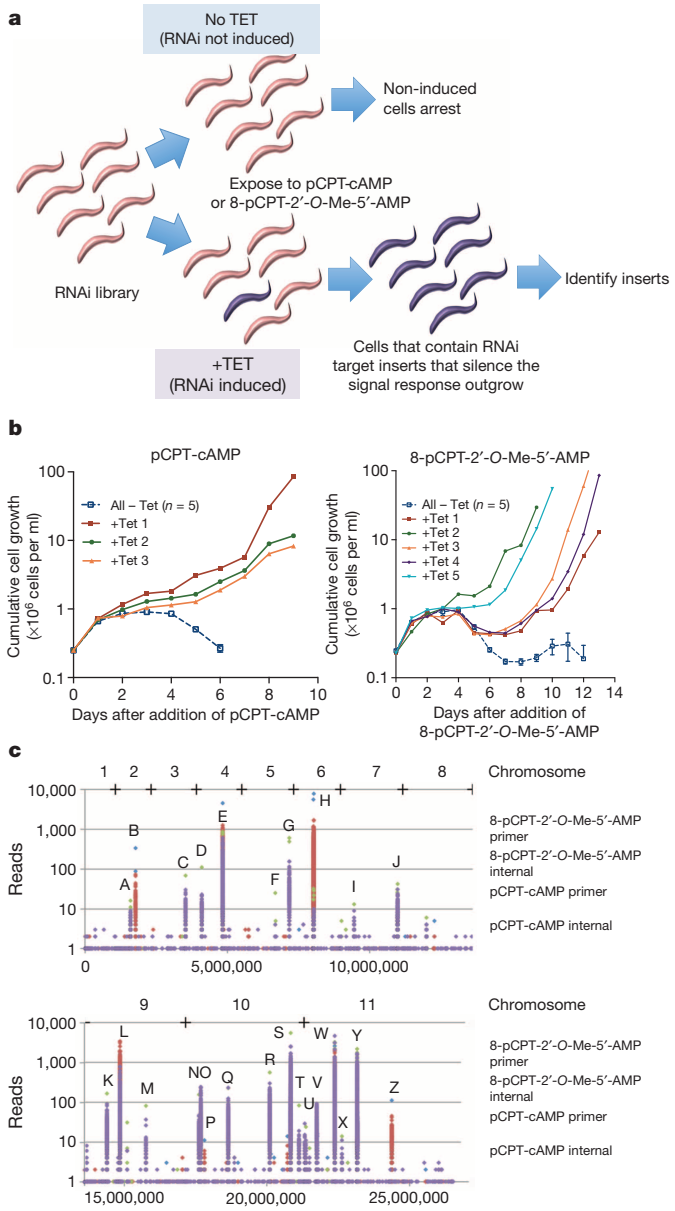


Figure 1 | Identification of trypanosome quorum sensing regulators. **a**, Selection for genes whose RNAi silencing renders trypanosomes resistant to pCPT-cAMP or 8-pCPT-2′-*O*-Me-5′-AMP, identifying molecules that promote stumpy formation. **b**, RNAi libaries were exposed to pCPT-cAMP or 8-pCPT-2′-*O*-Me-5′-AMP, RNAi being induced (1 μg ml^{-1} tetracycline), or not. The curves for uninduced samples are combined for clarity (mean ± s.e.m., $n = 5$). **c**, Ion Torrent read-density from the selected parasites aligned to the trypanosome genome. Because amplicons were fragmented before sequencing, reads with and without the flanking primers are shown.

adenylate kinase, GMP synthase and IMP dehydrogenase targets are also predicted to rebalance purine levels within the parasites.

Signal transduction pathway genes were also unambiguously targeted by RNAi in the selected populations (Supplementary Table 2, Supplementary Table 3 and Supplementary Data Sets 2 and 3). Potentially linking AMP balance to downstream cellular effects, an AMPK/SNF1/KIN11 homologue target (Tb927.3.4560) was identified, as was a MEK kinase (Tb927.2.2720), and predicted cell-cycle regulators of the NEK kinase (Tb927.10.5930/40/50; these three tandem genes being indistinguishable by RNAi phenotyping) and Dyrk/YAK kinase (Tb927.10.15020) families, the latter being required for cellular quiescence (G0) in yeast[23] and *Dictyostelium discoideum*[24]. A dual-specificity phosphatase (Tb927.7.7160) and members of the protein phosphatase 1 gene family (*PP1-4*, *PP1-5*, *PP1-6*[25]; Tb927.4.3620/30/40) were also selected, genes whose knockdown generates only limited cell growth defects in proliferative procyclic forms[25]. Finally, regulatory and effector molecules were represented by the *RBP7* RNA-binding proteins (*RBP7A* and *RBP7B*, Tb927.10.12090/12100, indistinguishable by RNAi), whereas a number of hypothetical proteins with no detectable homologies to any gene were also identified. Overall, this indicated that representatives at many stages in the pCPT-cAMP/8-pCPT-2′-*O*-Me-5′-AMP response pathway had been selected, from signal processing, through signal transduction to regulatory effector molecules.

To validate the identified genes, independent monomorphic and pleomorphic cell lines were initially generated targeting 12 discrete members (Supplementary Table 2). Nine genes were analysed in detail (Extended Data Figs 2b, c), these representing different steps ('signal processing', 'signal transduction', 'effector molecules') in the predicted signal response pathway, and two hypothetical proteins (*HYP1*, Tb927.11.6600; *HYP2*, Tb927.9.4080) of unknown function, although *HYP2* has a DksA zinc finger motif involved in prokaryotic ribosomal

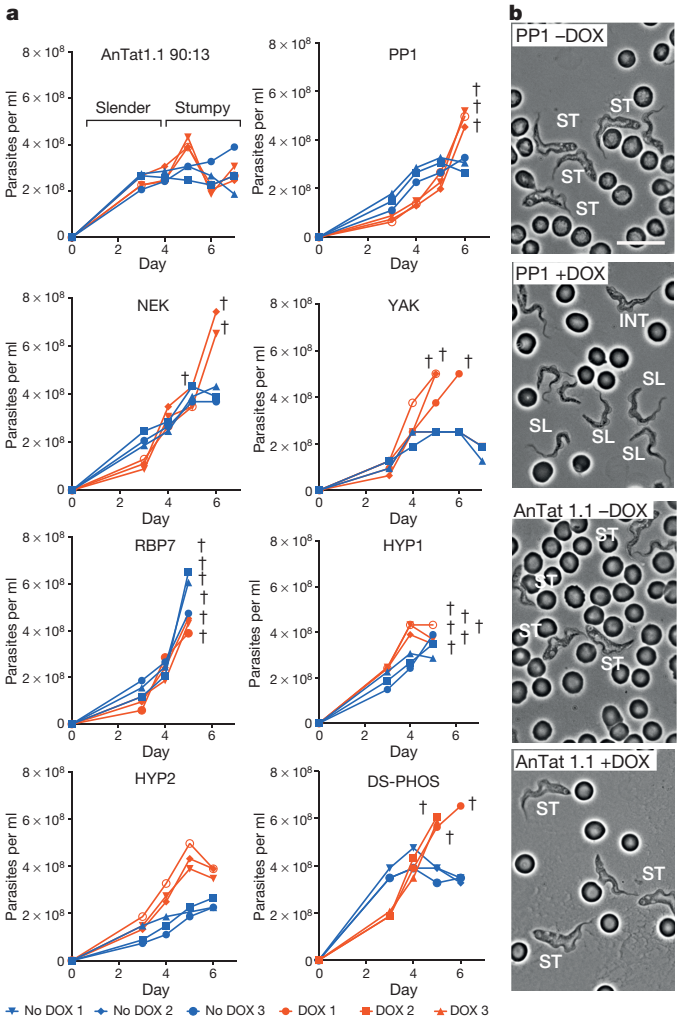


Figure 2 | RNAi to the identified genes prevents growth control and morphological transformation. **a**, *In vivo* growth of pleomorphic RNAi lines targeting distinct genes identified from the genome-wide screen for quorum-sensing-signal resistance. RNAi was induced by provision of doxycycline (DOX) to the infected animals ($n = 3$, red lines), with parallel infections remaining uninduced ($n = 3$, blue lines). Infections were terminated when the ascending parasitaemias were predicted to become lethal within 12 h (daggers). **b**, Morphology of *PP1* RNAi cells and parental *T. brucei* AnTat1.1 90:13 cells (each grown in mice ± doxycycline) at 6 days post infection. The induced *PP1* cells remained predominantly slender in morphology. Cells with slender (SL), intermediate (INT) or stumpy (ST) morphology are labelled. Bar, 15 μm.

RNA transcriptional responses to nutritional status and quorum sensing. Several generated a growth inhibition when targeted by RNAi, indicating roles in other important cellular processes (*PP1*, *HYP1*, *ADSL*, *ADSS*; Supplementary Table 2 and Extended Data Fig. 2b, c), although for *ADSS* this was alleviated in monomorphs by pCPT-cAMP, probably owing to the restoration of the purine balance by the analogue (Extended Data Fig. 2c). Several targeted genes showed evidence of increased resistance to pCPT-cAMP-mediated growth inhibition validating their selection *in vitro* (*PP1*, *NEK*, *YAK*, *RBP7A/B*, *HYP2*, *DS-PHOS*; Extended Data Fig. 2b).

To analyse the physiological relevance of the identified genes in developmental quorum-sensing, nine pleomorphic RNAi cell lines were investigated for their response to the SIF signal *in vivo*. Figure 2 shows that the parental AnTat 1.1 90:13 line generated a highly enriched population of arrested stumpy forms, these accumulating from day 4 onwards (Fig. 2a). For *ADSL* and *ADSS*, parasitaemias were strongly suppressed for at least 5 days, matching their growth characteristics *in vitro* (Extended Data Figs 2c, 3). In contrast, all of the other target RNAi lines exhibited abrogated or delayed stumpy formation over 4–6 days, with mice requiring euthanasia at the exceptionally high parasitaemias generated in the case of *PP1*, *NEK*, *YAK* and *DS-PHOS* (Fig. 2a). The induced parasites also retained a slender morphology or showed delayed progression to stumpy morphology compared to the control or uninduced infections (Fig. 2b and Extended Data Fig. 4a). In two cases (*RBP7*, *HYP1*), the cell lines generated elevated growth when uninduced, reflecting leaky RNAi for these lines (Extended Data Fig. 4b). Such leaky RNAi would be positively selected for genes involved in density-dependent cell-cycle arrest. Confirming the reduction of stumpy formation in each cell line *in vivo*, cell-cycle analysis during the course of the parasitaemias revealed reduced accumulation of cells with a 1 kinetoplast and 1 nucleus (1K1N) configuration, indicating loss or a delay of the cell-cycle arrest in G1/G0 characteristic of stumpy forms (Fig. 3a). Detailed analysis of the molecular characteristics of the populations confirmed that the parasites showed reduced or delayed expression of the stumpy-specific surface protein PAD1[26] (*PP1*-depleted cells are shown in Fig. 3b and Extended Data Fig. 5a) as well as reduced mitochondrial elaboration (Extended Data Fig. 5b). As expected for a slender-enriched population, the *PP1*-depleted cells were less able to differentiate to procyclic forms after exposure to cis-aconitate (CCA), indicated by reduced procyclin expression (Fig. 3c, Extended Data Figs 5c and 6a) and kinetoplast repositioning (Extended Data Fig. 6b). Hence, by the key parameters, RNAi targeting of genes predicted to operate at different steps in a signalling pathway prevented or delayed stumpy formation *in vivo*, confirming their involvement in physiological quorum sensing. With respect to an effector function, the overexpression of *RBP7B* (Tb927.10.12100) (Fig. 4a) promoted premature cell-cycle arrest (Fig. 4b, c and Extended Data Fig. 7) and increased capacity for differentiation to procyclic forms in pleomorphic lines (Fig. 4d), albeit incompletely in the population, similar to the *RBP6*-mediated regulation of development in tsetse forms[27]. Transcriptome analysis revealed few widespread changes in gene expression upon perturbed *RBP7* expression (Extended Data Figs 8 and 9; Supplementary Data Set 4), although RNA regulators and procyclin transcripts were elevated in *RBP7* overexpressing cells whereas histones were downregulated compared to *RBP7*-depleted cells (Extended Data Fig. 9; Supplementary Data Set 4), consistent with their cell-cycle arrest and differentiation competence.

By using a stringent genome-wide *in vitro* selection, these experiments have provided a first identification of the molecules required to promote the development of trypanosome transmission stages in the mammalian bloodstream (Supplementary Table 2 and Fig. 4e). Although inhibitors of stumpy formation[14–16] would not be identified in the screens,

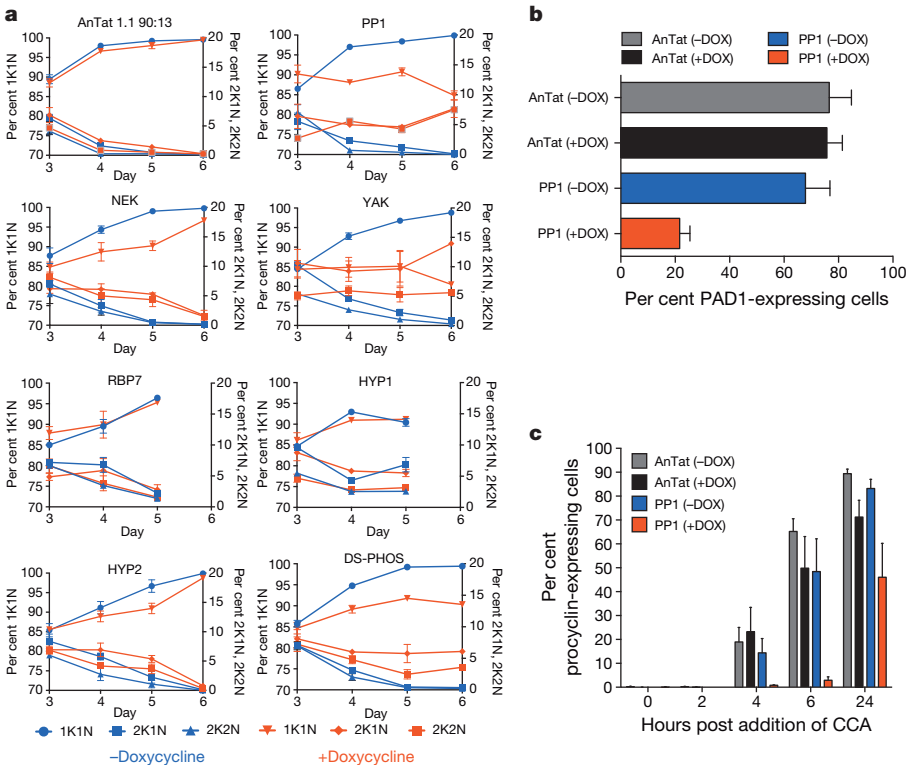


Figure 3 | Silencing the identified genes reduces G1 arrest and differentiation competence. **a**, Cell-cycle status of pleomorphic RNAi lines. $n = 6$; mean ± s.e.m. Percentage 1 kinetoplast (K), 1 nucleus (N) (G0/G1, plus S-phase cells; left *y* axis), 2K1N (G2-phase cells) or 2K2N (post-mitotic cells) (both right *y* axes) are shown. Test genes showed a significant difference (GLMM, $P < 0.001$) in comparison to AnTat1.1 90:13 + doxycycline on at least one day of infection. **b**, PAD1 expression on day 6 post-infection ($n = 3$ per group, mean ± s.e.m.). *PP1* RNAi cells show reduced *PAD1* expression (GLM, $F_{1,4} = 22.35$, $P = 0.009$). **c**, *PP1*-depleted cells show significantly reduced procyclin expression during differentiation (GLM, $F_{1,4} = 10.87$, $P = 0.030$). Bars represents mean ± s.e.m.; $n = 3$.

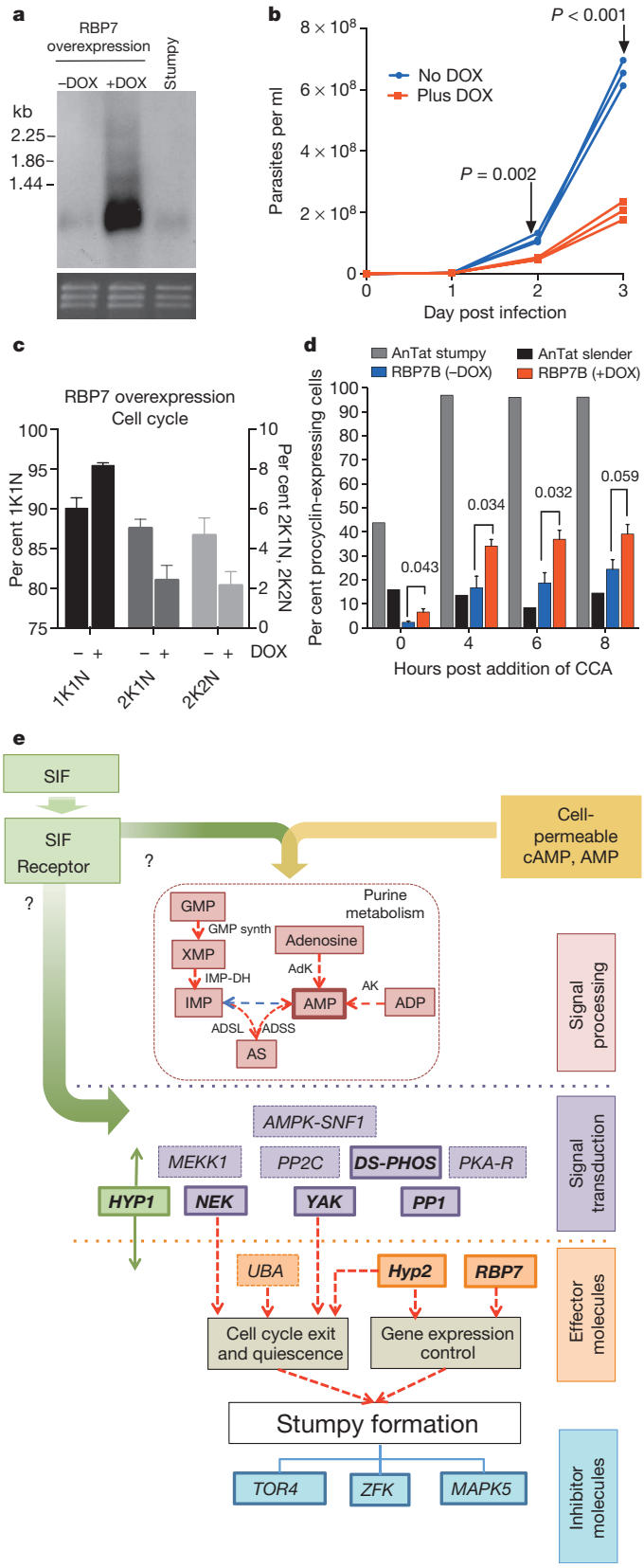


Figure 4 | RBP7 drives cell-cycle arrest and differentiation competence. **a**, Inducible overexpression of *RBP7B* mRNA on day 3 post-infection. Stumpy RNA is also shown; ethidium bromide stained rRNA indicates loading. **b**, AnTat1.1 90:13 induced with doxycycline (red lines) to overexpress *RBP7B* show reduced parasitaemia in mice ($n = 3$ per group). GLM, $F_{1,4} = 55.25$, $P = 0.002$ and $F_{1,4} = 233.1$, $P < 0.001$ on day 2 and 3, respectively. **c**, Cells accumulate in G1 upon *RBP7B* ectopic expression. Values shown are mean ± s.e.m. at day 3 post-infection; $n = 3$ per group. GLM, $F_{1,4} = 15.1$, $P = 0.018$ (1K1N); $F_{1,4} = 8.9$, $P = 0.041$ (2K1N); $F_{1,4} = 5.17$, $P = 0.085$ (2K2N). **d**, Parasites isolated on day 3 post-infection were exposed to 6 mM *cis*-aconitate (CCA) and EP-procyclin expression monitored by flow cytometry. At 0 h enhanced cold induction of EP-procyclin expression is seen in the induced population. $n = 3$; GLM; $F_{1,4} = 8.54$, $P = 0.043$ (0 h); $F_{1,4} = 9.99$, $P = 0.034$ (4 h); $F_{1,4} = 10.36$, $P = 0.032$ (6 h) $F_{1,4} = 6.84$, $P = 0.059$ (8 h). **e**, Schematic of the proposed SIF-signalling pathway in *T. brucei*. Major identified components (Supplementary Table 2) are shown; those in bold italics are experimentally confirmed. The order and potential branching in the pathway is unknown as is the position of pathway inhibitors, *TOR4* (Tb927.1.1930), *ZFK* (Tb927.11.9270) and *MAPK5* (Tb927.6.4220). AdK, adenosine kinase; AK, adenylate kinase; AS, adenylosuccinate; DH, dehydrogenase; Synth; synthase.

the reproducibility of the enriched amplicons in independent selections and the use of Ion Torrent based deep sequencing indicates that many genes involved in the promotion of stumpy formation have been identified. Moreover, this set is significantly enriched for genes whose RNAi reduces differentiation to procyclic forms in monomorphic lines ($P = 0.0008$ χ^2 test; Extended Data Fig. 10), indicating that monomorphs need to progress through a stumpy-like (quiescent) form to differentiate to procyclic forms. Although the use of membrane-permeable analogues prevents identification of a SIF receptor at the parasite surface, our results reveal that selecting resistance to AMP analogues identifies not only purine salvage enzymes, but also genes important in the physiological SIF signalling pathway (Fig. 4e). Interestingly, the *AMPK/SNF1/KIN11* homologue (Tb927.3.4560) is a potential AMP/ATP energy sensor that could inhibit trypanosomal TORC4, whose activity is proposed to prevent stumpy formation[16]. Combined with the confirmation of downstream transduction components such as *PP1*, *NEK*, *YAK* kinase and a dual-specificity phosphatase, our analyses reveal that quorum-sensing-signalled production of stumpy forms in trypanosomes shares components with quiescence regulation in mammalian stem cells[28] and the starvation responses and developmental transitions of unicellular eukaryotes[29].

This assembly of the molecular regulators of quorum sensing provides the first detailed catalogue of an environmental signalling pathway in trypanosomes, providing molecular insight into microbial sociality relevant to both virulence and transmission in a major eukaryotic pathogen. As drivers of the irreversible arrest of stumpy forms in the mammalian bloodstream, these molecules also represent novel therapeutic targets via quorum sensing interference[30], whose pharmacological activation would generate a stringent anti-virulence effect.

METHODS SUMMARY

***In vitro* pCPT-cAMP resistance validation.** Pleomorphic cells were seeded at 1×10^5 cells per ml. RNAi was induced in one flask (1 μg ml^{-1} doxycycline) while the other was left uninduced. After 24 h, each flask was split, one being exposed to 100 μM of pCPT-cAMP (Sigma). Assays were performed in triplicate, growth being monitored every 24 h. Cells were maintained at ≤ 1×10^6 cells ml^{-1} using HMI-9, supplementing with doxycycline and pCPT-cAMP, where needed.

***In vivo* analysis of quorum sensing.** Six female age-matched cyclophosphamide-treated MF1 mice were inoculated intraperitoneally. One group ($n = 3$) was provided with doxycycline (200 μg ml^{-1} in 5% sucrose) in their drinking water immediately pre-inoculation, the other group ($n = 3$) received 5% sucrose only. Parasitaemias were scored over 5–7 days with humane end points conforming to UK Home Office requirements.

Cell-cycle analysis. Methanol-fixed blood smears were re-hydrated in PBS for 10 min and stained with DAPI (100 ng ml^{-1}) for 5 min. 250 cells were analysed per slide.

***In vitro* differentiation to procyclic forms.** Parasites were incubated at 3×10^6 cells ml^{-1} in SDM79 with 6 mM cis-aconitate (CCA), 27 °C. Samples were collected for flow cytometry at 0–24 h.

Flow cytometry. Approximately 3×10^6 cells were fixed in 2% formaldehyde/0.05% glutaraldehyde for >1 h at 4 °C. Subsequently, the cell suspension was pelleted, washed twice with PBS and re-suspended in 200 μl mouse anti-EP-procyclin (Cedar Lane, catalogue no. CLP001A; 1:500) or rabbit anti-PAD1[26] (1:100). After washing and staining with secondary antibody (1:1,000 anti-mouse–FITC and anti-rabbit–Cy5) cells were analysed using a Becton Dickinson LSRII Flow cytometer and analysed using FlowJo software (Tree Star). Unstained cells and secondary antibody-only stained cells provided negative controls.

Statistical analyses. Statistical analysis was carried out with Proc Mixed SAS version 9.3.1 or Minitab version 16, using a general linear model (GLM), general linear mixed model (GLMM), or a χ^2 test.

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 15 May; accepted 11 November 2013.
Published online 15 December 2013.

1. MacGregor, P., Szoor, B., Savill, N. J. & Matthews, K. R. Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. *Nature Rev. Microbiol.* **10,** 431–438 (2012).
2. Goldenberg, S. & Avila, A. R. Aspects of *Trypanosoma cruzi* stage differentiation. *Adv. Parasitol.* **75,** 285–305 (2011).
3. Aly, A. S., Vaughan, A. M. & Kappe, S. H. Malaria parasite development in the mosquito and infection of the mammalian host. *Annu. Rev. Microbiol.* **63,** 195–221 (2009).
4. Matthews, K. R. Controlling and coordinating development in vector-transmitted parasites. *Science* **331,** 1149–1153 (2011).
5. Baker, D. A. Malaria gametocytogenesis. *Mol. Biochem. Parasitol.* **172,** 57–65 (2010).
6. Frank, S. A. Models of parasite virulence. *Q. Rev. Biol.* **71,** 37–78 (1996).
7. Gjini, E., Haydon, D. T., Barry, J. D. & Cobbold, C. A. Critical interplay between parasite differentiation, host immunity, and antigenic variation in trypanosome infections. *Am. Nat.* **176,** 424–439 (2010).
8. Mackinnon, M. J. & Read, A. F. Virulence in malaria: an evolutionary viewpoint. *Phil. Trans. R. Soc. B* **359,** 965–986 (2004).
9. Barrett, M. P. & Croft, S. L. Management of trypanosomiasis and leishmaniasis. *Br. Med. Bull.* **104,** 175–196 (2012).
10. Robertson, M. Notes on the polymorphism of *Trypanosoma gambiense* in the blood and its relation to the exogenous cycle in *Glossina palpalis*. *Proc. R. Soc. Lond. B* **85,** 241–248 (1912).
11. Vickerman, K. Developmental cycles and biology of pathogenic trypanosomes. *Br. Med. Bull.* **41,** 105–114 (1985).
12. Vassella, E., Reuner, B., Yutzy, B. & Boshart, M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J. Cell Sci.* **110,** 2661–2671 (1997).
13. Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21,** 319–346 (2005).
14. Vassella, E. *et al.* Deletion of a novel protein kinase with PX and FYVE-related domains increases the rate of differentiation of *Trypanosoma brucei*. *Mol. Microbiol.* **41,** 33–46 (2001).
15. Domenicali Pfister, D. *et al.* A mitogen-activated protein kinase controls differentiation of bloodstream forms of *Trypanosoma brucei*. *Eukaryot. Cell* **5,** 1126–1135 (2006).
16. Barquilla, A. *et al.* Third target of rapamycin complex negatively regulates development of quiescence in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA* **109,** 14399–14404 (2012).
17. Laxman, S., Riechers, A., Sadilek, M., Schwede, F. & Beavo, J. A. Hydrolysis products of cAMP analogs cause transformation of *Trypanosoma brucei* from slender to stumpy-like forms. *Proc. Natl Acad. Sci. USA* **103,** 19194–19199 (2006).
18. MacGregor, P. & Matthews, K. R. Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in *Trypanosoma brucei*. *Nucleic Acids Res.* **40,** 7705–7717 (2012).
19. Gould, M. K. *et al.* Cyclic AMP effectors in African trypanosomes revealed by genome-scale RNA interference library screening for resistance to the phosphodiesterase inhibitor CpdA. *Antimicrob. Agents Chemother.* **57,** 4882–4893 (2013).
20. Alsford, S. *et al.* High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.* (2011).
21. Rothberg, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475,** 348–352 (2011).
22. Vodnala, M. *et al.* Adenosine kinase mediates high affinity adenosine salvage in *Trypanosoma brucei*. *J. Biol. Chem.* **283,** 5380–5388 (2008).
23. Garrett, S., Menold, M. M. & Broach, J. R. The *Saccharomyces cerevisiae YAK1* gene encodes a protein kinase that is induced by arrest early in the cell cycle. *Mol. Cell. Biol.* **11,** 4045–4052 (1991).
24. Souza, G. M., Lu, S. & Kuspa, A. YakA, a protein kinase required for the transition from growth to development in *Dictyostelium*. *Development* **125,** 2291–2302 (1998).
25. Li, Z., Tu, X. & Wang, C. C. Okadaic acid overcomes the blocked cell cycle caused by depleting Cdc2-related kinases in *Trypanosoma brucei*. *Exp. Cell Res.* **312,** 3504–3516 (2006).
26. Dean, S., Marchetti, R., Kirk, K. & Matthews, K. A surface transporter family conveys the trypanosome differentiation signal. *Nature* **459,** 213–217 (2009).
27. Kolev, N. G., Ramey-Butler, K., Cross, G. A., Ullu, E. & Tschudi, C. Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science* **338,** 1352–1353 (2012).
28. Pietras, E. M., Warr, M. R. & Passegue, E. Cell cycle regulation in hematopoietic stem cells. *J. Cell Biol.* **195,** 709–720 (2011).
29. Zaman, S., Lippman, S. I., Zhao, X. & Broach, J. R. How *Saccharomyces* responds to nutrients. *Annu. Rev. Genet.* **42,** 27–81 (2008).
30. Rasko, D. A. & Sperandio, V. Anti-virulence strategies to combat bacteria-mediated disease. *Nature Rev. Drug Discov.* **9,** 117–128 (2010).

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

Acknowledgements This work was funded by a Wellcome Trust Programme grant (088293MA) to K.M. and by a Wellcome Trust strategic award (095831MA) to the Centre for Immunity, Infection and Evolution. D.H. is supported by a Wellcome Trust senior investigator award (100320/Z/12/Z). We thank M. Chase-Topping for statistical support.

Author Contributions Conceived the study (K.M.), developed and carried out the screen (P.M., A.C., D.H., K.M.), cloned and analysed the screen outputs (B.M.M., P.M., A.C., A.I., D.H., K.M.), derived RNAi lines (B.M.M.) and analysed these *in vitro* (B.M.M.) and *in vivo* (B.M.M., J.Y.). Genomic analyses were performed by A.I.; F.R. optimized culture methods for pleomorphic cells and performed MitoTracker assays. Northern blots were performed by P.M., cell-cycle scoring and analysis was carried out by B.M.M. The paper was written by K.M., B.M.M., P.M. and A.I.

Author Information Data in this submission has been submitted to the NCBI GEO database with accession code GSE46501. 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. (keith.matthews@ed.ac.uk).

LETTER

doi:10.1038/nature12861

Mutational and fitness landscapes of an RNA virus revealed through population sequencing

Ashley Acevedo[1], Leonid Brodsky[2] & Raul Andino[1]

RNA viruses exist as genetically diverse populations[1]. It is thought that diversity and genetic structure of viral populations determine the rapid adaptation observed in RNA viruses[2] and hence their pathogenesis[3]. However, our understanding of the mechanisms underlying virus evolution has been limited by the inability to accurately describe the genetic structure of virus populations. Next-generation sequencing technologies generate data of sufficient depth to characterize virus populations, but are limited in their utility because most variants are present at very low frequencies and are thus indistinguishable from next-generation sequencing errors. Here we present an approach that reduces next-generation sequencing errors and allows the description of virus populations with unprecedented accuracy. Using this approach, we define the mutation rates of poliovirus and uncover the mutation landscape of the population. Furthermore, by monitoring changes in variant frequencies on serially passaged populations, we determined fitness values for thousands of mutations across the viral genome. Mapping of these fitness values onto three-dimensional structures of viral proteins offers a powerful approach for exploring structure–function relationships and potentially uncovering new functions. To our knowledge, our study provides the first single-nucleotide fitness landscape of an evolving RNA virus and establishes a general experimental platform for studying the genetic changes underlying the evolution of virus populations.

To overcome the limitations of next-generation sequencing error, we developed circular sequencing (CirSeq), wherein circularized genomic RNA fragments are used to generate tandem repeats that then serve as substrates for next-generation sequencing (for DNA adaptation, see ref. 4). The physical linkage of the repeats, generated by 'rolling circle' reverse transcription of the circular RNA template, provides sequence redundancy for a genomic fragment derived from a single individual within the virus population (Fig. 1a and Extended Data Fig. 1). Mutations that were originally present in the viral RNA will be shared by all the repeats. Differences within the linked repeats must originate from enzymatic or sequencing errors and can be excluded from the analysis computationally. A consensus generated from a three-repeat tandem reduces the theoretical minimum error probability associated with current Illumina sequencing by up to 8 orders of magnitude, from 10^{-4} to 10^{-12} per base. This accuracy improvement reduces sequencing error to far below the estimated mutation rates of RNA viruses (10^{-4} to 10^{-6}) (ref. 5), allowing capture of a near-complete distribution of mutant frequencies within RNA virus populations.

We used CirSeq to assess the genetic composition of populations of poliovirus replicating in human cells in culture. Starting from a single viral clone, poliovirus populations were obtained following 7 serial passages (Fig. 2a). At each passage, 10^6 plaque forming units (p.f.u.) were used to infect HeLa cells at low multiplicity of infection (m.o.i. ~0.1) for a single replication cycle (8 h) at 37 °C (Methods).

We assessed the accuracy of CirSeq relative to conventional next-generation sequencing by estimating overall mutation frequencies as a function of sequence quality (Fig. 1b). The observed mutation frequency using CirSeq analysis was significantly lower than that using conventional analysis of the same data (Fig. 1b). In contrast to conventional next-generation sequencing, the mutation frequency in the CirSeq consensus was constant over a large range of sequencing quality scores (Fig. 1b and Extended Data Fig. 2, quality scores from 20 to 40). The mutation frequency obtained in the stable range of the CirSeq analysis is similar to previously reported mutation frequencies in poliovirus populations—approximately 2×10^{-4} mutations per nucleotide[3,6] (Fig. 2b and Extended Data Table 1).

We also compared transition-to-transversion ratios (ts:tv) obtained by CirSeq and conventional next-generation sequencing. Although purine (A/G) to purine, or pyrimidine (C/T) to pyrimidine transitions (ts) are the most commonly observed mutations in most organisms[7], error stemming from Illumina sequencing exhibits substantial purine to pyrimidine or pyrimidine to purine transversion (tv) bias[8]. This bias is reduced using CirSeq, as resulting ts:tv ratios are significantly higher than in the conventional repeat analysis (Fig. 1c). Notably, even if conventional next-generation data are filtered at high sequence quality (that is, quality scores over 30), the ts:tv ratio is still up to 10 times lower than that obtained with CirSeq. Thus, filtering conventional data fails to eliminate most sequencing errors (Fig. 1c). Our results indicate that CirSeq efficiently reduces errors generated during sequencing, producing mutation frequencies and ts:tv ratios consistent with the high values expected for poliovirus[6,9,10].

Using these results, we selected an average quality score of 20 as a threshold for further CirSeq analysis. This threshold corresponds to an estimated error probability of 10^{-6} (see Methods), setting a limit of detection for minor genetic variants two orders of magnitude below the expected average mutation frequency for RNA viruses. In comparison, the same quality threshold of 20, generally accepted for conventional analysis of next-generation sequencing data, limits variant detection to a minimum of 1% (ref. 11), two orders of magnitude higher than the average mutation frequency of many RNA viruses.

With an average coverage of more than 200,000 reads per position (Extended Data Fig. 3a), we detected on average more than 16,500 variants, ~74% of all possible variant alleles, per population per passage (Fig. 2b and Extended Data Table 1). Many alleles were detected for virtually all positions in the genome: mutations for all three alternative alleles (from the remaining three possible alternative nucleotides) were detected at 45.7% of genome positions; mutations for two of three were detected at 42% of positions; and mutations for only one alternative allele were detected at 12.2% of positions. The vast majority of variants are homogenously distributed at low frequencies between 10^{-3} and 10^{-5}, with very few populating the range between 1 and 10^{-3} (Fig. 2c). Thus, we can infer that the structure of a virus population replicating in the stable environment used here, is characterized by a sharp peak, representing the population consensus sequence, surrounded by a dense array of diverse variants present at very low frequencies (Extended Data Fig. 5a).

Mutation rates are central to evolution, as the rate of evolution is determined by the rate at which mutations are introduced into the

[1]Department of Microbiology and Immunology, University of California, San Francisco, California 94122–2280, USA. [2]Tauber Bioinformatics Research Center and Department of Evolutionary & Environmental Biology, University of Haifa, Mount Carmel, Haifa 31905, Israel.

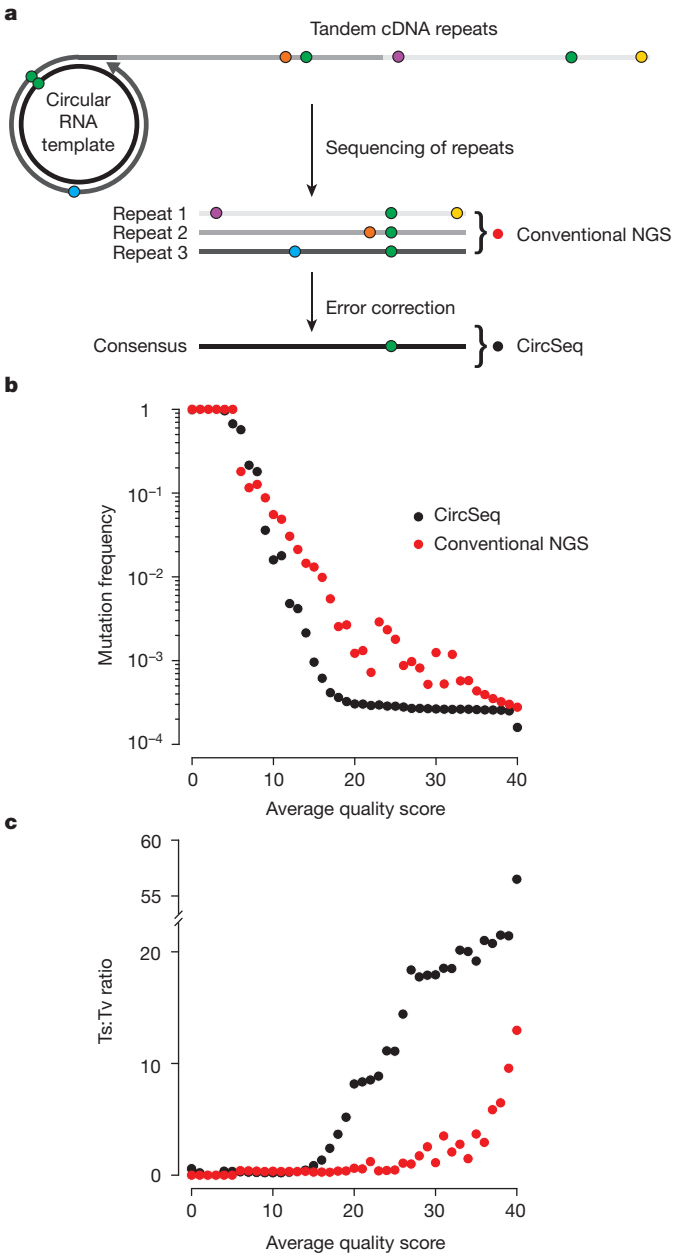


Figure 1 | CirSeq substantially improves data quality. **a**, Schematic of the CirSeq concept. Circularized genomic fragments serve as templates for rolling-circle replication, producing tandem repeats. Sequenced repeats are aligned to generate a majority logic consensus (Methods). Green symbols represent true genetic variation. Other coloured symbols represent random sequencing error. NGS, next-generation sequencing. **b**, **c**, Comparison of overall mutation frequency (**b**) and transition:transversion ratio (**c**) for repeats analysed as three independent sequences (red circles) or as a consensus (black circles). High-quality scores indicate low error probabilities. Quality scores are represented as averages because the consensus quality score is the product of quality scores from each repeat. Data was obtained from a single passage.

population[12,13]. Determination of virus mutation rates is difficult and often unreliable because accuracy depends on observing rare events[5]. We employed CirSeq to measure the rates for each type of mutation occurring during poliovirus replication *in vivo*. To do so, we estimated the frequency of lethal mutations, which are produced anew in each generation at a frequency equal to the mutation rate[14]. These included mutations producing stop codons within the virus polyprotein or those causing amino acid substitutions at catalytic sites of the essential viral enzymes 2A, 3C and 3D[15–17]. We find that mutation rates vary by more than two orders of magnitude depending on mutation type, transitions averaging 2.5×10^{-5} to 2.6×10^{-4} substitutions per site and transversions averaging 1.2×10^{-6} to 1.5×10^{-5} substitutions per site (Fig. 3). Even within these groups, transitions or transversions, the rates of the various nucleotide changes differ by an order of magnitude (Fig. 3). These nucleotide-specific differences in mutation rate likely reflect the molecular mechanism of viral polymerase fidelity, which may ultimately provide a means for the directionality of evolution. For example, C to U and G to A transitions accumulate up to 10 times faster than U to C and A to G; this inequality may provide a mechanistic basis for Dollo's law of irreversibility[18] because the likelihood of moving in one direction in sequence space is not equivalent to the reverse. Our analysis of mutation rates is consistent with biochemical estimations[9] and provides a physiological view of how the spectrum of mutation rates contribute to the genetic diversity of virus populations.

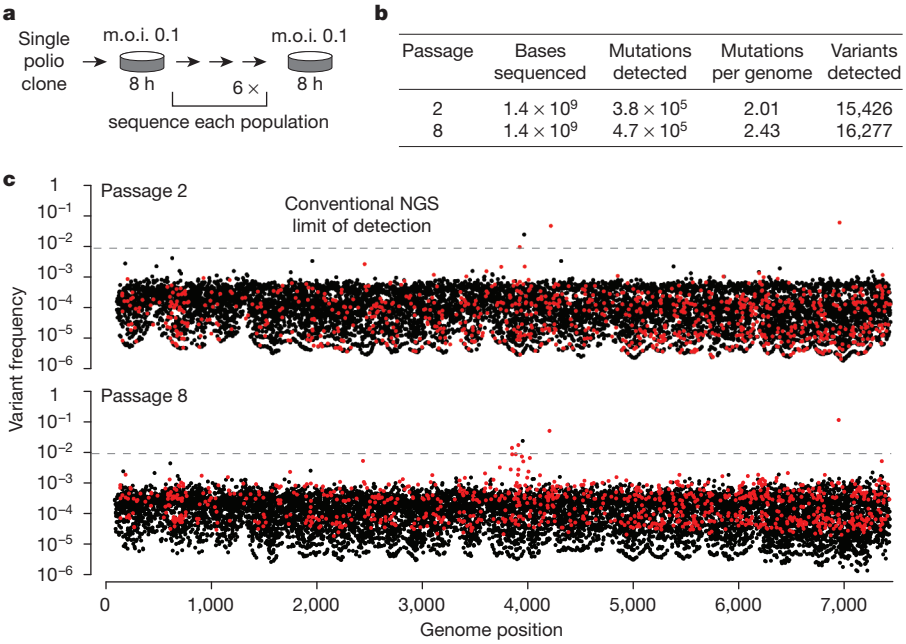
We next measured the fitness of each allele in the population by determining the change in mutation frequency for each variant over the course of seven passages (Fig. 2a). Variant frequency is governed by mutation and selection[19], assuming that our experimental conditions (low m.o.i. and large population size at each passage) minimize genetic drift and complementation. We employed a simple model based on classical population genetics to estimate fitness:

$$\frac{a_t}{A_t} = \frac{a_{t-1}}{A_{t-1}} \bullet w_{rel} + \mu_{t-1} \quad (1)$$

where a and A are the counts of variant and wild type alleles, respectively, w_{rel} is the relative fitness of a to A (ratio of growth rates), t is time in generations (infection cycles) and μ is the specific rate of mutation from A to a. We measured proportions of A and a over the seven passages and, using mutation rates we previously determined (Fig. 3), calculated w_{rel} for mutations across the viral genome. The current length limitations of next-generation sequencing preclude CirSeq from providing direct information about haplotypes. Accordingly, our fitness measurements represent the average relative fitness of the population of haplotypes containing a variant allele compared to the population of haplotypes containing the wild-type allele at that position (see Supplementary Information).

Overall, the distribution of mutational fitness effects we obtained (Fig. 4a) is highly consistent with previous small-scale analyses of RNA viruses[20–22], validating CirSeq as a robust method for large-scale fitness measurement. In our analysis, the non-lethal distribution of mutational fitness effects for synonymous mutations is centred near neutrality (Fig. 4a), reflecting the predominantly neutral effects anticipated for synonymous mutations. In contrast, the distribution of non-lethal mutational fitness effects for non-synonymous mutations encompasses primarily deleterious mutations, consistent with previous findings[21–23].

Notably, despite the expectation that synonymous mutations will have relatively low impact on fitness, a significant fraction of synonymous changes were subject to strong selection, with 2% being highly beneficial (relative fitness >1.2) and 10% being lethal (Fig. 4a and Extended Data Fig. 6c). Synonymous mutations under strong selection are relatively evenly dispersed throughout the coding sequence, rather than clustered at known functional elements (Extended Data Fig. 6a). Given that the entire capsid-coding region can be deleted without disrupting replication or translation, indicating that this region contains no essential RNA structural elements, it is probable that RNA structure is not the primary driving force behind strong selection of synonymous mutants in poliovirus. Although it is possible that observed mutational fitness effects could be the result of codon usage or codon pair bias, in practice, deoptimization of these biases does not result in lethality based on single nucleotide substitutions[24,25]. Future studies will be necessary to elucidate the mechanisms modulated by these synonymous mutations. Furthermore, the variance in fitness for non-synonymous mutations was significantly larger ($P < 0.001$, Extended Data Fig. 6c) than for synonymous; indeed the largest beneficial fitness effects (not shown in Fig. 4a) were the result of non-synonymous



Passage	Bases sequenced	Mutations detected	Mutations per genome	Variants detected
2	1.4×10^9	3.8×10^5	2.01	15,426
8	1.4×10^9	4.7×10^5	2.43	16,277

Figure 2 | CirSeq reveals the mutational landscape of poliovirus. **a**, Experimental evolution paradigm. A single plaque was isolated, amplified and then serially passaged at low multiplicity of infection (m.o.i.). Low m.o.i. passages were amplified to produce sufficient quantities of RNA for library preparation (Methods). **b**, Summary of population metrics obtained by CirSeq. **c**, Frequencies of variants detected using CirSeq are mapped to nucleotide position with the genome for passages 2 and 8. The conventional next-generation sequencing limit of detection (1%) is indicated by dashed lines. Each position contains up to three variants. Variants are coloured based on relative fitness, black indicating lethal or detrimental and red indicating beneficial. Sampling error can affect variant frequencies (see Methods and Extended Data Fig. 4a, b).

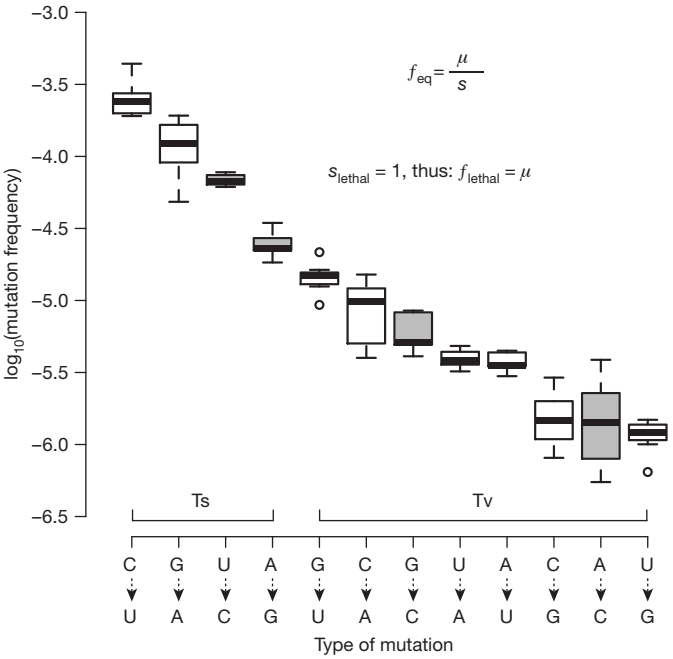


Figure 3 | Determination of *in vivo* mutation rates of poliovirus. **a**, The frequency of deleterious mutations at mutation–selection balance is the mutation rate (μ) over the deleterious selection coefficient (*s*), see inset. For lethal mutations, $s = 1$, thus their frequencies equal the mutation rate. Nonsense mutations and catalytic site substitutions were used to obtain lethal mutation frequencies, and thus mutation rates, for each mutation type. Grey boxes were measured using only catalytic site mutants. $n = 7$ (biological replicates), whiskers represent the lowest and highest datum within 1.5 inner quartile range of the lower and upper quartile, respectively.

substitutions. Notably, a large number of substitutions are beneficial (145 significantly beneficial mutations, see Methods), indicating the potential for a highly dynamic population structure, where selection for minor genetic components constantly drives the population to new regions of sequence space, even in a relatively constant environment.

The genome-wide distribution of mutational fitness effects does not apply uniformly to each protein as non-synonymous mutations exhibit distinct mutational fitness effects distributions in structural genes (those encoding the viral capsid) and non-structural genes (encoding enzymes and factors involved in viral replication) (Fig. 4b, Extended Data Fig. 6b for synonymous). Although non-structural genes show slightly lower mean mutational fitness effects when considering lethal mutants, they have significantly larger variance in mutational fitness effects ($P < 0.001$, Extended Data Fig. 6c), indicating that these proteins may have intrinsic differences in their tolerance of mutations. These differences may relate to biophysical properties, like stability constraints[26], or the density of functional residues, for example, non-structural proteins often play multifunctional roles and participate in a greater number of host–pathogen interactions[27].

To investigate further the relationship between mutational fitness effects and protein structure and function, we mapped fitness values onto the three-dimensional structure of the well characterized poliovirus RNA-dependent RNA polymerase[28]. We find a remarkable agreement between our fitness data and known structure–function relationships in this enzyme (see Supplementary Information and Extended Data Table 2). For example, many detrimental mutations map to residues associated with RNA binding and catalysis in the central chamber of the polymerase (Fig. 4d, red). Intriguingly, two clusters of beneficial mutations, discontinuous on the genome sequence, mapped to uncharacterized and structurally contiguous regions on the surface of the polymerase (Fig. 4c, blue). Our data suggest that this domain must be functionally relevant to viral replication, as it is clearly tuned by evolution over the course of passaging. Such genome-wide fitness calculations enabled by CirSeq, combined with structural information, can provide high-definition, bias-free insights into structure–function relationships, potentially revealing novel functions for viral proteins

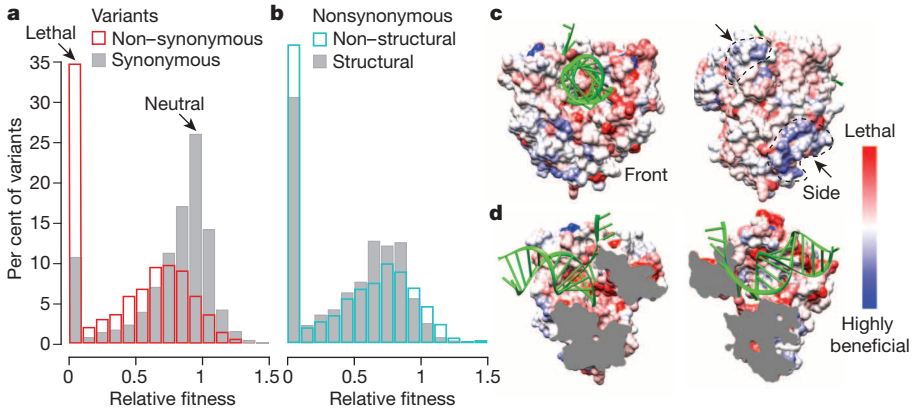


Figure 4 | Fitness landscape defines structure–function relationships. **a**, **b**, Distributions of fitness for synonymous (grey) and non-synonymous (red) mutations (**a**) and for non-synonymous mutations in structural (grey) and non-structural (blue) genes (**b**). Fitness was determined as described in Methods. C > U and G > A transitions were excluded as we observed indications of hypermutation for these variants. The proportion of lethal variants for each group is likely higher, as not all possible variants were detected. Variants with fitness >1.5 are not shown. **c**, **d**, The most fit non-synonymous variant observed for each codon was mapped onto the viral polymerase (3OL6)[28] using a red (lethal) to white (neutral) to blue (beneficial) scale. RNA is coloured green. Front and side views show two positively selected surfaces (marked by arrows) (**c**) and split view shows negative selection along active core and RNA binding sites (**d**).

and RNA structures, as well as nuanced insights into a viral genome's phenotypic space. Such analyses have the power to reveal protein residues or domains that directly correspond to viral functional plasticity and may significantly inform our structural and mechanistic understanding of host–pathogen interactions.

The analytical approach we describe provides an opportunity to examine and quantify evolutionary dynamics at nucleotide resolution on a genome-wide scale and to integrate evolutionary information with structural and physiological data. Such large-scale measurements of fitness are a fundamental step in understanding the effects of mutation on phenotype and evolutionary trajectory. Modelling the evolutionary dynamics of infection, transmission, host-switching and drug resistance may be central for developing innovative strategies for drug and vaccine design, personalized treatment and the containment of emerging viruses.

METHODS SUMMARY

Viral populations were obtained by serial passaging of a single poliovirus clone at m.o.i. of 0.1. Populations were amplified *in vivo* before library preparation to increase the ratio of viral to cellular RNA. RNA extracted from amplified populations was polyA purified, Zn^{2+} fragmented, size selected (Extended Data Fig. 3b), circularized, reverse transcribed and then cloned by standard mRNA sequencing library preparation methods (Extended Data Fig. 1).

Libraries were sequenced 323 cycles on an Illumina MiSeq. Custom analysis software, using Bowtie 2 (ref. 29) for sequence mapping, was developed to identify and align repeats, generate a consensus by majority logic and recalculate estimated error probabilities.

Consensus data was filtered at average quality score 20, where the estimated error probability is 10^{-6} ($10^{-2} \times 10^{-2} \times 10^{-2}$ for three repeats). The statistical significance of mutations detected was determined by a one-sided binomial test in R using the average estimated error probability at each genome position as the null probability of success. The accuracy of frequencies (Extended Data Fig. 4) was estimated using the standard error of a binomial distribution.

Fitness was determined using variant frequencies over seven passages (Extended Data Fig. 7) and a regression model, equation (1), describing the change in frequency of variants over time based on their selection and the accumulation of *de novo* mutations, assuming that the counts of the variant allele are negligible compared to wild type and that selection is constant over the series of passages (Extended Data Fig. 5b). Genetic drift was accounted for in our fitness calculations by simulating random fluctuations in variant frequencies in our fitness model (Extended Data Fig. 8). The highest relative fitness value of non-synonymous mutations observed at each codon of the viral polymerase was mapped to the polymerase structure (Protein Data Bank accession code 3OL6)[28] using UCSF Chimera[30].

A complete description of the materials and methods used to generate this data and its result is provided in the 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 12 April; accepted 11 November 2013.
Published online 27 November 2013.

1. Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13,** 735–744 (1978).
2. Burch, C. L. & Chao, L. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406,** 625–628 (2000).
3. Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439,** 344–348 (2006).
4. Lou, D. I. *et al.* High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl Acad. Sci. USA.* (in the press).
5. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84,** 9733–9748 (2010).
6. Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl Acad. Sci. USA* **98,** 6895–6900 (2001).
7. Wakeley, J. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol. Evol.* **11,** 158–162 (1996).
8. Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36,** e105 (2008).
9. Freistadt, M. S., Vaccaro, J. A. & Eberle, K. E. Biochemical characterization of the fidelity of poliovirus RNA-dependent RNA polymerase. *Virol. J.* **4,** 44 (2007).
10. Arnold, J. J. & Cameron, C. E. Poliovirus RNA-dependent RNA polymerase ($3D^{pol}$): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg^{2+}. *Biochemistry* **43,** 5126–5137 (2004).
11. Radford, A. D. *et al.* Application of next-generation sequencing technologies in virology. *J. Gen. Virol.* **93,** 1853–1868 (2012).
12. Orr, H. A. The rate of adaptation in asexuals. *Genetics* **155,** 961–968 (2000).
13. Kimura, M. *The Neutral Theory of Molecular Evolution* 55–97 (Cambridge Univ. Press, 1983).
14. Cuevas, J. M., González-Candelas, F., Moya, A. & Sanjuán, R. Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus *in vivo*. *J. Virol.* **83,** 5760–5764 (2009).
15. Hämmerle, T., Hellen, C. U. & Wimmer, E. Site-directed mutagenesis of the putative catalytic triad of poliovirus 3C proteinase. *J. Biol. Chem.* **266,** 5412–5416 (1991).
16. Hellen, C. U. T., Lee, C.-K. & Wimmer, E. Determinants of substrate recognition by poliovirus 2A proteinase. *J. Virol.* **66,** 3330–3338 (1992).
17. Gohara, D. W. *et al.* Poliovirus RNA-dependent RNA polymerase ($3D^{pol}$): structural, biochemical, and biological analysis of conserved structural motifs A and B. *J. Biol. Chem.* **275,** 25523–25532 (2000).
18. Gould, S. J. Dollo on Dollo's law: irreversibility and the status of evolutionary laws. *J. Hist. Biol.* **3,** 189–212 (1970).
19. Haldane, J. B. S. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Math. Proc. Camb. Philos. Soc.* **23,** 838–844 (1927).
20. Cuevas, J. M., Domingo-Calap, P. & Sanjuán, R. The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol. Biol. Evol.* **29,** 17–20 (2012).

21. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nature Rev. Genet.* **8,** 610–618 (2007).
22. Sanjuán, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101,** 8396–8401 (2004).
23. Chao, L. Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348,** 454–455 (1990).
24. Mueller, S., Papamichail, D., Coleman, J. R., Skiena, S. & Wimmer, E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.* **80,** 9687–9696 (2006).
25. Coleman, J. R. *et al.* Virus attenuation by genome-scale changes in codon pair bias. *Science* **320,** 1784–1787 (2008).
26. Tokuriki, N. & Tawfik, D. Protein dynamism and evolvability. *Science* **324,** 203–207 (2009).
27. Jäger, S. *et al.* Global landscape of HIV–human protein complexes. *Nature* **481,** 365–370 (2012).
28. Gong, P. & Peersen, O. B. Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl Acad. Sci. USA* **107,** 22505–22510 (2010).
29. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
30. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

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

Acknowledgements We thank J. Frydman, S. Bianco, H. Dawes, K. Ehmsen and members of the Andino laboratory for critical reading of the manuscript and G. Schroth, M. Harrison, P. Wassam and T. Collins for technical advice. This work was financially supported by a National Science Foundation graduate research fellowship to A.A., NIAID AI091575, AI36178 and AI40085 to R.A., and DARPA Prophecy to R.A. and L.B.

Author Contributions R.A. and A.A. conceived and designed the experiments. A.A. performed experiments and sequencing. A.A. and L.B. analysed the data and performed statistical analyses. R.A. and A.A. wrote the manuscript.

Author Information Sequencing data has been deposited in the NCBI Sequence Read Archive under accession number PRJNA222998. Software complementary to this analysis is available at http://andino.ucsf.edu. 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.A. (raul.andino@ucsf.edu).

LETTER

doi:10.1038/nature12862

Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity

John W. Schoggins[1]†, Donna A. MacDuff[2], Naoko Imanaka[1], Maria D. Gainey[3], Bimmi Shrestha[4], Jennifer L. Eitson[5], Katrina B. Mar[5], R. Blake Richardson[5], Alexander V. Ratushny[6,7], Vladimir Litvak[6], Rea Dabelic[8], Balaji Manicassamy[9], John D. Aitchison[6,7], Alan Aderem[6], Richard M. Elliott[10]†, Adolfo García-Sastre[11,12,13], Vincent Racaniello[8], Eric J. Snijder[14], Wayne M. Yokoyama[3], Michael S. Diamond[2,4], Herbert W. Virgin[2] & Charles M. Rice[1]

The type I interferon (IFN) response protects cells from viral infection by inducing hundreds of interferon-stimulated genes (ISGs), some of which encode direct antiviral effectors[1–3]. Recent screening studies have begun to catalogue ISGs with antiviral activity against several RNA and DNA viruses[4–13]. However, antiviral ISG specificity across multiple distinct classes of viruses remains largely unexplored. Here we used an ectopic expression assay to screen a library of more than 350 human ISGs for effects on 14 viruses representing 7 families and 11 genera. We show that 47 genes inhibit one or more viruses, and 25 genes enhance virus infectivity. Comparative analysis reveals that the screened ISGs target positive-sense single-stranded RNA viruses more effectively than negative-sense single-stranded RNA viruses. Gene clustering highlights the cytosolic DNA sensor cyclic GMP-AMP synthase (*cGAS*, also known as *MB21D1*) as a gene whose expression also broadly inhibits several RNA viruses. *In vitro*, lentiviral delivery of enzymatically active cGAS triggers a STING-dependent, IRF3-mediated antiviral program that functions independently of canonical IFN/STAT1 signalling. *In vivo*, genetic ablation of murine cGAS reveals its requirement in the antiviral response to two DNA viruses, and an unappreciated contribution to the innate control of an RNA virus. These studies uncover new paradigms for the preferential specificity of IFN-mediated antiviral pathways spanning several virus families.

To identify IFN-induced effectors targeting diverse viruses, we used our established flow cytometry-based ISG screening platform (Fig. 1a, see Methods)[4]. We screened over 350 common ISGs for inhibitory or enhancing effects on 14 viruses, including one double-stranded DNA (dsDNA) virus, six positive-sense single-stranded RNA (+ssRNA) viruses and seven negative-sense single-stranded RNA (−ssRNA) viruses (Extended Data Table 1 and Fig. 1 legend for abbreviations). Viruses were screened in HeLa cells, Huh7 hepatoma cells or human *STAT1*$^{-/-}$ fibroblasts[14]. Notably, infection of most +ssRNA viruses was inhibited by greater than 50% when any of multiple ISGs were expressed, whereas screens for vaccinia virus (VV), human metapneumovirus, respiratory syncytial virus, measles virus and Bunyamwera virus had few or no genes that inhibited virus infection by more than 50% (Fig. 1b and Supplementary Tables 1 and 2).

Confirmatory assays were performed on selected ISGs to verify the primary screening hits. A total of 159 assays representing 96 unique genes were performed (Fig. 2a, b). Of these, 125 assays representing 72 unique genes yielded results that were consistent with the primary screens. We identified 47 inhibitory and 25 enhancing ISGs. Of these, 25 inhibitory ISGs suppressed infectivity by more than 50%, and 4 enhancing ISGs increased infectivity by more than 150%. We also confirmed by plaque assay that ISGs with antiviral effects against green fluorescent protein (GFP)-tagged poliovirus could inhibit the parental non-GFP strain (Extended Data Fig. 1). Comparative analysis of the confirmatory assays on RNA viruses indicated that the ectopically

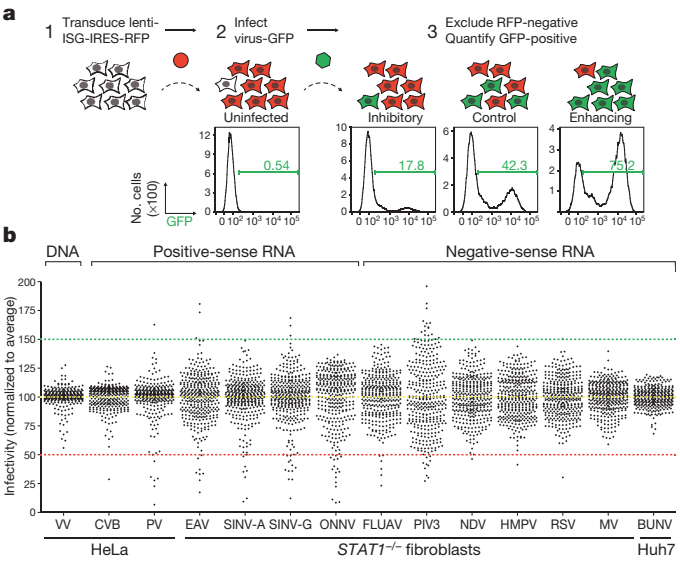


Figure 1 | Flow cytometry-based screens for identifying inhibitory or enhancing ISGs against 14 viruses. a, Schematic of the ectopic expression screen showing cells transduced with lentiviral vectors expressing an inhibitory ISG (*IFITM3*) and red fluorescent protein (RFP), a control (Fluc), or an enhancing ISG (*MCOLN2*). Cells were infected with influenza A virus expressing GFP, and GFP-positive cells were quantified by flow cytometry. **b**, Dot plots of virus infectivity in the presence of expressed ISGs. Data sets were normalized to the average of each screen, which is indicated by a yellow dotted line. The 50% inhibitory and 150% enhancing effects are denoted by red and green dotted lines, respectively. Bunyamwera virus, BUNV; Coxsackie B virus, CVB; equine arterivirus, EAV; influenza A virus, FLUAV; human metapneumovirus, HMPV; measles virus, MV; Newcastle disease virus, NDV; o'nyong-nyong virus, ONNV; human parainfluenza virus type 3, PIV3; poliovirus, PV; respiratory syncytial virus, RSV; Sindbis virus AR86, SINV-A; Sindbis virus Girdwood, SINV-G.

[1]Laboratory of Virology and Infectious Disease, The Rockefeller University, New York, New York 10065, USA. [2]Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. [3]Rheumatology Division, Department of Medicine, and Howard Hughes Medical Institute, Washington University School of Medicine, St Louis, Missouri 63110, USA. [4]Infectious Diseases Division, Department of Medicine and Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. [5]Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [6]Seattle Biomedical Research Institute, Seattle, Washington 98109, USA. [7]Institute for Systems Biology, Seattle, Washington 98109, USA. [8]Department of Microbiology and Immunology, Columbia University, New York, New York 10032, USA. [9]Department of Microbiology, University of Chicago, Chicago, Illinois 60637, USA. [10]School of Biology, University of St Andrews, St Andrews, Scotland KY16 9ST, UK. [11]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [12]Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [13]Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [14]Department of Medical Microbiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. †Present addresses: Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA (J.W.S.); MRC-University of Glasgow Centre for Virus Research, Glasgow, Scotland G61 1QH, UK (R.M.E.).

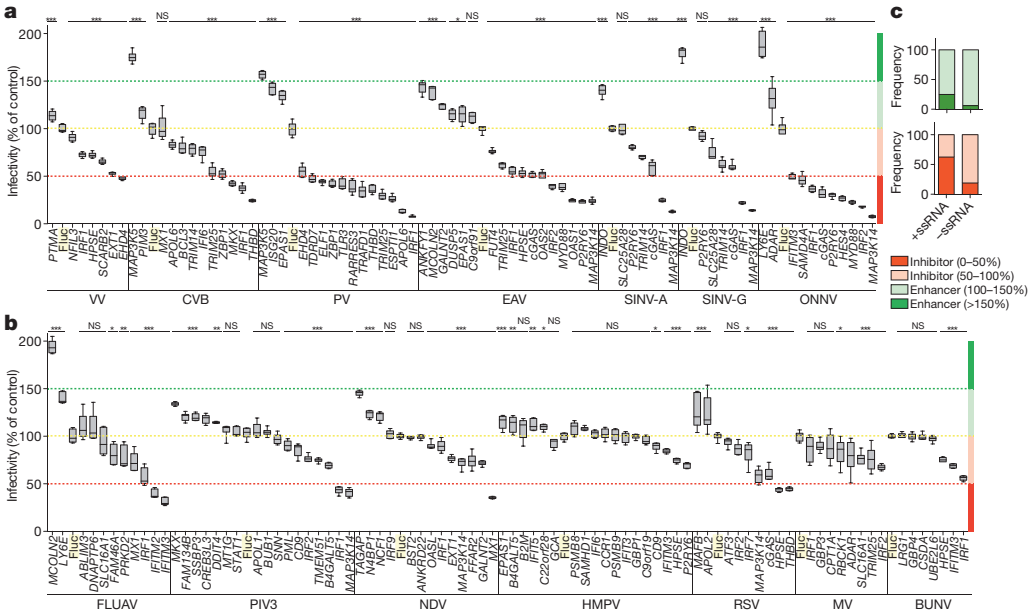


Figure 2 | Confirmatory assays for selected inhibitory and enhancing ISGs. **a**, **b**, Independent confirmatory assays against DNA and +ssRNA viruses (**a**) and –ssRNA viruses (**b**) were carried out using new lentivirus stocks. Data were normalized to a Fluc control, highlighted in a yellow box and by dotted line. Data are presented as box and whisker plots; whiskers extend to show the highest and lowest values. Data represent six technical replicates to control for intra-assay variability. Statistical significance was determined by one-way analysis of variance (ANOVA) (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; NS, not significant). The 50% inhibitory and 150% enhancing effects are denoted by red and green dotted lines, respectively. **c**, Comparative analysis showing the frequency with which confirmed inhibitory or enhancing ISGs targeted +ssRNA or –ssRNA viruses.

expressed ISGs with the strongest inhibitory (<50%) and enhancing (>150%) effects were biased towards the +ssRNA viruses (Fig. 2c).

We next performed a hierarchical clustering analysis of the primary screening data to group viruses and ISGs with one another (Fig. 3a). We considered only viruses screened in *STAT1*$^{-/-}$ fibroblasts, including three +ssRNA viruses from previous studies: two flaviviruses, West Nile virus (WNV) and yellow fever virus, and one alphavirus, chikungunya virus[4]. The clustering data revealed a division of viruses into two major groups representing either +ssRNA and –ssRNA viruses. Within these groups, several related viruses, including the flaviviruses, alphaviruses and paramyxoviruses, clustered together. We repeated the analysis in the absence of 1, 2 or 6 of the more potent ISGs and obtained similar results, indicating that the clustering was not skewed by a selection of dominant genes (Extended Data Fig. 2a–e). In a second clustering analysis, viruses were grouped based on the presence of ISG names in a list of the top 30 genes from the primary screens (Supplementary Tables 1 and 2). This analysis revealed a similar division of +ssRNA and –ssRNA viruses (Extended Data Fig. 2f). A co-occurrence analysis of the top 20 antiviral genes from 7 +ssRNA and 5 –ssRNA virus screens further supported the hierarchical clustering studies (Extended Data Fig. 3). These data suggest that subsets of ISGs may target similar viruses and raise the possibility that therapeutics targeting specific antiviral host molecules may be broad spectrum across related viruses.

To distinguish direct ISG effectors from transcriptional regulators, we tested inhibitory ISGs for interferon-stimulated response element (ISRE)-dependent transcription. Of the 68 genes tested, only *IRF1*, *IRF2*, *TLR3* and *MYD88* directly activated an ISRE-driven reporter plasmid (Extended Data Fig. 4a). We also tested whether 4 ISGs with virus enhancing activity could impair IFN-mediated ISRE activation. In contrast to *SOCS1*, a known negative regulator of IFN signalling, none of the ISGs had an effect (Extended Data Fig. 4b). These data indicate that most ISG 'hits' do not affect ISRE-dependent gene transcription. They may have direct effector mechanisms or may regulate other pathways, as suggested by Gene Ontology analysis. (Supplementary Table 3).

The clustering analysis grouped the antiviral transcription factor IRF1 and cyclic GMP-AMP synthase cGAS (Fig. 3a). We first identified the gene encoding cGAS (formerly *C6orf150*) as antiviral in previous screens[4], and our current studies confirmed this with additional viruses. Because both cGAS and IRF1 are broadly antiviral, we proposed that cGAS, like IRF1, might upregulate antiviral gene transcription. We found that *STAT1*$^{-/-}$ fibroblasts transduced with lentiviruses expressing cGAS and IRF1, but not IRF7 or firefly luciferase (Fluc), had increased messenger RNA levels of the ISG *OAS2* (Fig. 3b). We extended these findings with microarray analysis and showed that lentiviral-mediated expression of cGAS induced 60 genes by at least twofold compared to Fluc control. (Fig. 3c and Extended Data Table 2). Many of these genes are ISGs, and more than half of them overlap with IRF1-induced transcripts in the same cellular background[4]. These results indicate that in *STAT1*$^{-/-}$ fibroblasts, lentiviral-mediated expression of cGAS induces an antiviral program independently of canonical IFN signalling.

During the course of these studies, murine cGAS was shown to be a cytosolic DNA-sensing enzyme that catalyses the production of cyclic GMP-AMP (cGAMP), a second-messenger activator of IFN antiviral responses[15,16]. The induction of IFN by cGAS appears to require a DNA, but not RNA, substrate to trigger a STING/IRF3 activation pathway. Nonetheless, we observed that cGAS induced an antiviral program that targeted several RNA viruses in *STAT1*$^{-/-}$ fibroblasts (Figs 2 and 3c), which are compromised in canonical IFN/STAT signalling[14]. We therefore proposed that lentiviral-driven cGAS expression triggers antiviral gene expression by direct STING/IRF3 activation. We confirmed STING expression in *STAT1*$^{-/-}$ fibroblasts (Fig. 3d), and showed that cells transduced with lentivirus expressing cGAS had a strong induction of phosphorylated IRF3 and *OAS2* mRNA compared to control cells (Fig. 3e). *OAS2* induction by cGAS was abrogated when STING expression was silenced with short interfering RNA (siRNA; Fig. 3f and Extended Data Fig. 5a), confirming a requirement for STING in the pathway. Consistent with this, IRF3 phosphorylation, *OAS2* mRNA induction and viral inhibition were not observed when lentiviruses expressing cGAS were used to transduce Huh7 cells, which lack detectable levels of STING (Fig. 3d, e). These data indicate that, in *STAT1*$^{-/-}$ fibroblasts, lentiviral-driven cGAS expression activates IRF3 through STING and establishes a transcriptional program that inhibits infection of several RNA viruses.

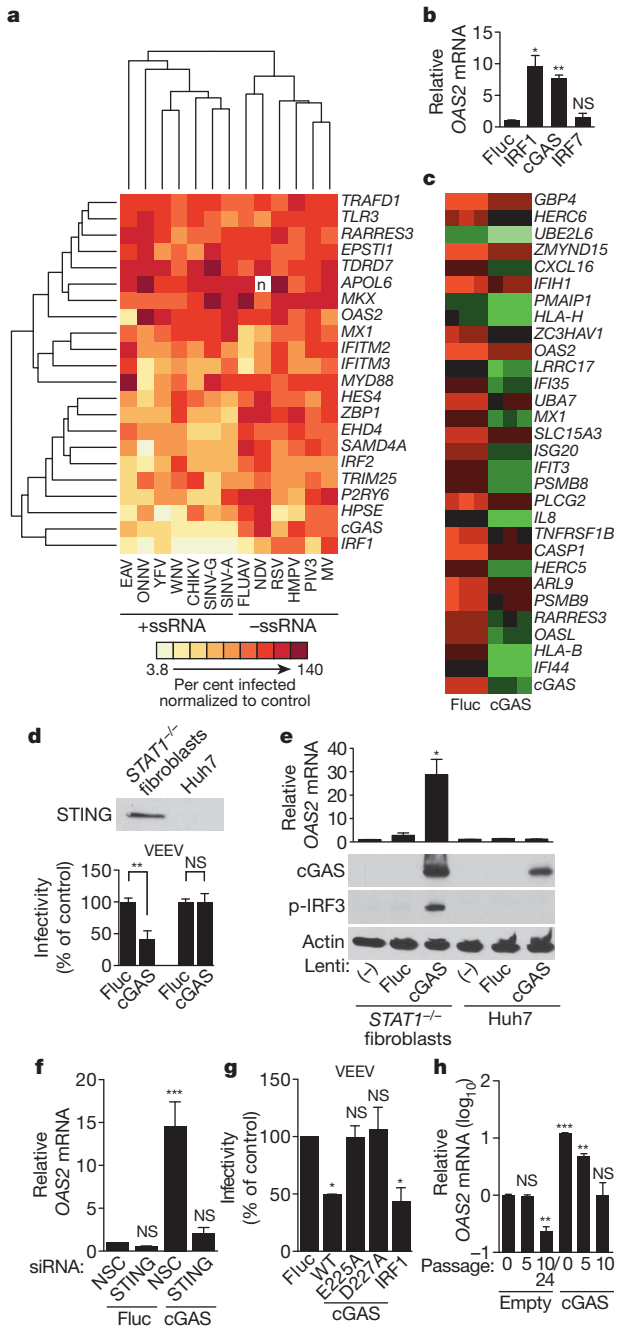


Figure 3 | cGAS activates an IRF3-driven antiviral program independently of canonical IFN/STAT1 signalling. **a**, Hierarchical clustering analysis of 22 ISGs and 13 viruses screened in *STAT1*$^{-/-}$ fibroblasts. CHIKV, chikungunya virus; YFV, yellow fever virus. **b**, *OAS2* gene expression in *STAT1*$^{-/-}$ fibroblasts transduced with lentiviruses expressing IRF7, IRF1, cGAS and Fluc. Data represent one of two experiments performed in duplicate. **c**, Microarray analysis of *STAT1*$^{-/-}$ fibroblasts transduced with lentiviruses expressing Fluc or cGAS. Data show a subset of genes (green) induced 2.5-fold with $P < 0.05$, $n = 3$ (using Benjamini–Hochberg false discovery rate correction). Data represent one of two independent experiments. **d**, Top, western blot of STING expression in *STAT1*$^{-/-}$ fibroblasts and Huh7 cells. Bottom, infectivity of Venezuelan equine encephalitis virus (VEEV) in *STAT1*$^{-/-}$ fibroblasts and Huh7 cells transduced with lentivirus expressing Fluc or cGAS. Virus infectivity was normalized to Fluc control. **e**, *OAS2* mRNA expression (top) and western blots (bottom) of cGAS, phosphorylated IRF3 (p-IRF3) and control actin in *STAT1*$^{-/-}$ fibroblasts and Huh7 cells transduced with lentiviruses expressing Fluc or cGAS. **f**, Antiviral gene expression in *STAT1*$^{-/-}$ fibroblasts that were depleted of STING by siRNA (Extended Data Fig. 5a) before transduction with lentiviruses expressing Fluc or cGAS. NSC, non-silencing siRNA control. **g**, Infectivity of VEEV in *STAT1*$^{-/-}$ fibroblasts transduced with Fluc control, wild type (WT) or point mutant (E225A, D227A) cGAS, or IRF1. Virus infectivity was normalized to Fluc control. **h**, *OAS2* mRNA induction in *STAT1*$^{-/-}$ fibroblasts stably expressing cGAS or an empty cassette. RNA samples were processed at the indicated cell passage number. In **d**–**g**, data represent the means of two or three independent experiments performed in triplicate. In **h**, data represent one of two independent experiments with similar results. Error bars represent s.d. Statistical significance was determined by *t*-test or one-way ANOVA (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; NS, not significant).

We next probed the mechanism of cGAS activation by performing genetic analyses to identify functional domains and residues (Extended Data Fig. 5b). Deletion analyses of cGAS localized the antiviral activity to the carboxy-terminal domain, with the first 164 amino acids being dispensable (Extended Data Fig. 5c). Active site mutants (E225A, D227A) showed no antiviral activity (Fig. 3g) and were impaired in IRF3 phosphorylation and *OAS2* mRNA induction (Extended Data Fig. 5d). These results are in agreement with recent studies showing the requirement for these residues in the synthesis of cGAMP[15,17–19]. Our data indicate that the antiviral effect of cGAS requires an active enzyme, and by extension, an activating substrate. We proposed that the lentivirus itself provides the trigger. Accordingly, we predicted that once cells stabilize from transient lentiviral infection, cGAS expression from the provirus would be less activating as the cells were passaged. Indeed, over at least 10 passages, we observed a progressive decrease in *OAS2* levels in cGAS-expressing and control cells (Fig. 3h), despite continuous and high levels of *cGAS* mRNA and protein in cGAS-expressing cells (Extended Data Fig. 5e). These data suggest that transient delivery of lentivirus may trigger the formation of a DNA-based substrate that reacts with cGAS to activate IRF3. A recent report supports this hypothesis by showing that cGAS can sense reverse-transcribed retroviral DNA[20]. However, given the selectivity of this effect against several +ssRNA viruses, we cannot rule out other mechanisms of cGAS activation.

We next determined whether these *in vitro* studies predict physiologically relevant functions of antiviral molecules. We generated mice with a targeted deletion of *cGas* exon 2, which contains the active site (Extended Data Fig. 6a, b). Knockout mice bred in normal Mendelian ratios and showed no overt growth or developmental defects. Gene expression analysis from the spleen (Fig. 4a), lungs, and bone marrow-derived macrophages (BMMO) (Extended Data Fig. 6c) of wild-type and knockout mice confirmed reduced *cGas* mRNA (Fig. 4a). As cGAS is activated *in vitro* by DNA[15,17,19], we challenged mice with two DNA viruses, murine gammaherpesvirus 68 (MHV68) and VV. Viral titres of MHV68 were 2.0-fold higher in the spleen and 3.5-fold higher in the lungs of *cGas*$^{-/-}$ mice compared to wild-type mice (Fig. 4b, c). VV had a notable mortality phenotype, with all *cGas*$^{-/-}$ mice succumbing to infection, whereas 70% of wild-type mice recovered (Fig. 4d). We next infected BMMO from wild-type and *cGas*$^{-/-}$ mice with MHV68 or VV and observed increased titres of both viruses in *cGas*$^{-/-}$ cells (Fig. 4e, f). *cGas*$^{-/-}$ BMMO were also refractory to the >200-fold induction of *Ifnb* (also known as *Infb1*) mRNA observed in MHV68-infected wild-type BMMO (Fig. 4g). These data provide direct genetic evidence that cGAS is required for innate control of DNA viruses in mice. A recently published study used *cGas*-deficient 'gene-trap' mice[21], which served as our starting point before excision of *cGas* exon 2 by sequential crossings to FlpE-deleter and Cre-expressing mice (Extended Data 6a). This study demonstrated that *cGas*-deficient gene-trap mice were also more vulnerable to infection by a DNA virus, herpes simplex virus 1. Thus, two variants of mice lacking cGAS establish a role for this sensor in the antiviral immunity to DNA viruses.

Our *in vitro* studies linked cGAS antiviral function to RNA virus inhibition through IRF3, and initial evidence suggests that lentivirus is the trigger. However, some RNA viruses were not targeted by this lentivirus/cGAS/IRF3 axis (Fig. 2a, b), prompting us to explore whether endogenous cGAS modulates RNA virus infection. Notably, *cGas*$^{-/-}$ mice were more vulnerable to lethal WNV infection compared to wild-type mice (Fig. 4h). We did not detect an increase in viral burden

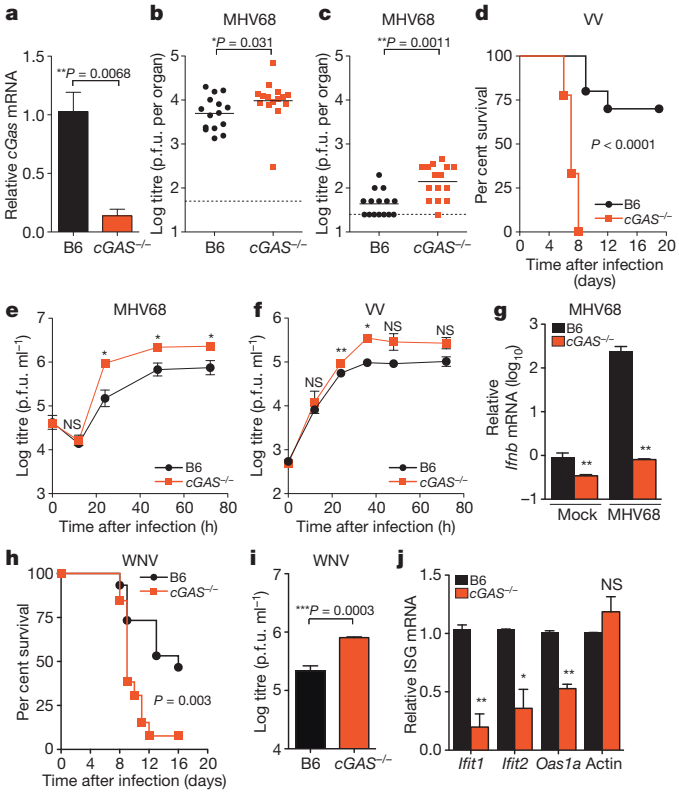


Figure 4 | Requirement for cGAS in controlling viral infection *in vivo*. **a**, Quantitative PCR with reverse transcription (qRT–PCR) of relative *cGas* expression in spleens of wild-type (B6) and $cGas^{-/-}$ mice. $n = 3$ mice per group. Error bars represent s.e.m. Statistical significance was determined by *t*-test. **b**, **c**, Viral titres in spleen (**b**) and lungs (**c**) of wild-type and $cGas^{-/-}$ mice after infection with 10^6 plaque-forming units (p.f.u.) of MHV68, $n = 15$ mice per group. Statistical significance was determined by Mann–Whitney test. **d**, Survival curves of wild-type and $cGas^{-/-}$ mice infected with 8000 p.f.u. of VV, $n = 10$ (wild type), $n = 9$ ($cGas^{-/-}$). Statistical significance was determined by log rank test. **e**, **f**, **i**, Viral titres from BMMO after infection with 10 multiplicity of infection (m.o.i.) MHV68 (**e**), 0.1 m.o.i. VV (**f**) or 0.1 m.o.i. WNV (**i**). Data represent three (**e**, **f**) or four (**i**) independent experiments performed in triplicate. Error bars represent s.e.m. Statistical significance was determined by *t*-test. **g**, qRT–PCR of relative *Ifnb* expression in BMMO infected with MHV68 for 6 h. Data represent means of three experiments performed in triplicate. Error bars represent s.d. Statistical significance was determined by *t*-test. **h**, Survival curves of wild-type and $cGas^{-/-}$ mice infected with 100 p.f.u. of WNV, $n = 10$ (wild type), $n = 9$ ($cGas^{-/-}$). **j**, qRT–PCR of baseline gene expression (*Ifnb*, *Ifit1*, *Ifit2*, *Oas1a*, *actin*) in wild-type and $cGas^{-/-}$ BMMO. Data represent means of two experiments performed in triplicate. Error bars represent s.d. Statistical significance was determined by *t*-test. In all panels, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; NS, not significant.

in brains of $cGas^{-/-}$ mice (Extended Data Fig. 7), although extensive time courses and tissue profiling were not performed. However, when we infected wild-type and $cGas^{-/-}$ BMMO with WNV, we detected a modest yet significant fourfold increase in viral titres in $cGas^{-/-}$ cells (Fig. 4i). We assessed the kinetics of WNV-mediated activation of BMMO by monitoring mRNA induction of *Ifnb*, several ISGs (*Ifit1*, *Ifit2*, *Oas1a*), chemokines (*Ccl5* and *Cxcl10*) and cytokines (*Tnfa*, *Il6*, *Il1b*). WNV induced most of these genes to similar levels in both wild-type and $cGas^{-/-}$ cells (data not shown). However, basal mRNA levels of *Ifnb*, all ISGs, and chemokines were significantly reduced in uninfected $cGas^{-/-}$ BMMO (Fig. 4g, j and Extended Data Fig. 8a). Activation of $cGas^{-/-}$ BMMO by agonists of RIG-I-like receptors and RNA-activated Toll-like receptors was also modestly impaired (Extended Data Fig. 8b). Together, these studies implicate a role for cGAS in controlling an RNA virus and in regulating basal immune responses. cGAS may, therefore, set the antiviral tone of the cell. We propose that, in the absence of cGAS, basal mRNA levels of some antiviral genes are reduced, making cells more vulnerable to some RNA viruses. As WNV RNA is predominantly controlled by RIG-I-like-receptor-mediated signalling through IRF3 (refs 22, 23), cGAS may be triggered by endogenous ligands to confer antiviral effects against RNA viruses. Alternatively, cGAS or protein complexes containing cGAS may have a more flexible functionality with respect to nucleic acid triggering than previously anticipated[17,19], such that some viral RNA species can trigger cGAMP production and downstream antiviral responses.

The studies presented here validate the utility of the ISG screening platform to identify critical molecules in innate immunity and lay a foundation for further studies on mechanisms of novel antiviral molecules. Our *in vivo* data indicate that cGAS is pivotal in protecting the host from both DNA and RNA viruses, underscoring an unappreciated role for this key antiviral molecule in the innate immune response.

METHODS SUMMARY

The ISG library and screening platform have been described previously[4]. Before performing ISG screens, dose responses and time courses were performed to determine the amount of virus needed to infect approximately 25–50% of the cells within the first viral life cycle. This dosing allowed the detection of ISGs that inhibit or enhance virus infectivity. Primary screens were analysed by quantifying the infectivity (GFP positivity) of GFP-expressing viruses in ISG-expressing cells. The data were normalized to the average of all data points in the screen. Selected ISGs were chosen for confirmatory assays that were performed using independent lentiviral stocks. For hierarchical clustering analysis, viruses and ISGs were grouped using R or MATLAB statistical software. For gene expression analyses, cells were transduced with lentiviral stocks and total RNA was analysed for gene induction by qRT–PCR, or by microarray using Illumina BeadArray technology. To characterize cGAS signalling in siRNA knockdown cells, or in cells transduced with lentiviruses, cellular protein lysates were analysed by western blot with antibodies to detect cGAS, STING, phosphorylated IRF3 or actin. To study the role of cGAS *in vivo*, we obtained mice with a gene-trap cassette at the *cGas* locus. Gene-trapped mice were bred to FlpE-expressing mice to remove the targeting cassette, followed by crossing to Cre-expressing mice to generate $cGas^{-/-}$ mice. Knockout mice were infected in parallel with congenic C57BL/6 (B6) control mice with MHV68, VV or WNV. Titres of MHV68 and WNV from organs of infected mice were determined on NIH-3T12 and Vero cells, respectively. Mice infected with VV and WNV were monitored for weight loss and/or lethality. BMMO from wild-type and $cGas^{-/-}$ mice were infected with MHV68, VV or WNV. Viral titres were determined by plaque assay, or mRNA induction was assessed by qRT–PCR.

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 21 June; accepted 7 November 2013.
Published online 27 November 2013.

1. Der, S. D., Zhou, A., Williams, B. R. & Silverman, R. H. Identification of genes differentially regulated by interferon α, β, or γ using oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **95,** 15623–15628 (1998).
2. de Veer, M. J. *et al.* Functional classification of interferon-stimulated genes identified using microarrays. *J. Leukoc. Biol.* **69,** 912–920 (2001).
3. Schoggins, J. W. & Rice, C. M. Interferon-stimulated genes and their antiviral effector functions. *Curr. Opin. Virol.* **1,** 519–525 (2011).
4. Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472,** 481–485 (2011).
5. Liu, S. Y., Sanchez, D. J., Aliyari, R., Lu, S. & Cheng, G. Systematic identification of type I and type II interferon-induced antiviral factors. *Proc. Natl Acad. Sci. USA* **109,** 4239–4244 (2012).
6. Metz, P. *et al.* Identification of type I and type II interferon-induced effectors controlling hepatitis C virus replication. *Hepatology* **56,** 2082–2093 (2012).
7. Fusco, D. N. *et al.* A genetic screen identifies interferon-α effector genes required to suppress hepatitis C virus replication. *Gastroenterology* **144,** 1438–1449 (2013).
8. Zhao, H. *et al.* A functional genomic screen reveals novel host genes that mediate interferon-alpha's effects against hepatitis C virus. *J. Hepatol.* **56,** 326–333 (2012).
9. Zhang, Y., Burke, C. W., Ryman, K. D. & Klimstra, W. B. Identification and characterization of interferon-induced proteins that inhibit alphavirus replication. *J. Virol.* **81,** 11246–11255 (2007).
10. Wilson, S. J. *et al.* Inhibition of HIV-1 particle assembly by 2′,3′-cyclic-nucleotide 3′-phosphodiesterase. *Cell Host Microbe* **12,** 585–597 (2012).
11. Schoggins, J. W. *et al.* Dengue reporter viruses reveal viral dynamics in interferon receptor-deficient mice and sensitivity to interferon effectors *in vitro*. *Proc. Natl Acad. Sci. USA* **109,** 14610–14615 (2012).

12. Karki, S. *et al.* Multiple interferon stimulated genes synergize with the zinc finger antiviral protein to mediate anti-alphavirus activity. *PLoS ONE* **7,** e37398 (2012).
13. Meng, X. *et al.* C7L family of poxvirus host range genes inhibits antiviral activities induced by type I interferons and interferon regulatory factor 1. *J. Virol.* **86,** 4538–4547 (2012).
14. Dupuis, S. *et al.* Impaired response to interferon-α/β and lethal viral disease in human STAT1 deficiency. *Nature Genet.* **33,** 388–391 (2003).
15. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
16. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339,** 826–830 (2013).
17. Kranzusch, P. J., Lee, A. S.-Y., Berger, J. M. & Doudna, J. A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.* **3,** 1362–1368 (2013).
18. Gao, P. *et al.* Cyclic [G(2′,5′)pA(3′,5′)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153,** 1094–1107 (2013).
19. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498,** 332–337 (2013).
20. Gao, D. *et al.* Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341,** 903–906 (2013).
21. Li, X. D. *et al.* Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* **341,** 1390–1394 (2013).
22. Suthar, M. S. *et al.* IPS-1 is essential for the control of West Nile virus infection and immunity. *PLoS Pathog.* **6,** e1000757 (2010).
23. Errett, J. S., Suthar, M. S., McMillan, A., Diamond, M. S. & Gale, M. Jr. The essential, non-redundant roles of RIG-I and MDA5 in detecting and controlling West Nile virus infection. *J. Virol.* **87,** 11416–11425 (2013).

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

Acknowledgements We thank the following investigators for contributing viral molecular clones or viral stocks: R. Cattaneo (MV), P. Collins (PIV3, HMPV, RSV), I Frolov (VEEV), M. Heise (SINV), S. Higgs (ONNV), P. Traktman (VV), J. L. Whitton (CVB). We thank E. Jouanguy and J.-L. Casanova for *STAT1*$^{-/-}$ fibroblasts. We acknowledge the support of C. Zhao, X. Wang and W. Zhang at The Rockefeller University Genomics Resource Center. We thank E. Castillo, B. Flatley, A. Webson, E. Duan and A. McLees for laboratory support, and D. Kraemalmeyer for managing mouse colonies. This work was supported in part by National Institutes of Health grants AI091707 to C.M.R., AI057158 to I. Lipkin (Northeast Biodefense Center, subcontract to C.M.R.), DK095031 to J.W.S., AI057160 to H.W.V., HHSN272200900041CU19 to M.S.D. and H.W.V., AI104972 to M.S.D., AI083025, AI095611 and CEIRS contract HHSN266200700010C to A.G-S., GM076547 and GM103511 to J.D.A. and AI057160 to W.M.Y., an investigator of the Howard Hughes Medical Institute. M.D.G. is supported by T32 AR007279. Additional funding was provided by the Greenberg Medical Research Institute, the Starr Foundation and the Ronald A. Shellow, M.D. Memorial Fund (C.M.R.).

Author Contributions J.W.S., D.A.M., M.S.D., H.W.V. and C.M.R. designed the project. J.W.S., D.A.M., N.I., M.D.G., B.S., R.D., J.L.E., K.B.M. and R.B.R. performed the experimental work. J.W.S., V.L. and A.V.R. performed clustering analyses. J.W.S., D.A.M., M.S.D., H.W.V. and C.M.R. analysed the results and wrote the manuscript. B.M., A.A., J.D.A., R.M.E., A.G.-S., V.R., M.D.G., W.M.Y., E.J.S. and M.S.D. contributed reagents and technical expertise.

Author Information Microarray Data were submitted to NCBI Gene Expression Omnibus (GEO) with accession number GSE52241. 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 C.M.R. (ricec@rockefeller.edu) or J.W.S. (john.schoggins@utsouthwestern.edu).

METHODS

Viruses and cells. Huh7, HeLa and 293T cells were maintained in DMEM (Invitrogen) with 10% FCS and 0.1 mM non-essential amino acids. NIH-3T12 cells were grown in DMEM supplemented with 5% FCS, 100 U penicillin per ml, 100 μg streptomycin per ml and 2 mM L-glutamine. *STAT1*$^{-/-}$ fibroblasts (an SV40 large T antigen immortalized human skin fibroblast line) were grown in RPMI (Invitrogen) with 10% FCS. The construction, characterization and generation of viral stocks for the following viruses have been previously described: CVB-GFP (derived from infectious clone pMKS1-GFP)[24], PV-GFP (strain P1M, derived from infectious clone pPVM-2A144-GFP)[25], EAV-GFP (derived from infectious clone pEAV211-GFP2aT)[26], SINV-A-GFP and SINV-G-GFP (derived from infectious clones pS300-GFP and pG100-GFP)[27], ONNV-GFP (derived from infectious clone pONNV.GFP)[28], VEEV-GFP (derived from pTC83-GFP infectious clone)[4], FLUAV-GFP (based on strain PR8)[29], PIV3-GFP (based on strain JS)[30], NDV-GFP (based on strain Hitchner B1)[31], HMPV-GFP[32] (based on isolate CAN97-83), RSV-GFP (based on strain A2)[32], MV-GFP (MVvac2-GFP, based on vaccine strain, Edmonston lineage measles virus)[33] and BUNV-GFP[34] (based on rBUN-del7GFP). VV-GFP was propagated in BSC-40 cells. Viral stocks were prepared by three freeze–thaw cycles, followed by centrifugation at 1,000*g* to remove cellular debris. VV Western Reserve was obtained from the ATCC, propagated in Vero cells and purified by ultracentrifugation through a 36% sucrose cushion. MHV68 clone WUMS was obtained from the ATCC and propagated in NIH-3T12 cells. The WNV strain was isolated and passaged as described previously[35].

Plasmids and molecular cloning. The production of the lentiviral-based ISG expression library has been described in detail[4]. To characterize human cGAS (MB21D1/C6orf150), we used the pENTR.C6orf150 (Genecopoeia, NCBI accession AK097148) plasmid as a starting point for all modifications. Standard PCR was used to generate mutants of cGAS that were progressively deleted of amino acids from the N or C termini. Overlap extension PCR was used to generate point mutants (E225A, D227A) in the wild-type protein. All mutant cGAS sequences were moved into the lentivirus by Gateway cloning, using pENTR.C6orf150 plasmids and pTRIP.CMV.IVSB.IRES.TagRFP-DEST in an LR reaction (Invitrogen) as previously described[4]. Primer sequences for mutagenesis are available upon request.

Lentivirus production and transduction assays. Lentiviral stocks were generated in 293T cells by co-transfection of with plasmids expressing (1) the TRIP.CMV.IVSb.ISG.ires.TagRFP lentivirus; (2) HIV gag-pol; and (3) the vesicular stomatitis virus glycoprotein (VSV-G) in a ratio of 1:0.8:0.2. For puromycin-selectable lentiviruses, we used the SCRPSY lentiviral backbone, which has been described previously[11]. Supernatants were collected at 48 h and 72 h, pooled, cleared by centrifugation at 1,000*g* and stored at −80 °C. For transduction assays, Huh7, HeLa or *STAT1*$^{-/-}$ fibroblasts were seeded into 24-well plates at a density of 7×10^4 cells per well and transduced with lentiviral pseudoparticles by spinoculation at 1,000*g* for 45 min at 37 °C in medium containing 3% FBS, 20 mM HEPES and 4 μg ml^{-1} polybrene. For confirmatory experiments (Fig. 2), new lentiviral stocks were generated for selected ISGs that had *Z*-scores less than −1.5 or greater than 2.0 in the initial screens. The less stringent cutoff of $Z < -1.5$ was chosen to include more ISGs from –ssRNA screens. Confirmatory experiments were performed under the same infection conditions as described above. The data from the confirmatory assays was stratified according to the frequency and relative magnitude with which ISGs affected RNA virus infectivity, using progressive 50% cutoffs to delineate strong versus modest effectors.

Virus infections. Before ISG screens, all GFP reporter viruses were optimized for infection in their respective target cells. Dose response and time course assays were carried out to determine the optimal volume of virus needed to infect 25–50% (approximately 0.5 m.o.i.) of the cell population during the first round of replication, before onset of viral spread. ISG screens and confirmatory assays were carried out under these optimized conditions, and cells were infected with each virus for the following time periods: VV-GFP (8 h), CVB-GFP (6 h), PV-GFP (8 h), EAV-GFP (19 h), SINV-GFP (10 h), VEEV-GFP (6 h), ONNV-GFP (17 h), FLUAV-GFP (8 h), PIV3-GFP (24 h), NDV-GFP (8 h), HMPV-GFP (18 h), RSV-GFP (23 h), MV (24 h), BUNV (11 h). For FLUAV and HMPV, trypsin was not added to the infected cells, thereby preventing release of virions from the cell surface and blocking viral spread.

Bioinformatics (clustering, co-occurrence and Gene Ontology). To cluster ISGs with respect to the viruses they inhibit, 22 ISGs that inhibited at least one virus by more than 50% in confirmatory assays were selected. Replication data from the primary screens for each of these 22 genes was compiled. A web-based tool, http://www.hiv.lanl.gov/content/sequence/HEATMAP/heatmap.html, which uses *heatmap.2* of the *gplots* package of the R statistical computing and graphics software environment, was used to generate a heatmap based on the similarity of ISG effects on virus infectivity. In brief, a data set in which ISGs were set as columns and viruses set as rows, with replication values corresponding to each position, was uploaded to the server. The algorithm used the 'average' cluster method and 'Euclidean' distance method. The heatmap was created with nine colours representing the quantitative range of virus infectivity, and dendrograms were generated to show the hierarchical relationships of both ISGs and viruses.

In a second clustering analysis, selected viruses were clustered hierarchically according to the appearance of the top 30 ISG names in the lists generated from the primary screening data (see Supplementary Tables 1 and 2). Each gene list was transformed to an *n*-dimensional binarized vector, where *n* represents the number of unique gene names in the cumulative gene list for all selected viruses ($n = 176$ for the selected 12 viruses). Thus, each position of this vector corresponds to a unique gene from this cumulative list. A value of 1 or 0 at each position of this vector indicates the presence or absence, respectively, of the ISG in the current list. The binarized vectors for selected viruses were clustered using the *linkage* function of the MATLAB Statistics Toolbox with the weighted average distance (WPGMA) ('weighted') selected as the algorithm for computing distance between clusters and one minus the sample correlation between points ('correlation') as the distance metric. The dendrogram was constructed using *dendrogram* function of the MATLAB Statistics Toolbox. The dendrogram represents U-shaped lines connecting viruses in the hierarchical tree according to the clustering of binarized vectors (see above) of selected viruses. The height of each U represents the distance between the two viruses being connected.

The top 20 inhibitory ISGs from 12 screens were analysed for co-occurrence. A gene appearing in all lists was assigned a frequency of 1, or a fractional percentage of 1 if the gene appeared in fewer lists. The data were stratified by +ssRNA and −ssRNA viruses.

Gene ontology (GO) analysis was performed on the top 30 inhibitory genes from 12 screens, followed by statistical enrichment analysis using the Enrichment widget of STRING, as described[36]. In this analysis, GO terms associated with a known pathway were assigned to each protein in the list. The *P* values were determined by a hypergeometric test, corrected using the Benjamini–Hochberg false discovery rate procedure, and ranked based on enrichment value. We also included 'electronic inferred annotations' in the *P* values calculation to increase statistical significance. GO terms with $P < 0.05$ were compiled in Supplementary Table 3. Full GO analyses for each virus are available upon request.

siRNA-mediated gene silencing. STING was depleted in *STAT1*$^{-/-}$ fibroblasts by siRNA-mediated gene silencing. Four individual siRNAs (Qiagen) at 20 nM were tested for knockdown using HiPerFect Transfection Reagent (Qiagen) according to the manufacturer's reverse transfection protocol. Cells were collected 48 h after transfection and STING protein levels were monitored by western blot, as described below. One of the four siRNAs reduced STING expression to nearly undetectable protein levels, and was chosen for subsequent functional assays.

RNA and protein detection from cell cultures. For gene expression studies, total RNA from *STAT1*$^{-/-}$ fibroblasts transduced with lentiviral vectors was isolated 48 h after transduction using an RNeasy Mini Kit (Qiagen). 50 ng total RNA was analysed by qRT–PCR using QuantiFast SYBR Green RT–PCR kit with commercially available QuantiTect Primers specific for *OAS2* and *RPS11* housekeeping control (Qiagen) according to the manufacturer's instructions. Reactions were run on a Roche 480 Light Cycler or ABI 7500 Fast Real Time PCR System, and gene expression was calculated using the $\Delta\Delta C_T$ method. In separate experiments, total RNA was processed for microarray analysis using BeadArray technology (Illumina) as described previously[4]. For protein expression studies, cells were lysed in radioimmunoprecipitation assay buffer containing Complete Protease Inhibitor Cocktail (Roche) and PhosSTOP Phosphatase Inhibitor Cocktail (Roche). Protein concentration of cell lysates was determined by Bradford assay (Pierce). Lysates were separated on 4–20% SDS–PAGE gradient gels (BioRad), blotted to nitrocellulose membrane (Amersham) and processed by western blotting. Blots were blocked overnight in 5% milk in 1 × TBS (50 mM Tris-Cl, pH 7.5, 150 mM NaCl.) with 0.05% Tween-20 (TBS-T), followed by incubation with primary and secondary antibodies in 1% milk in TBS-T for 1 h and 30 min, respectively. Proteins were visualized by incubating blots with enhanced chemiluminescent substrate ECL (Pierce) and exposing blots to autoradiography film (Denville Scientific). Antibodies used in the study include: anti-STING (R&D Systems MAB7169), anti-phosho-IRF3 (Abcam ab76493), anti-MB21D1 (Sigma HPA031700), anti-actin (Abcam ab6276), goat anti-rabbit horseradish peroxidase (HRP) and goat anti-mouse HRP (Pierce).

ISRE reporter assays. To test ISGs for activation of ISRE-dependent transcription, 1.5×10^4 293T cells in individual wells of a 96-well plate were transduced with lentiviruses expressing ISGs. 16–18 h later, transduced cells were transfected with the pISG54.ISRE-Fluc plasmid using X-treme Gene 9 Transfection reagent (Roche) following standard protocols. Cells were collected 24 h after transfection using 1 × cell culture lysis buffer (Promega) and Fluc activity was monitored using the Luciferase Assay System (Promega). To test for ISG-mediated suppression of IFN activity, a similar protocol was used, except that ISG and ISRE-Fluc plasmids

were co-transfected before treatment with recombinant IFN-α (PBL Interferon Source).

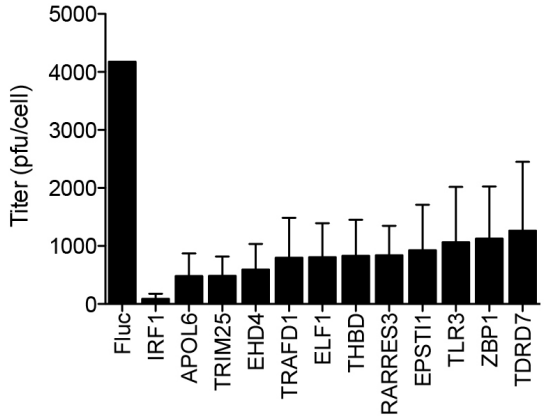
Infection of mice. Mb21d1^{tm1a} mice were obtained from the EUCOMM consortium and bred to FLPe-expressing mice (B6-Tg(CAG-FLPe)36, provided by the RIKEN BRC through the National Bio-Resource Project of the MEXT, Japan[37]) to generate conditional knockout mice by removing the targeting cassette. Conditional knockout mice were bred to Cre-expressing mice (B6.C-Tg(CMV-cre)1Cgn/J, Jackson Laboratories) to generate mice with a deletion of exon 2, which includes the catalytic residues E211 and D213. Mice were backcrossed to C57BL/6J (B6) mice (Jackson Laboratories) to remove Cre allele. All mice were bred and maintained in a specific-pathogen-free barrier facility at Washington University in St Louis, Missouri, in accordance with federal and institutional guidelines. To determine MHV68 titres in tissues, mice were infected between 8 and 9 weeks of age with 10^6 p.f.u. of MHV68 by intraperitoneal injection in 0.3 ml PBS. Upon euthanization, organs were placed in 1 ml of complete DMEM and frozen at −80 °C. To determine susceptibility to VV Western Reserve, 8–9-week-old mice were anesthetized with ketamine/xylazine before intranasal inoculation with 8,000 p.f.u. in 50 μl MEM. Mice were also infected with 100 p.f.u. of WNV (strain New York 1999) in 50 μl via a subcutaneous route. Tissues were collected at day 8 after infection and analysed for viral burden by plaque assay on Vero cells. For these studies, 10 or 15 mice were used per experiment in order to achieve reliable statistics. All mice were monitored daily for weight loss and lethality, and mice that became moribund were euthanized.

Bone-marrow-derived macrophage infections. Primary bone-marrow-derived macrophages were prepared as described previously[38]. Cells were allowed to differentiate for 7 days, and then adherent cells were scraped and seeded in tissue culture-treated plates. For MHV68 experiments, cells were infected with MHV68 at a m.o.i. of 10 for 1 h with occasional rocking at 37 °C and 5% CO_2. For viral growth curves, cells were washed three times with medium and incubated in DMEM supplemented with 10% FBS and 2 mM L-glutamine for the indicated period of time at 37 °C and 5% CO_2, before being frozen at −80°C. For gene expression analysis, the inoculum was replaced with complete DMEM and the cells incubated for 6 h before being lysed in TRIzol reagent for total RNA extraction. For VV experiments, macrophages were infected at a m.o.i. of 0.1 in DMEM without serum. One hour later, cells were washed once with PBS, and incubated in DMEM supplemented with 2% FBS. At indicted times post infection, cells and media were frozen/thawed twice and then supernatants were serially diluted and plaqued on monolayers of Vero cells. For WNV experiments, cells were infected in 12-well plates at a m.o.i. of 0.1 or 3. Virus was collected from supernatants at specific times and titrated by plaque assay on Vero cells. For activation experiments, cells were stimulated with 2 μg ml^{-1} imiquimod (Invivogen), 2 μg ml^{-1} polyI:C (Invivogen) or transfected with 1 μg polyI:C using TransIT-LT1 (Mirus) for the indicated amount of time. Total RNA was collected from BMMO to assess gene expression levels by qRT–PCR, as described below.

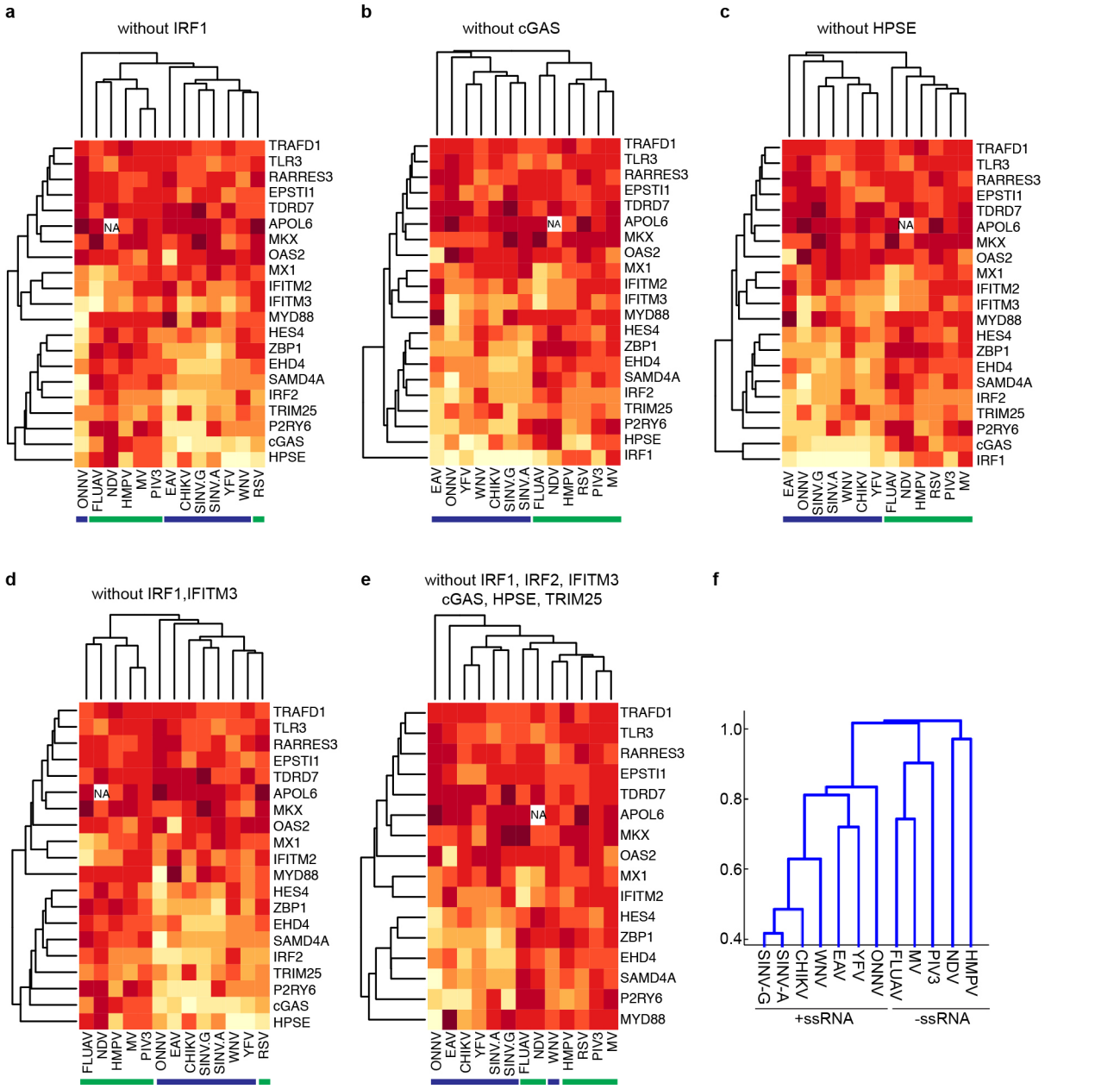
Plaque assays. To determine the effects of ISGs on poliovirus production, HeLa cells were transfected in a 24-well plated with 500 ng lentiviral plasmids encoding ISGs using FugeneHD (Roche). Forty-eight hours after transfection, cells were infected with P1M (10 m.o.i.) for 16 h. Lysates were collected and viral titres determined by plaque assay on HeLa cell monolayers. Infections were performed at 37 °C for 1 h with occasional rocking before cells were overlaid with medium containing DMEM, 0.2% $NaHCO_3$, 5% bovine calf serum, 1% penicillin-streptomycin and 0.8% Noble agar (Sigma). After 48 h incubation, plaques were visualized with crystal violet. Plaque assays to determine MHV68 and WNV viral titres from BMMO or mouse organs were performed on NIH-3T12 and Vero cells, respectively. Organs were thawed and homogenized with sterile 1.0 mm zirconia/silica beads and a mini-beadbeater (BioSpec Products) before dilution and plating onto cells. Infection was performed at 37 °C for 1 h with occasional rocking before cells were overlaid with medium containing 2% methylcellulose. After a 1-week incubation, plaques were visualized with 3% neutral red solution.

Determination of mRNA transcript levels in mouse cells and organs. Spleens were homogenized and macrophages were lysed in TRIzol reagent (Invitrogen) and processed according to the manufacturer's instructions to isolate total RNA. RNA samples were treated with DNase I (Ambion) before first-strand cDNA synthesis with ImProm-II (Promega) and oligo(dT)$_{15}$. qPCR was performed on a StepOnePlus machine using Power SYBR Green master mix (Applied Biosystems) and primers specific for *cGas* (5′-ACGAGAGCCGTTTTATCTCGTACCC-3′ and 5′-TGTCCGGAAGATTCACAGCATGTTT-3′) or ribosomal protein S29 (*RPS29*; 5′-AGCAGCTCTACTGGAGTCACC-3′ and 5′-AGGTCGCTTAGTCCAACTTAATG-3′). TaqMan quantitative PCR was performed using primers and probes specific for *Ifnb* (5′-CTGGAGCAGCTGAATGGAAAG-3′, 5′-CTTCTCCGTCATCTCCATAGGG-3′ and probe 5′-/56-FAM/CAACCTCACCTACAGGGCGGACTTCAAG/36-TAMSp/-3′[39], *Ifit1* (5′-GAGCCAGAAAACCCTGAGTACA-3′, 5′-AGAAATAAAGTTGTCATCTAAATC-3′ and probe 5′-/56-FAM/ACTGGCTATGCAGTCGTAGCCTATCGCC/36-TAMSp/-3′), *Ifit2* (5′-CTGAAGCTTGACGCGGTACA-3′, 5′-ACTTGGGTCTTTCTTTAAGGCTTCT-3′ and probe 5′-/56-FAM/AAAACCAAGCAATGGCGCTGGTTG/36-TAMSp/-3′), *Oas1a* (5′-TGAGCGCCCCCCATCT-3′, 5′-CATGACCCAGGACATCAAAGG-3′ and probe 5′-/56-FAM/AGGAGGTGGAGTTTGATGTGCTG/36-TAMSp/-3′), *Il6* (5′-GCCAGAGTCCTTCAGAGAGATACA-3′ and 5′-CTTGGTCCTTAGCCACTCCTTC-3′), *Tnfa* (5′-GGGTGATCGGTCCCCAAAGG-3′ and 5′-CTCCACTTGGTGGTTTGCTACGA-3′), *Il1b* (5′-GCACACCCACCCTGCAG-3′ and 5′-AACCGTTTTTCCATCTTCTTCTT-3′), *Ccl5* (5′-CAAGTGCTCCAATCTTGCAGTC-3′ and 5′-TTCTCTGGGTTGGCACACAC-3′), *Cxcl10* (5′-AGTGCTGCCGTCATTTTCTG-3′, 5′-ATTCTCACTGGCCCGTCA T-3′ and probe 5′-/56-FAM/AGTCCCACTCAGACCCAGCAGG/36-TAMSp/-3′) with AmpliTaq Gold (Applied Biosystems). Transcript levels were analysed using the $\Delta\Delta C_t$ method, with *RPS29* as the reference gene. qPCR products were confirmed by melt curve and/or agarose gel electrophoresis.

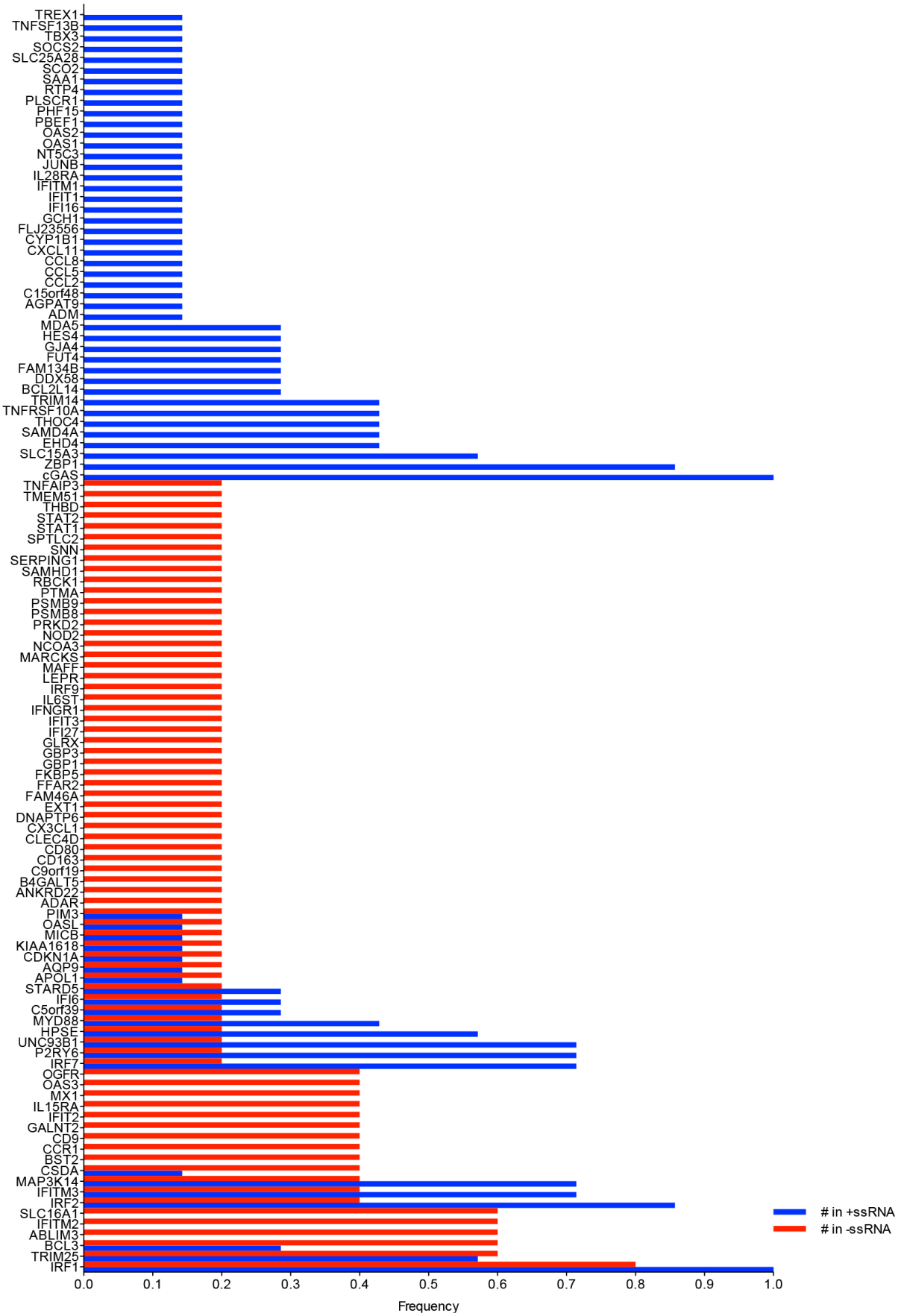
24. Feuer, R., Mena, I., Pagarigan, R., Slifka, M. K. & Whitton, J. L. Cell cycle status affects coxsackievirus replication, persistence, and reactivation in vitro. *J. Virol.* **76,** 4430–4440 (2002).
25. Teterina, N. L., Levenson, E. A. & Ehrenfeld, E. Viable polioviruses that encode 2A proteins with fluorescent protein tags. *J. Virol.* **84,** 1477–1488 (2010).
26. van den Born, E., Posthuma, C. C., Knoops, K. & Snijder, E. J. An infectious recombinant equine arteritis virus expressing green fluorescent protein from its replicase gene. *J. Gen. Virol.* **88,** 1196–1205 (2007).
27. Simmons, J. D., Wollish, A. C. & Heise, M. T. A determinant of Sindbis virus neurovirulence enables efficient disruption of Jak/STAT signaling. *J. Virol.* **84,** 11429–11439 (2010).
28. Brault, A. C. *et al.* Infection patterns of o'nyong nyong virus in the malaria-transmitting mosquito, *Anopheles gambiae*. *Insect Mol. Biol.* **13,** 625–635 (2004).
29. Manicassamy, B. *et al.* Analysis of *in vivo* dynamics of influenza virus infection in mice using a GFP reporter virus. *Proc. Natl Acad. Sci. USA* **107,** 11531–11536 (2010).
30. Zhang, L. *et al.* Infection of ciliated cells by human parainfluenza virus type 3 in an *in vitro* model of human airway epithelium. *J. Virol.* **79,** 1113–1124 (2005).
31. Park, M. S. *et al.* Newcastle disease virus (NDV)-based assay demonstrates interferon-antagonist activity for the NDV V protein and the Nipah virus V, W, and C proteins. *J. Virol.* **77,** 1501–1511 (2003).
32. Biacchesi, S. *et al.* Recovery of human metapneumovirus from cDNA: optimization of growth in vitro and expression of additional genes. *Virology* **321,** 247–259 (2004).
33. del Valle, J. R. *et al.* A vectored measles virus induces hepatitis B surface antigen antibodies while protecting macaques against measles virus challenge. *J. Virol.* **81,** 10597–10605 (2007).
34. Shi, X., van Mierlo, J. T., French, A. & Elliott, R. M. Visualizing the replication cycle of bunyamwera orthobunyavirus expressing fluorescent protein-tagged Gc glycoprotein. *J. Virol.* **84,** 8460–8469 (2010).
35. Ebel, G. D., Carricaburu, J., Young, D., Bernard, K. A. & Kramer, L. D. Genetic and phenotypic variation of West Nile virus in New York, 2000–2003. *Am. J. Trop. Med. Hyg.* **71,** 493–500 (2004).
36. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41,** D808–D815 (2013).
37. Kanki, H., Suzuki, H. & Itohara, S. High-efficiency *CAG-FLPe* deleter mice in C57BL/6J background. *Exp. Anim.* **55,** 137–141 (2006).
38. Hwang, S. *et al.* Nondegradative role of Atg5-Atg12/Atg16L1 autophagy protein complex in antiviral activity of interferon gamma. *Cell Host Microbe* **11,** 397–409 (2012).
39. Daffis, S., Samuel, M. A., Keller, B. C., Gale, M., Jr. & Diamond, M. S. Cell-specific IRF-3 responses protect against West Nile virus infection by interferon-dependent and -independent mechanisms. *PLoS Pathog.* **3,** e106 (2007).



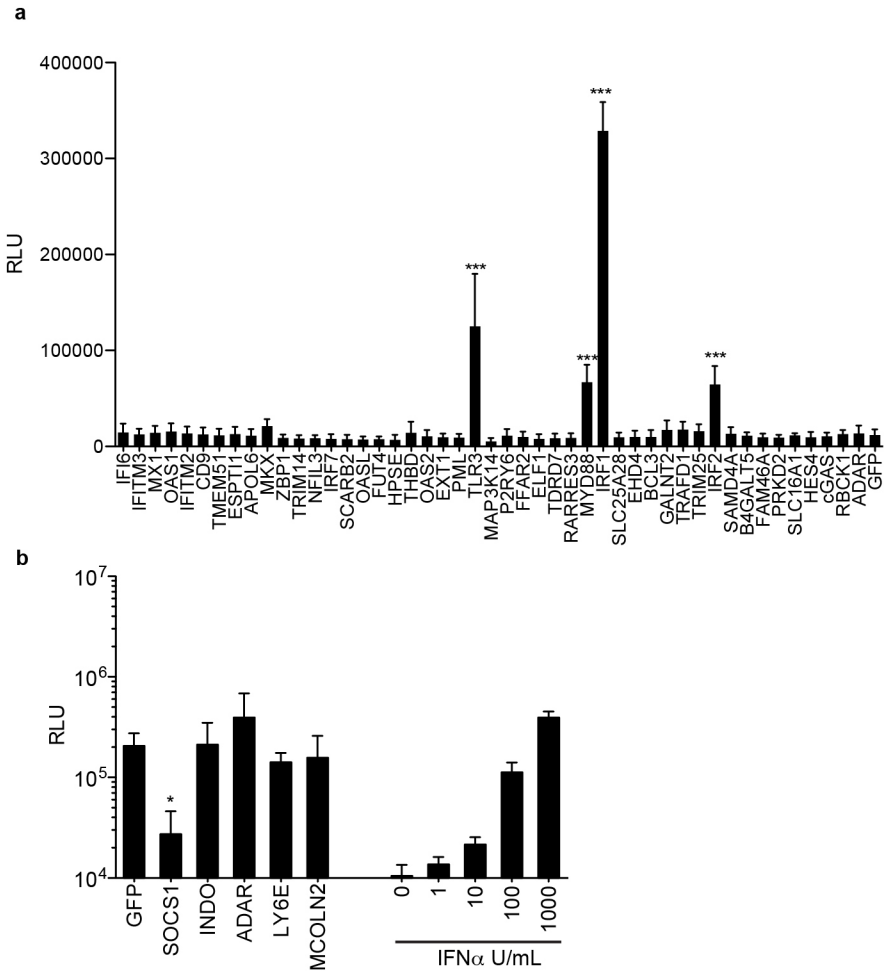
Extended Data Figure 1 | Antiviral effects of ISGs on virus production of a non-GFP poliovirus. HeLa cells were transfected with plasmids encoding ISGs and 48 h later infected with P1M (10 m.o.i.) for 16 h. Lysates were collected and viral titres determined by plaque assay on HeLa cell monolayers, as described in Methods. Plaque assays were performed in duplicate. Data represent the average of three independent experiments. Error bars represent s.d.



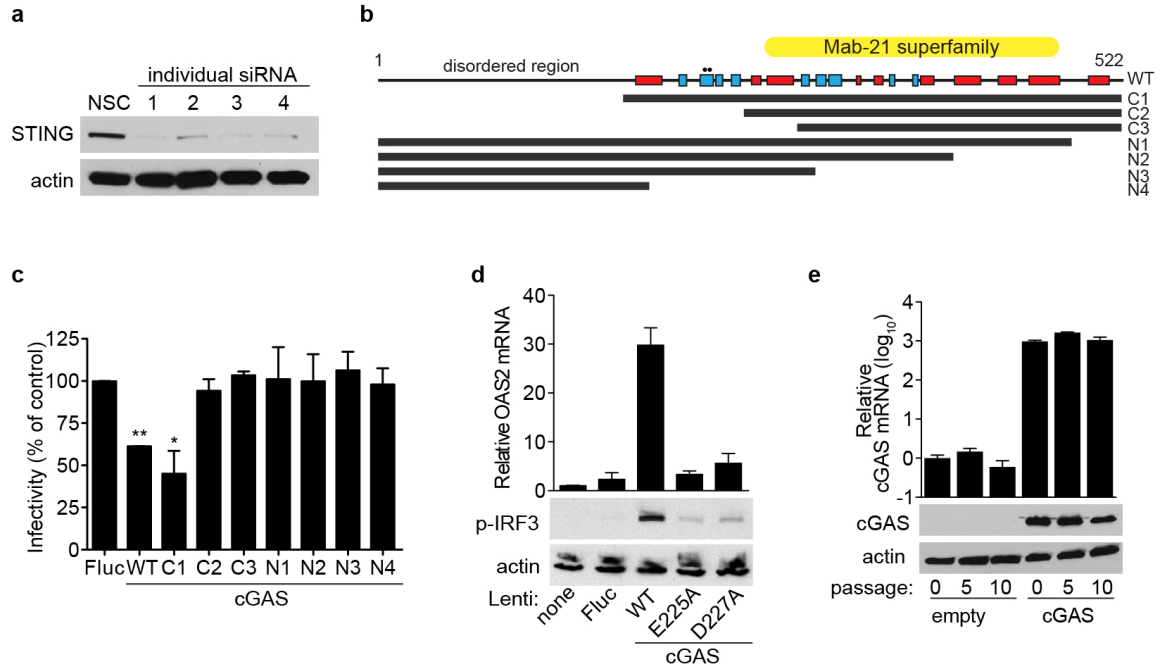
Extended Data Figure 2 | Hierarchical clustering. **a**–**e**, Analyses were performed as described in Methods. In each cluster, one or more ISGs were removed from the analysis in Fig. 3a to determine whether virus clustering is driven by a subset of one or more dominant genes. Blue and green bars underscore +ssRNA and −ssRNA viruses, respectively. **f**, The top 30 ISG inhibitors from the primary screens were compiled as a gene list and transformed to a binarized vector for clustering using MATLAB Statistics Toolbox (see Methods). A dendrogram was generated from the clustering analysis to show how viruses relate to each other with respect to the ISGs that target them.



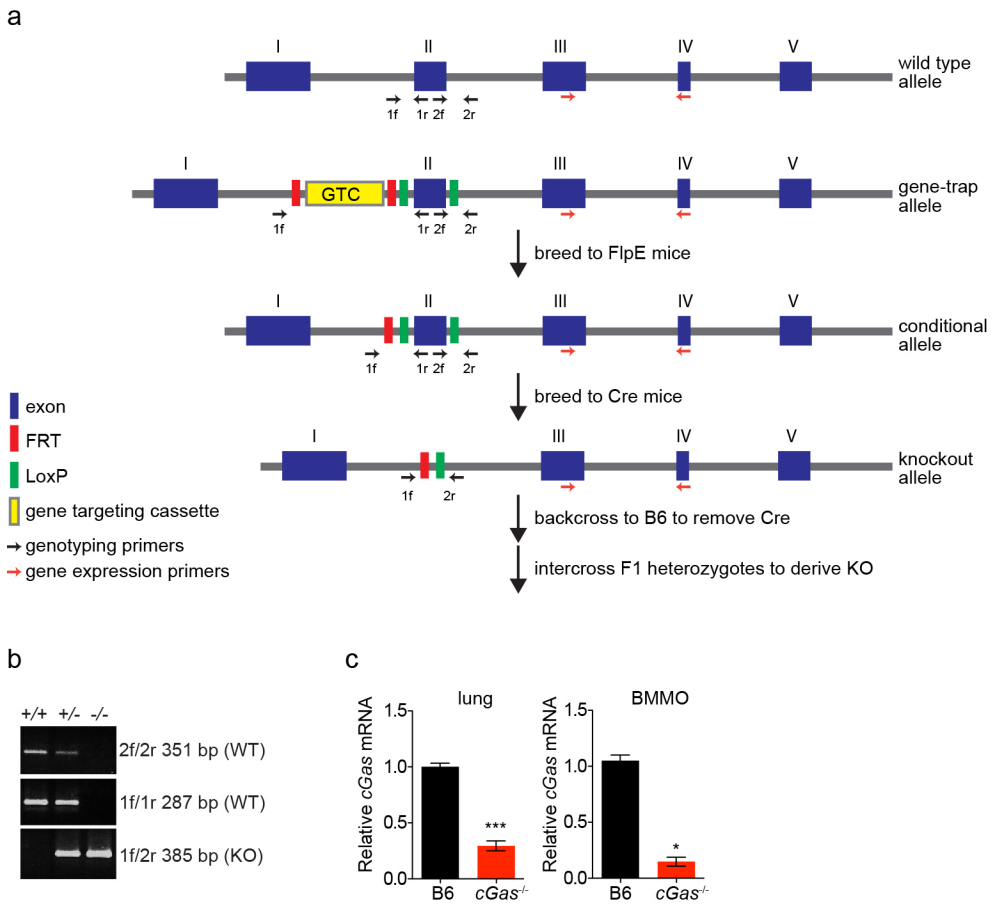
Extended Data Figure 3 | Co-occurrence of top 20 antiviral ISGs from primary screens. ISGs were assigned a frequency on the basis of the number of times the gene appeared in a list of the 20 most inhibitory genes from 7 +ssRNA or 5 −ssRNA virus screens. A frequency of 1 indicates an occurrence of 100% across all gene lists. Co-occurrence is reflected by adjacent red and blue bars.



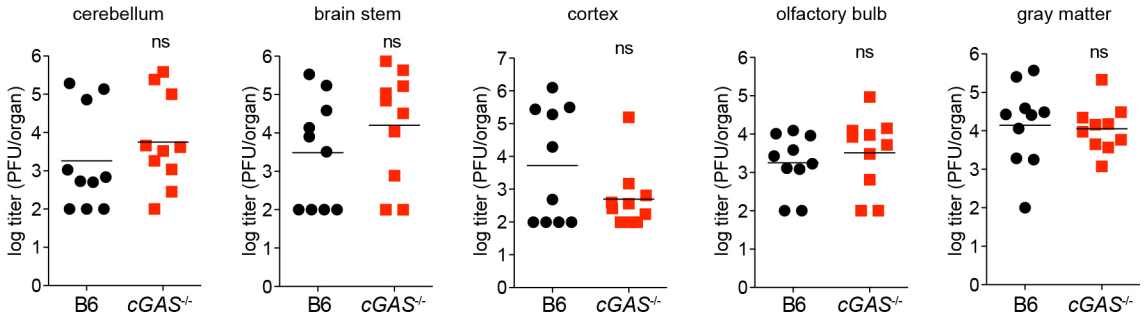
Extended Data Figure 4 | Effects of ISGs on ISRE-dependent transcription. **a**, 293 cells were transduced with lentiviruses expressing ISGs, followed by transfection with an ISRE reporter plasmid expressing Fluc. Cells were assayed for Fluc activity 24 h after transfection. **b**, 293 cells were co-transfected with ISG-expressing plasmids and an ISRE reporter plasmid. The cells were then treated overnight with 1,000 U ml^{-1} interferon-α (IFN-α), followed by Fluc activity assay. Data represent the average of three independent experiments performed in triplicate. Error bars represent s.d. Statistical significance was determined by one-way ANOVA or *t*-test. *$P < 0.05$, ***$P < 0.001$.



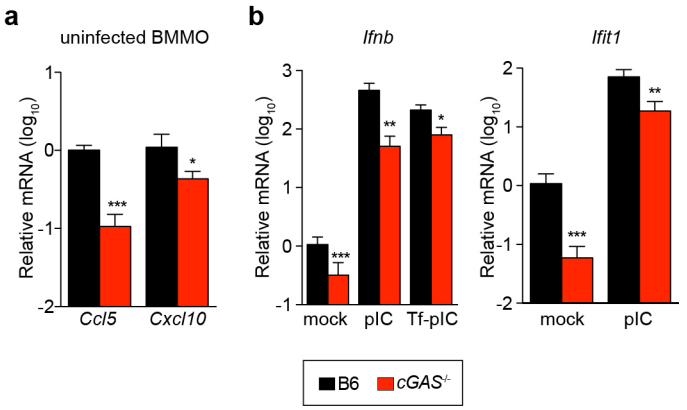
Extended Data Figure 5 | cGAS mechanistic studies. **a**, *STAT1*$^{-/-}$ fibroblasts were transfected with individual siRNAs targeting *STING*. Cell lysates were processed 48 h after transfection for western blot with anti-STING- or anti-actin-specific antibodies. From these results, siRNA no. 1 was chosen for additional studies. **b**, Schematic of cGAS protein sequence and truncation mutants. Red box, α-helix; blue box, β-sheet. Circles denote catalytic residues E225A and D227A. **c**, *STAT1*$^{-/-}$ fibroblasts were transduced with lentivirus expressing control or cGAS (wild type and truncation mutants). Cells were infected 48 h after transduction with VEEV-GFP and infectivity was monitored by FACS. Data represent the mean of two independent experiments. Error bars represent s.d. Statistical significance was determined by one-way ANOVA. *$P < 0.05$, **$P < 0.01$. **d**, *STAT1*$^{-/-}$ fibroblasts were transduced with lentivirus expressing control or cGAS (wild type and mutants). Cells were collected 48 h after transduction and total RNA was analysed for *OAS2* mRNA induction relative to *RPS11* (top), or protein lysates were analysed for phospho-IRF3 and actin expression by western blot (bottom). **e**, *STAT1*$^{-/-}$ fibroblasts were transduced with a puromycin-selectable lentivirus expressing cGAS and placed under selection. At various passages, cells were collected and total RNA was analysed for *OAS2* mRNA induction relative to *RPS11* (top), or protein lysates were analysed for cGAS and actin expression by western blot (bottom). Western blot and *cGAS* mRNA data represent one of two independent experiments, each showing similar results. *OAS2* mRNA data are presented as the average of two independent experiments, each performed in triplicate. Error bars represent s.d.



Extended Data Figure 6 | Gene-targeting strategy to create *cGas* knockout mice. **a**, Mice expressing a *cGas* exon 2 gene-trap cassette were crossed to FlpE-expressing mice to generate conditional knockouts. These mice were then crossed to Cre-expressing mice to generate the knockout allele with a deletion of exon 2, which contains the cGAS catalytic sites. Mice were backcrossed to remove Cre, and $cGas^{+/-}$ mice were intercrossed to derive $cGas^{-/-}$ mice. **b**, PCR products from genomic DNA of $cGas^{+/+}$, $cGas^{+/-}$ and $cGas^{-/-}$ mice using primers outlined in **a**. **c**, qRT–PCR of relative *cGas* expression in lungs (left) or BMMO (right) from wild-type B6 and $cGas^{-/-}$ mice. Data from lung represent means of three mice per group. Data from BMMO were derived from two independent experiments performed in triplicate. Error bars represent s.e.m. Statistical significance was determined by *t*-test. $^{*}P < 0.05$, $^{***}P < 0.001$.



Extended Data Figure 7 | Viral burden in mice infected with WNV. Wild-type or *cGas*$^{-/-}$ mice were infected with WNV and viral titres in several regions of the brain were determined by plaque assay. $n = 10$ mice per group. Statistical significance was determined by *t*-test.



Extended Data Figure 8 | Role for *cGAS* in BMMO activation. a, BMMO from wild-type and $cGas^{-/-}$ mice were analysed for baseline expression of chemokines *Ccl5* and *Cxcl10* by RT–PCR. **b,** BMMO from wild-type and $cGas^{-/-}$ mice were treated with polyIC (pIC) or transfected with polyIC (Tf-pIC) and *Ifnb* and *Ifit1* levels were determined by qRT–PCR. In both panels, gene expression levels are relative to the housekeeping gene *RPS29* and normalized to mock-treated wild-type cells. Data represent two experiments performed in triplicate. Error bars represent s.d. Statistical significance was determined by *t*-test. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

Extended Data Table 1 | Classification of viruses screened in this study

Virus	Genome content	Family	Genus
vaccinia virus	dsDNA	*Poxviridae*	*Orthopoxvirus*
coxsackie B virus	(+)ssRNA	*Picornaviridae*	*Enterovirus*
poliovirus	(+)ssRNA	*Picornaviridae*	*Enterovirus*
equine arterivirus	(+)ssRNA	*Arteriviridae*	*Arterivirus*
Sindbis virus*	(+)ssRNA	*Togaviridae*	*Alphavirus*
O'nyong nyong virus	(+)ssRNA	*Togaviridae*	*Alphavirus*
influenza A virus	(-)ssRNA†	*Orthomyxoviridae*	*Influenzavirus A*
human parainfluenza virus type 3	(-)ssRNA	*Paramyxoviridae*	*Respirovirus*
Newcastle disease virus	(-)ssRNA	*Paramyxoviridae*	*Avulavirus*
human metapneumovirus	(-)ssRNA	*Paramyxoviridae*	*Metapneumovirus*
respiratory syncytial virus	(-)ssRNA	*Paramyxoviridae*	*Pneumovirus*
measles virus	(-)ssRNA	*Paramyxoviridae*	*Morbillavirus*
Bunyamwera virus	(-)ssRNA†	*Bunyaviridae*	*Orthobunyavirus*

* Two strains of SINV, AR86 and Girdwood, were included in the screens.
† Segmented −ssRNA genomes; all other −ssRNA viruses have non-segmented genomes.
dsDNA, double-stranded DNA; +ssRNA, positive-sense single-stranded RNA; −ssRNA, negative-sense single-stranded RNA.

Extended Data Table 2 | List of genes induced twofold or more in *STAT1*$^{-/-}$ fibroblasts transduced with lentivirus expressing cGAS compared to Fluc control

Gene	FC	p value	Accession
PARP3	2.0	0.040	NM_001003931.1
LAMP3	2.0	0.024	NM_014398.2
CEACAM1	2.0	0.015	NM_001024912.1
GBP1	2.0	0.013	NM_002053.1
C6orf192	2.1	0.011	NM_052831.2
PODXL	2.1	0.024	NM_001018111.1
VNN2	2.1	0.043	NM_004665.2
C1R	2.1	0.040	NM_001733.4
CDCP1	2.1	0.005	NM_022842.3
TDRD7	2.1	0.023	NM_014290.1
SAMD9	2.1	0.025	NM_017654.2
PARP12	2.1	0.014	NM_022750.2
CASZ1	2.1	0.015	NM_017766.3
APOBEC3G	2.2	0.024	NM_021822.1
IFIT3	2.2	0.046	NM_001031683.1
ABI3BP	2.2	0.028	NM_015429.2
LOC728855	2.2	0.041	NR_024510.1
LGALS3BP	2.2	0.036	NM_005567.2
IFI44L	2.2	0.023	NM_006820.1
ZC3HAV1	2.2	0.018	NM_020119.3
DHX58	2.2	0.039	NM_024119.2
OAS2	2.3	0.023	NM_001032731.1
ACSS1	2.3	0.025	NM_032501.2
HLA-DRA	2.3	0.017	NM_019111.3
CADPS2	2.3	0.009	NM_017954.9
USP18	2.4	0.028	NM_017414.3
TMEM62	2.4	0.022	NM_024956.3
BTN3A2	2.4	0.018	NM_007047.3
LRRC17	2.4	0.023	NM_001031692.1
EPSTI1	2.4	0.037	NM_033255.2
PARP14	2.5	0.028	NM_017554.1
HLA-F	2.5	0.007	NM_018950.1
IFIT3	2.5	0.032	NM_001549.2
APOBEC3G	2.5	0.018	NM_021822.1
GIMAP2	2.5	0.018	NM_015660.2
GBP4	2.6	0.036	NM_052941.3
DDX60	2.6	0.023	NM_017631.4
ITGA2	2.6	0.028	NM_002203.3
UBE2L6	2.6	0.029	NM_004223.3
CASP1	2.6	0.034	NM_033294.2
HERC6	2.6	0.047	NM_017912.3
UBE2L6	2.6	0.011	NM_004223.3
ZMYND15	2.7	0.025	NM_032265.1
CXCL16	2.7	0.011	NM_022059.1
PMAIP1	2.8	0.013	NM_021127.1
IFIH1	2.8	0.027	NM_022168.2
HLA-H	2.8	0.015	NR_001434.1
ZC3HAV1	2.9	0.037	NM_024625.3
OAS2	2.9	0.029	NM_016817.2
LRRC17	2.9	0.023	NM_005824.1
IFI35	3.0	0.015	NM_005533.2
MX1	3.1	0.012	NM_002462.2
SLC15A3	3.1	0.007	NM_016582.1
PSMB8	3.2	0.013	NM_148919.3
ISG20	3.2	0.002	NM_002201.4
PSMB8	3.3	0.007	NM_148919.3
IFIT3	3.3	0.007	NM_001031683.1
PSMB8	3.5	0.002	NM_004159.4
PLCG2	3.5	0.007	NM_002661.2
IL8	3.5	0.013	NM_000584.2
CASP1	3.7	0.012	NM_033294.2
TNFRSF1B	3.7	0.029	NM_001066.2
HERC5	3.7	0.007	NM_016323.2
ARL9	3.8	0.013	NM_206919.1
PSMB9	3.9	0.006	NM_002800.4
RARRES3	3.9	0.012	NM_004585.3
OASL	4.4	0.007	NM_198213.1
HLA-B	5.0	0.005	NM_005514.5
IFI44	5.2	0.007	NM_006417.3
cGAS	9.4	0.007	NM_138441.2

FC, fold change.

LETTER

doi:10.1038/nature12756

In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features

Yiliang Ding[1,2,3]*, Yin Tang[1,3,4]*, Chun Kit Kwok[2,3]*, Yu Zhang[4,5], Philip C. Bevilacqua[2,3,6] & Sarah M. Assmann[1,3,4,6]

RNA structure has critical roles in processes ranging from ligand sensing to the regulation of translation, polyadenylation and splicing[1–4]. However, a lack of genome-wide *in vivo* RNA structural data has limited our understanding of how RNA structure regulates gene expression in living cells. Here we present a high-throughput, genome-wide *in vivo* RNA structure probing method, structure-seq, in which dimethyl sulphate methylation of unprotected adenines and cytosines is identified by next-generation sequencing. Application of this method to *Arabidopsis thaliana* seedlings yielded the first *in vivo* genome-wide RNA structure map at nucleotide resolution for any organism, with quantitative structural information across more than 10,000 transcripts. Our analysis reveals a three-nucleotide periodic repeat pattern in the structure of coding regions, as well as a less-structured region immediately upstream of the start codon, and shows that these features are strongly correlated with translation efficiency. We also find patterns of strong and weak secondary structure at sites of alternative polyadenylation, as well as strong secondary structure at 5′ splice sites that correlates with unspliced events. Notably, *in vivo* structures of messenger RNAs annotated for stress responses are poorly predicted *in silico*, whereas mRNA structures of genes related to cell function maintenance are well predicted. Global comparison of several structural features between these two categories shows that the mRNAs associated with stress responses tend to have more single-strandedness, longer maximal loop length and higher free energy per nucleotide, features that may allow these RNAs to undergo conformational changes in response to environmental conditions. Structure-seq allows the RNA structurome and its biological roles to be interrogated on a genome-wide scale and should be applicable to any organism.

Most existing RNA structure mapping has been performed *in vitro*[5–8]. Among RNA structure probing reagents, dimethyl sulphate (DMS) can penetrate cells and has been used to map structures of high-abundance RNAs *in vivo* in various organisms[9–12]. DMS methylates the base-pairing faces of A and C of RNA in loops, bulges, mismatches and joining regions. The base-pairing status of U and G nucleotides can be inferred from structural mapping of As and Cs, because constraining even some nucleotides substantially improves predictions of other regions[13]. However, a method for genome-wide study of RNA structure *in vivo* has been lacking. Here we combine DMS methylation with next-generation sequencing to establish structure-seq, an *in vivo* quantitative measurement of genome-wide RNA secondary structure at nucleotide resolution.

We optimized DMS treatment conditions for *Arabidopsis* etiolated seedlings (Extended Data Fig. 1a), and then generated two independent biological replicates of (+)DMS and (−)DMS libraries (Fig. 1). DMS-induced methylation sites were highly reproducible (Pearson correlation coefficient (PCC) of 0.91 for the two (+)DMS libraries (Extended Data Table 1a)). Nucleotide modification in the (+)DMS library was specific to As and Cs (Extended Data Fig. 1b). Notably, 98% of the combined 206 million sequence reads were mappable to the *Arabidopsis* genome; these reads include diverse classes of RNAs, with a predominance of mRNAs and ribosomal RNAs (Extended Data Fig. 1c and Extended Data Table 1b, c). The reverse transcriptase stops are evenly distributed along the transcripts, with no 3′ bias (Extended Data Fig. 1d). In particular, 10,781 transcripts had sufficient coverage at nucleotide resolution to obtain secondary-structure constraints (Extended

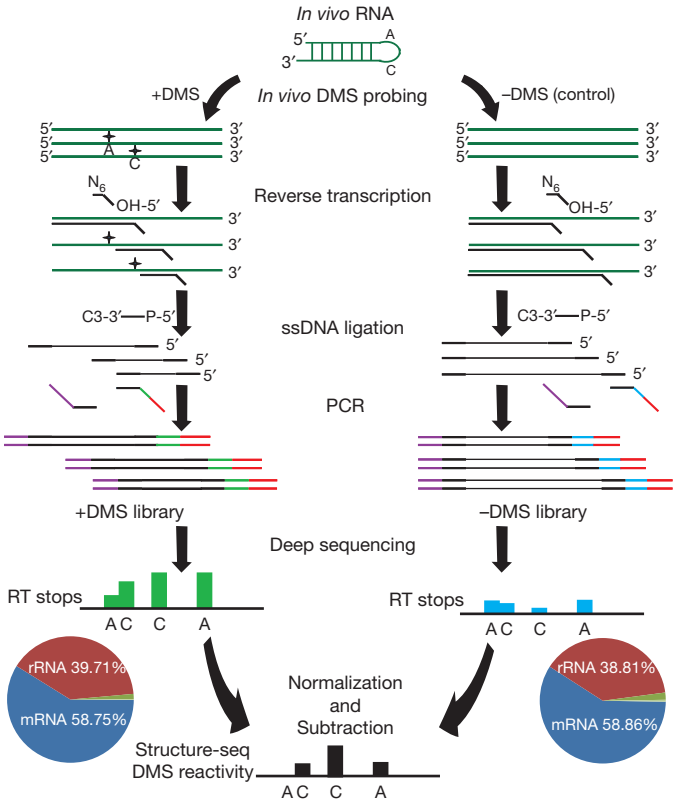


Figure 1 | Overview of structure-seq. *Arabidopsis* seedlings are treated with DMS. Reverse transcription is performed using random hexamers (N_6) with adaptors (thicker black lines). Reverse transcriptase stalls one nucleotide before DMS-modified As and Cs[11] (black crosses). Single-stranded (ss) DNA ligation attaches a single-stranded DNA linker (thicker black line) to the 3′ end. Double-stranded DNA is generated by PCR (purple line, forward primer; green–red line, unique index (green) and universal portion (red) of reverse primer). A (−)DMS library is prepared in parallel. Deep sequencing is performed with different indices for (+)DMS and (−)DMS libraries. Counts of the reverse transcriptase (RT) stops are normalized and subtracted. Pie charts depict percentages of RNA types for the (+)DMS (left) and (–)DMS (right) libraries. Green portions represent other RNA types plus unmappable reads (see Extended Data Table 1b, c).

[1]Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [2]Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [3]Center for RNA Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [4]Bioinformatics and Genomics Graduate Program, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [5]Department of Statistics, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [6]Plant Biology Graduate Program, Pennsylvania State University, University Park, Pennsylvania 16802, USA.
*These authors contributed equally to this work.

Data Fig. 2a). Abundance of individual mRNAs in structure-seq correlated well with mRNA abundance from RNA-seq analyses[14] (Extended Data Fig. 2b, c).

To validate *in vivo* structure-seq, we mapped DMS reactivities of 18S rRNA (Fig. 2a and Extended Data Fig. 3). Overall, the reactivities are consistent (Extended Data Fig. 3) with structure mapping of 30S subunit-bound 16S rRNA[15] and with the phylogenetically derived structures[16], which are the evolutionarily conserved structures and are the closest models of *in vivo*, protein-associated structure[17]. Further, comparison of DMS modifications from structure-seq with those from conventional gel-based *in vivo* structure probing yielded strong agreement for all regions of 18S rRNA tested (PCCs of 0.78 (Fig. 2b, c), 0.71 (Extended Data Fig. 4a, b) and 0.68 (Extended Data Fig. 4c, d)), as well as for a randomly chosen mRNA, *CAB1* (At1g29930) (Extended Data Fig. 4e, f, g). We thus conclude that structure-seq accurately probes RNA structures *in vivo* on a genome-wide basis. Importantly, complete coverage can be provided in a single experiment even for long transcripts, which is not the case for conventional gel-based methods.

We accordingly investigated global features and discovered several notable genome-wide *in vivo* RNA structural properties of *Arabidopsis* mRNAs (Fig. 3). We found that the average DMS reactivity of untranslated regions (UTRs) is significantly higher than that of coding sequences (CDS) (Extended Data Fig. 5a). The ~5 nucleotides (nt) immediately upstream of the start codon show particularly high DMS reactivity, which indicates less structure (Extended Data Fig. 5a). These findings agree with previous findings in yeast and *Arabidopsis* UTRs *in vitro*[5,6] and with *in silico* predictions in mouse and human[18]. Unstructured regions near start codons may facilitate ribosome binding and translation initiation. To evaluate this hypothesis, we ranked our mRNAs according to their polyribosome association on the basis of previous *in vivo* polyribosome profiling in *Arabidopsis* seedlings[19]. The unstructured region upstream of the start codon was enriched in high translation efficiency mRNAs and was absent in low translation efficiency mRNAs (Fig. 3a). Although a related observation was made *in vitro* for yeast[5], our data demonstrate that this is a genuine *in vivo* phenomenon, and extend these results to the plant kingdom.

When DMS reactivity along the CDS was averaged across mRNAs in our data set (see Methods for details), a periodic trend was revealed. A discrete Fourier transform applied to the CDS gave a period of 3, whereas periodicity was absent in UTR regions (Fig. 3a insets and

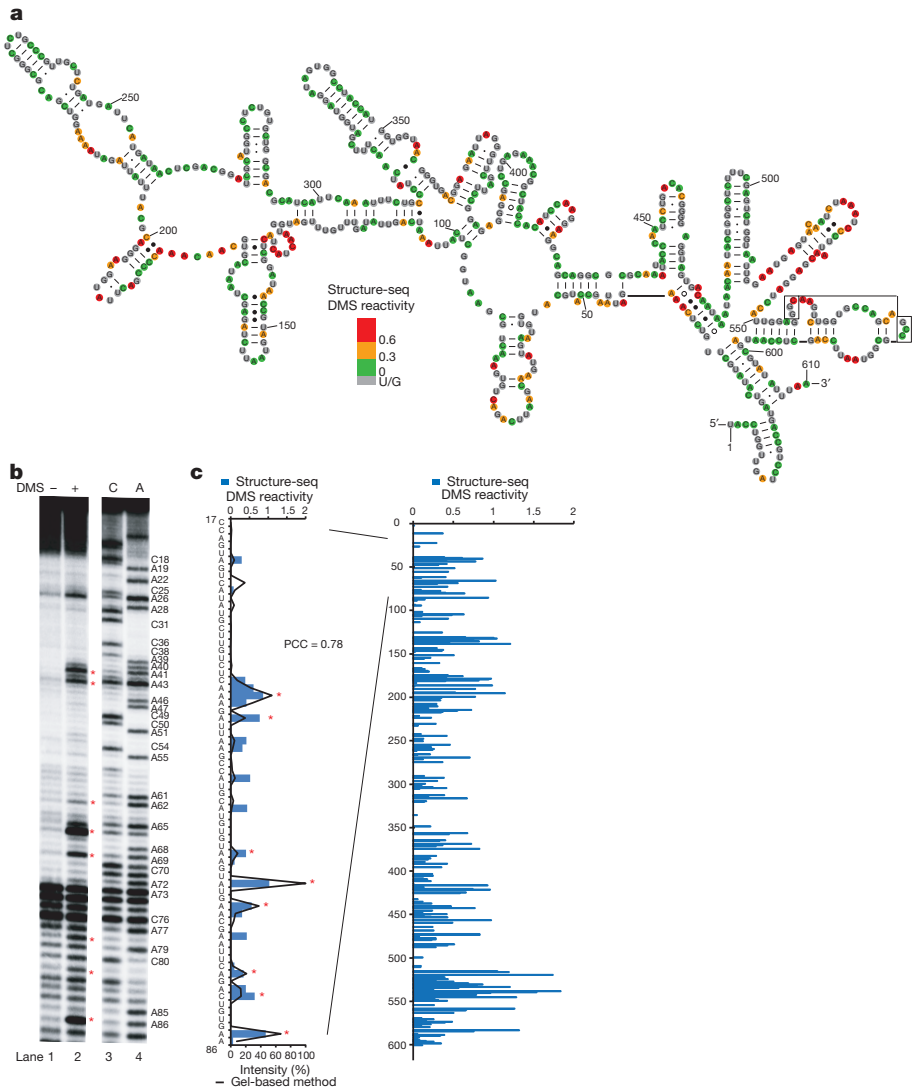


Figure 2 | Structure-seq accurately maps 18S rRNA and agrees with gel-based *in vivo* structure probing. **a**, Nucleotides 1–610 of the phylogenetic 18S rRNA structure[16], colour-coded according to structure-seq DMS reactivity. **b**, Nucleotides 17–86 of 18S rRNA structure-mapped by gel-based probing. Lanes 1–2, (−)DMS and (+)DMS treatments; lanes 3–4, C/A sequencing. **c**, Comparison of structure-seq (blue bars) and gel-based probing (black line, normalized to 0–100%) yields a PCC of 0.78. Structure-seq reactivity for nucleotides 1−610 is shown on the right. The red asterisks indicate nucleotides that have significant DMS modifications from both methods, and are also shown in panel **b**.

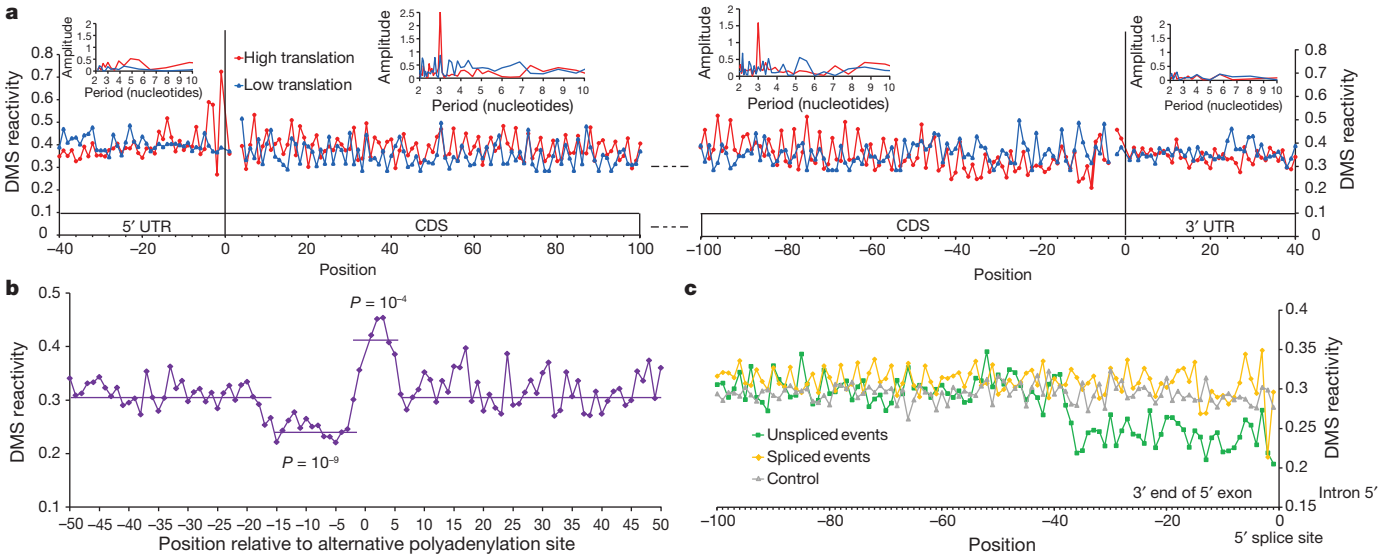


Figure 3 | Structure-seq reveals new features of mRNA secondary structures that prevail *in vivo*. **a**, RNA structure associated with translation. DMS reactivities of selected regions (5′ UTR, 40 nt upstream of the start codon; CDS, 100 nt downstream of the start codon and 100 nt upstream of the stop codon; 3′ UTR, 40 nt downstream of the stop codon) across high (red) or low (blue) translation efficiency mRNAs were averaged. mRNAs were aligned by their start/stop codons (vertical black lines). Discrete Fourier transforms (insets) of average DMS reactivity of selected regions across high (red) or low (blue) translation efficiency mRNAs shows structural periodicity only in high translation efficiency CDS. **b**, RNA structures associated with alternative polyadenylation. DMS reactivities 50 nt upstream and downstream of alternative polyadenylation sites (indicated by 0) were averaged (violet). One region with significantly lower DMS reactivity (−15 to −2 nt, $P = 10^{-9}$, Student's *t*-test) and one region with significantly higher DMS reactivity (−1 to 5 nt, $P = 10^{-4}$, Student's *t*-test) are highlighted. **c**, RNA structure associated with alternative splicing. DMS reactivities along 100 nt of the 3′ end of the 5′ exon were averaged from each of unspliced (green) and spliced (yellow) events. For unspliced events, the significance of the difference in average DMS reactivity between the 40 nt upstream of the 5′ splice site and the remaining 60-nt upstream region was $P = 10^{-25}$ (Student's *t*-test). For spliced events, the *P* value was > 0.05. Absence of structure in a nucleotide composition control is in grey.

Extended Data Fig. 5b, c). This represents the first *in vivo* demonstration of triplet periodicity in the structure of the CDS in a multicellular organism. Observation of an *in vivo* triplet periodicity in CDS structure in plants, as well as its presence in both *in vivo* (from ribosome profiling)[20] and *in vitro* (ribosome-free)[5] yeast data sets, and its proposed presence in mammals[18], suggests that periodic structure may have evolved as a universal regulatory feature of translated portions of mRNAs.

Our genome-wide *in vivo* structurome allowed us to evaluate the hypothesis that robustness of the periodic structure signal might influence translation. Notably, the periodic signal was intensified in high translation efficiency transcripts and absent from low translation efficiency transcripts[5] (Fig. 3a insets and Extended Data Fig. 5d). Further analysis revealed that differential presence of periodic structure between these two mRNA populations did not arise from differential codon usage or differential nucleotide bias in any of the three codon positions (Extended Data Fig. 5e). Our results thus reveal a hidden code in *in vivo* RNA structure that influences polyribosome association and, by inference, translation[21].

Alternative polyadenylation has been observed for ~60% of *Arabidopsis* mRNAs[22]. We assessed DMS modification 50 nt upstream and downstream of the known[22] alternative polyadenylation cleavage sites for the corresponding 5,959 mRNAs in our RNA structurome. For alternative polyadenylation, RNA secondary structure upstream of the cleavage site from nt −15 to −2 showed significantly lower DMS reactivity than the average reactivity throughout the 100-nt region, indicating more structure *in vivo* in the U- and A-rich upstream region (Fig. 3b and Extended Data Fig. 6a). This finding provides genome-wide support for a regulatory role of RNA structure in this region, in line with an early mutagenesis study of polyadenylation efficiency on one selected RNA assayed *in vitro*[23]. We also found that nt −1 to 5 had significantly higher DMS reactivity than average (Fig. 3b). This leads to a structured–unstructured pattern (Fig. 3b) that is not simply due to nucleotide composition (Extended Data Fig. 6b, c). These results, newly revealed by structure-seq, suggest that structural elements near the cleavage site may help to regulate alternative polyadenylation.

Alternative splicing has been proposed to be regulated by RNA secondary structure[24,25]. We considered a previous compilation of alternative splicing events in *Arabidopsis* seedlings[26] and identified, for each mRNA in our data set, whether introns were spliced out or whether alternative splicing (including exon skipping and intron retention) occurred. Notably, we found significantly lower DMS reactivity in the region ~40 nt upstream of the 5′ splice site for the unspliced events (Fig. 3c). This structural pattern was not found in the spliced events or in a nucleotide composition control (Fig. 3c), nor was it apparent at the 3′ splice site (Extended Data Fig. 6d). Secondary structure at the 5′ splice site appears to disfavour the first step of splicing, providing a regulatory mechanism for alternative splicing.

Current *in silico* structure prediction based on thermodynamics estimates a set of probable RNA structures, but constraints from experimental data significantly improve predictions[13,27]. Individual nucleotide DMS reactivities for each of the 10,623 mRNAs with ≥ 1 reverse transcriptase stop/nucleotide provided a rich data set (Fig. 4a) to compare RNA structure predictions with and without inclusion of *in vivo* DMS-guided structural constraints. First, we compared *in silico*-predicted structures and our *in vivo* structures with available *in vitro* structures[6]. We find that *in vitro* and *in vivo* structures differ, and that *in vitro* structures are more similar to *in silico* structures than are *in vivo* structures (Extended Data Fig. 7a). Next, using RNAstructure[27], we calculated for each of the 10,623 mRNAs the positive predictive value (PPV)[28], which indicates the proportion of base pairs in the *in vivo* DMS-constrained RNA structure that also appear in the *in silico*-predicted RNA structure. Most mRNAs did not fold *in vivo* according to *in silico*-predicted structures, as is evident from the broad PPV distribution (Fig. 4b). Such poor correlation could, in theory, be explained by mRNA association with proteins that block DMS reactivity *in vivo*. This hypothesis was not supported, however, as low

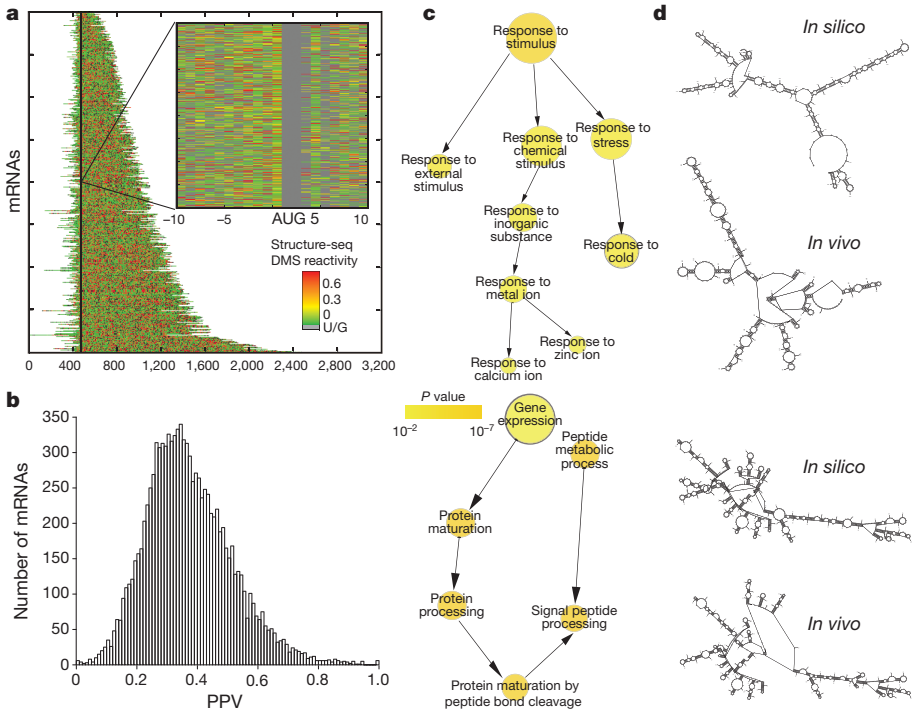


Figure 4 | Structure-seq provides *in vivo* RNA structure information at nucleotide resolution across 10,623 mRNAs and reveals correlations between RNA structure and biological function. **a**, DMS reactivity of each of 10,623 mRNAs. **b**, PPV distribution for *in vivo* versus *in silico* structures of 10,623 mRNAs; a higher PPV value indicates less difference. **c**, mRNAs with low PPV are enriched in functional annotations related to stress and stimulus responses; mRNAs with high PPV are enriched in basic biological functions. Gene Ontology categories over-represented in the 5% of 10,623 mRNAs with lowest and highest PPV are shown at the top and bottom, respectively. **d**, *In silico* and *in vivo* structures of one illustrative low PPV mRNA (top), *RCI2A* (At3g05880), are highly dissimilar, whereas such structures for one illustrative high PPV mRNA (bottom), *S24* peptidase (At1g52600), are highly similar.

reactivity did not correlate with low PPV, nor was PPV correlated with mRNA length (Extended Data Fig. 7b, c). The results of Fig. 4b and Extended Data Fig. 7a demonstrate the critical contribution of *in vivo* constraints in prediction of the RNA secondary structures that prevail in living cells. This is also illustrated by an improvement in predicting the phylogenetic structure of 18S rRNA when *in vivo* constraints are used (Extended Data Table 2).

We next asked whether genome-wide relationships exist between *in vivo* mRNA structures and biological functions of the encoded proteins. Intriguingly, the Gene Ontology annotations of those transcripts in the lowest 5% of the PPV distribution are enriched in annotations of biological functions related to stress and stimulus responses[29] (Fig. 4c and Extended Data Fig. 8a, b). For example, mRNAs of cold and metal ion stress-response genes folded significantly differently *in vivo* from their unconstrained *in silico* predictions (Fig. 4c, d and Extended Data Fig. 8a, b). Interestingly, these stresses are known to affect RNA structure and thermostability[27,30]. By contrast, genes involved in basic biological functions such as gene expression, protein maturation and processing, and peptide metabolic processes show little difference in their *in vivo*-constrained and *in silico*-predicted RNA secondary structures, as indicated by their enrichment in the highest 5% of the PPV distribution (Fig. 4c, d and Extended Data Fig. 8a, b). Speculatively, mRNAs related to cell maintenance and showing high PPV may have evolved to resist large conformational changes in order to maintain homeostasis.

Table 1 | RNA structural features differ between high and low PPV mRNAs

	Single-strand percentage	Maximum loop length of structure	Free energy per nucleotide
In silico	0.99	3.7×10^{-2}	7.73×10^{-3}
In vivo	5.80×10^{-19}	4.7×10^{-7}	3.07×10^{-34}

The significance of the difference for several RNA structural features was assessed between high PPV mRNAs and low PPV mRNAs. Each entry is the *P* value of a Student's *t*-test between the 5% of mRNAs with highest PPV and the 5% of mRNAs with lowest PPV. The comparisons were performed on *in silico*-predicted (without in *vivo* constraints) and *in vivo* (*in silico* prediction with constraints from our *in vivo* structure-seq data) structures. Small *P* values confirm that there are significant differences in RNA structural features between high- and low-PPV mRNAs.
(Pseudoknots are uncommon (~1 pseudoknot per 1,000 nt) in both high- and low-PPV mRNA data sets (calculated from the 1% mRNAs with highest PPV and the 1% mRNAs with lowest PPV). The *P* values for comparison of pseudoknot prevalence between these two groups are 0.48 and 0.31 for *in silico*-predicted and *in vivo* structures, respectively.)

We next compared several structural features between low and high PPV mRNAs. We found that the fraction of a mRNA's nucleotides with DMS reactivity greater than a 0.6 threshold is significantly higher in the low than in the high PPV mRNAs ($P = \sim 2 \times 10^{-42}$; two sample *t*-test), which provides experimental support independent of computational structure prediction that the low PPV mRNAs exist in multiple conformations and/or are less structured. The low PPV mRNAs, enriched in functions related to stress, also tend to have more single-stranded regions (consistent with higher average reactivity per nucleotide; $P = \sim 10^{-85}$; Student's *t*-test), longer maximum loop length and higher free energy per nucleotide when assessed *in vivo* (Table 1 and Extended Data Fig. 8b). These features might favour change in RNA structure in response to, for example, cold or metal ions, stress conditions with which these mRNAs are associated (Fig. 4c). In other words, stress-response RNAs may be more plastic, changing their structure in response to changing cellular conditions. As sessile organisms, plants face extreme environmental stresses; it will be of interest to ascertain whether the RNA structure–function relationships revealed in Fig. 4c, d prevail in other kingdoms.

In summary, we have established a high throughput, genome-wide method that profiles RNA secondary structure with high accuracy and nucleotide resolution *in vivo*. Our comprehensive study reveals new insights into how global native RNA structural characteristics regulate RNA processing and translation, and associates mRNA structural characteristics with functions of the encoded proteins. These trends are not discernible by studies on just one or a few RNAs, nor are they necessarily found in *in vitro* genome-wide studies. Structure-seq provides a broadly applicable method for the investigation of RNA structure–function relationships in living systems.

METHODS SUMMARY

Five-day-old *Arabidopsis thaliana* etiolated seedlings were treated with DMS, followed by dithiothreitol quench. Extracted poly(A)-selected RNA was reverse transcribed. First-strand complementary DNAs were ligated at their 3′ ends to a DNA linker and PCR was performed. Different barcode indices were used for the (+)DMS and (−)DMS libraries, which were subjected to Illumina sequencing. Two independent biological replicates were performed. Reads were mapped to the *Arabidopsis* transcriptome and genome using Bowtie (v.0.12.8). The natural log (ln) was taken of reverse transcriptase stops in both (+) and (−) DMS libraries, followed by normalization for abundance and length. Raw DMS reactivity was

calculated by subtracting from the normalized (+)DMS library values the normalized number of reverse transcriptase stops in the (−)DMS library, and further normalized (2–8% normalization) to obtain the final DMS reactivity of each nucleotide. PPVs were used to compare *in vivo*- and *in silico*-predicted structures for each mRNA. mRNAs with PPV values in the top and bottom 5% were subjected to Gene Ontology analysis using the hypergeometric test ($P < 0.01$ as significant).

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 8 April; accepted 9 October 2013.
Published online 24 November 2013.

1. Buratti, E. *et al.* RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol. Cell. Biol.* **24,** 1387–1400 (2004).
2. Cruz, J. A. & Westhof, E. The dynamic landscapes of RNA architecture. *Cell* **136,** 604–609 (2009).
3. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361,** 13–37 (2005).
4. Sharp, P. A. The centrality of RNA. *Cell* **136,** 577–580 (2009).
5. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467,** 103–107 (2010).
6. Li, F. *et al.* Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* **24,** 4346–4359 (2012).
7. Zheng, Q. *et al.* Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis*. *PLoS Genet.* **6,** e1001141 (2010).
8. Wan, Y. *et al.* Genome-wide measurement of RNA folding energies. *Mol. Cell* **48,** 169–181 (2012).
9. Senecoff, J. F. & Meagher, R. B. *In vivo* analysis of plant RNA structure: soybean 18S ribosomal and ribulose-1,5-bisphosphate carboxylase small subunit RNAs. *Plant Mol. Biol.* **18,** 219–234 (1992).
10. Wells, S. E., Hughes, J. M. X., Igel, A. H. & Ares, M. Use of dimethyl sulfate to probe RNA structure *in vivo*. *Methods Enzymol.* **318,** 479–493 (2000).
11. Zaug, A. J. & Cech, T. R. Analysis of the structure of Tetrahymena nuclear RNAs *in vivo*: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA* **1,** 363–374 (1995).
12. Zemora, G. & Waldsich, C. RNA folding in living cells. *RNA Biol.* **7,** 634–641 (2010).
13. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA* **101,** 7287–7292 (2004).
14. Oh, E., Zhu, J. Y. & Wang, Z. Y. Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses. *Nature Cell Biol.* **14,** 802–809 (2012).
15. Moazed, D., Stern, S. & Noller, H. F. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J. Mol. Biol.* **187,** 399–416 (1986).
16. Cannone, J. J. *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3,** 2 (2002).
17. Gutell, R. R., Lee, J. C. & Cannone, J. J. The accuracy of ribosomal RNA comparative structure models. *Curr. Opin. Struct. Biol.* **12,** 301–310 (2002).
18. Shabalina, S. A., Ogurtsov, A. Y. & Spiridonov, N. A. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* **34,** 2428–2437 (2006).
19. Branco-Price, C., Kawaguchi, R., Ferreira, R. B. & Bailey-Serres, J. Genome-wide analysis of transcript abundance and translation in *Arabidopsis* seedlings subjected to oxygen deprivation. *Ann. Bot. (Lond.)* **96,** 647–660 (2005).
20. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–223 (2009).
21. Branco-Price, C., Kaiser, K. A., Jang, C. J., Larive, C. K. & Bailey-Serres, J. Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in *Arabidopsis thaliana*. *Plant J.* **56,** 743–755 (2008).
22. Shen, Y. *et al.* Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. *Genome Res.* **21,** 1478–1486 (2011).
23. Loke, J. C. *et al.* Compilation of mRNA polyadenylation signals in *Arabidopsis* revealed a new signal element and potential secondary structures. *Plant Physiol.* **138,** 1457–1468 (2005).
24. Solnick, D. Alternative splicing caused by RNA secondary structure. *Cell* **43,** 667–676 (1985).
25. Jin, Y., Yang, Y. & Zhang, P. New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. *RNA Biol.* **8,** 450–457 (2011).
26. Filichkin, S. A. *et al.* Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* **20,** 45–58 (2010).
27. Lu, Z. J., Gloor, J. W. & Mathews, D. H. Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* **15,** 1805–1813 (2009).
28. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl Acad. Sci. USA* **106,** 97–102 (2009).
29. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25,** 25–29 (2000).
30. Misra, V. K. & Draper, D. E. The linkage between magnesium binding and RNA folding. *J. Mol. Biol.* **317,** 507–521 (2002).

Acknowledgements This research is supported by Human Frontier Science Program (HFSP) grant RGP0002/2009-C, the Penn State Eberly College of Science, and a Penn State Huck Institutes HITS grant to P.C.B. and S.M.A. We thank F. Pugh, Y. Li, A. Chan and K. Yen for help with Illumina sequencing; D. Mathews and A. Spasic for advice on RNA structure analysis; M. Axtell for reading of the manuscript; and P. Raghavan for access to the CyberSTAR server, funded by the National Science Foundation through grant OCI–0821527. We also thank L. Song, D. Chadalavada and S. Ghosh for discussions.

Author Contributions Y.D. and C.K.K. performed the experiments. Y.D., Y.T. and C.K.K. performed data analysis. Statistical analyses were designed by Y.Z. and Y.T., with input from all authors. Y.D., Y.T. and C.K.K. contributed equally to this work. All authors contributed ideas, discussed the results and wrote the manuscript.

Author Information Sequencing data are deposited in the Sequence Read Archive (SRA) on the NCBI website under the accession number SRP027216. 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 S.M.A. (sma3@psu.edu) or P.C.B. (pcb5@psu.edu).

METHODS

Plant materials and growth conditions. *Arabidopsis thaliana* seeds of the Columbia (Col-0) accession were sterilized with 70% (v/v) ethanol and plated on half-strength Murashige and Skoog medium. The plates were wrapped in foil and stratified at 4 °C for 3–4 days and then grown in a 22–24 °C growth chamber for 5 days.

***In vivo* DMS chemical probing.** All manipulations involving DMS were conducted in a chemical fume hood.

Five-day-old *A. thaliana* etiolated seedlings grown as described above were suspended intact and completely covered in 20 ml 1 × DMS reaction buffer in a 50 ml Falcon tube that contained 100 mM KCl, 40 mM HEPES (pH 7.5) and 0.5 mM $MgCl_2$. DMS was added to a final concentration of 0.75% (~75 mM) and allowed to react for 15 min at room temperature (~22 °C) with periodic swirling. This DMS concentration and reaction time allowed DMS to penetrate plant cells and modify the RNA *in vivo* with single-hit kinetics conditions. Single-hit kinetics conditions can be directly observed in the (+)DMS lanes of Fig. 2b and Extended Data Figs 1a and 4e, in which an intense full-length peak is observed for both rRNA and mRNA, and is confirmed in structure-seq data by the presence of transcripts with no internal reverse transcriptase stops. To quench the reaction, freshly prepared dithiothreitol (DTT) was added to a final concentration of 0.5 M, and after swirling for 2 min the reaction mixture was decanted and the seedlings were washed with ~2 × 50 ml deionized water. The seedlings were immediately frozen with liquid N_2 and ground into powder using a mortar and pestle pre-cleaned with RNase Zap (Ambion). Lysis buffer was added to the powder, and then the sample was subjected to total RNA extraction, following the protocol described in the RNeasy Plant Mini Kit (Qiagen).

Illumina library construction. *In vivo* total RNA isolation was followed by one round of poly(A) selection using the Poly(A) purist Kit (Ambion). The poly(A)-selected RNA (2 μg) was then treated with TURBO DNase (Ambion) following the manufacturer's protocol, followed by phenol chloroform extraction and ethanol precipitation. The RNA was re-suspended in RNase-free water and subjected to reverse transcription using the SuperScript III First Strand Kit (Invitrogen) and random hexamers fused with an Illumina TruSeq Adapter (5′-CAGACGTGT GCTCTTCCGATCNNNNNN-3′). The resultant first-strand cDNAs were then ligated at their 3′ ends to a ssDNA linker (5′-pNNNAGATCGGAAGAGCGTC GTGTAG-3′-Spacer, where '5′ p' is a 5′ phosphate and '3′-Spacer' is a 3-carbon linker) using CircLigase ssDNA Ligase (Epicentre), with slight modifications to the manufacturer's and literature procedures[31], as follows. In brief, the cDNA was re-dissolved in RNase-free water and reagents were added to yield the following final concentrations in a total volume of 20 μl: 70 μM ssDNA linker, 50 mM MOPS (pH 7.5), 10 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.05 mM ATP, 2.5 mM $MnCl_2$ and 200 U total CircLigase. The ligation was performed at 65 °C for 12 h and then the sample was heated at 85 °C for 15 min to deactivate the CircLigase. PCR amplification was performed on the ligated cDNA using Illumina TruSeq Primers (Illumina TruSeq forward primer, 5′-AATGATACGGCGACCACCGA GATCTACACTCTTTCCCTACACGACGCTCTTCCGATC T-3′; Illumina TruSeq reverse primer index 1, 5′-CAAGCAGAAGACGGCATACGAGATTGGTCAGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC-3′; Illumina TruSeq reverse primer index 2, 5′-CAAGCAGAAGACGGCATACGAGATGATCTGGTGACT GGAGTTCAGACGTGTGCTCTTCCGATC-3′). Three rounds of gel purification were performed to remove adapters and achieve a uniform size distribution of PCR products between 150 and 650 base pairs (bp) using both a 50-bp DNA Ladder and a 1 Kb Plus DNA Ladder (Invitrogen) as references. This, together with carefully measured loading DNA concentration, allowed an optimized cluster density to reduce unmappable reads (c.f. the manufacturer's protocol (Illumina)). Different barcode indices were used for the (+)DMS library and (−)DMS libraries. The dsDNA libraries were subjected to next-generation sequencing on an Illumina HiSeq 2000. An independent biological replicate was prepared in the same way and separately subjected to next-generation sequencing.

Illumina sequence mapping. Illumina sequencing read lengths of 37 nt were obtained and mapped to the *Arabidopsis thaliana* transcriptome and genome (TAIR v10 release 2010). Twenty-one nucleotides was determined to be the threshold length required for unique mapping of a sequencing read after the reads were linker trimmed at their 3′ ends. Up to three mismatches without any insertions or deletions were allowed to account for PCR and sequencing errors. Reads that could not be mapped or uniquely mapped to the genome were designated as 'not mappable'. Mapping of the reads was performed using Bowtie[32] (v0.12.8) (http://bowtie-bio.sourceforge.net/index.shtml).

As shown in Extended Data Table 1a, there is high correlation between the two (+)DMS libraries and between the two (−)DMS libraries from the biological replicates. Therefore, biological replicates were combined for further analysis.

Determination and normalization of DMS reactivity. To compare the (+)DMS and (−)DMS data sets and derive the final DMS reactivity for each nucleotide, the following three-step procedure was used:

Step 1. For a transcript, suppose $P_r(i)$ and $M_r(i)$ are the raw numbers of reverse transcriptase stops for nucleotide i (including all four bases) on the transcript in the (+)DMS and (−)DMS libraries (P and M, respectively), l is the length of the transcript and $P_r(0)$ and $M_r(0)$ are the raw numbers of full-length reverse transcriptase reads on the transcript in the (+)DMS and (−)DMS libraries, respectively.

For each nucleotide on each transcript, take the natural log (ln) of the number of reverse transcriptase stops mapped to that nucleotide position ($\ln[P_r(i)]$ or $\ln[M_r(i)]$) and divide the number by the average of the ln of reverse transcriptase stops per position, yielding equations (1) and (2). The average of the ln of reverse transcriptase stops per position is calculated as the sum of the ln of reverse transcriptase stops at each position (including all four bases (as random reverse transcriptase stalling can occur at any base) and full length reverse transcriptase reads) of the entire transcript, divided by the length of the transcript, as provided in the denominators of equations (1) and (2).

$$P(i) = \frac{\ln[P_r(i)]}{(\sum_{i=0}^{l} \ln[P_r(i)])/l} \quad (1)$$

Equation (1) is the normalized number of reverse transcriptase stops for nucleotide i in the (+)DMS library.

$$M(i) = \frac{\ln[M_r(i)]}{(\sum_{i=0}^{l} \ln[M_r(i)])/l} \quad (2)$$

Equation (2) is the normalized number of reverse transcriptase stops for nucleotide i in (−)DMS library.

Step 2. For each nucleotide, the raw DMS reactivity is calculated by subtracting the normalized number of reverse transcriptase stops for the nucleotide between (+) and (−)DMS libraries. All negative values are taken as 0 for the raw DMS reactivity.

$$\theta(i) = \max((P(i) - M(i)), 0) \quad (3)$$

Equation (3) gives the raw DMS reactivity for nucleotide i.

Step 3. Normalization (2–8% normalization[28]) is then performed on the raw DMS reactivity, $\theta(i)$, of all the nucleotides on all the transcripts to obtain the final DMS reactivity of each nucleotide. The reactivity is capped at seven[5].

In all of the figures in which the average DMS reactivity of a region is given, it is the average of the DMS reactivity of all adenine and cytosine nucleotides in that region, for all of the transcripts under consideration. Those transcripts that have no reverse transcriptase stops for any of the nucleotides are not used in further structure analyses, as they provide no structure information.

***In vivo* RNA structure analysis of the genome-wide transcriptome using DMS reactivity.** Global *in vivo* mRNA structure trends. We determined global transcriptome trends in mRNA structure by averaging DMS reactivity from selected regions of mRNAs: the 5′ UTR region (the first 40 nt upstream of the start codon); the CDS-beginning region (the first 100 nt downstream of the start codon); the CDS-ending region (the 100 nt upstream of the stop codon); and the 3′ UTR region (the first 40 nt downstream of the stop codon). There were 22,721 unique mRNAs (including splice variants) that had at least 40 nt in both the 5′ UTR region and the 3′ UTR region and at least 200 nt in the CDS; these mRNAs were analysed for global trends (Extended Data Fig. 5a).

We analysed the global mRNA structure of polyribosome-associated mRNAs defined in a previous study[21], ranking the transcripts according to their polyribosome-associated mRNA abundance relative to their mRNA abundance. We selected the top 5% (1,136 mRNAs) and the bottom 5% (1,136 mRNAs) of mRNAs from the ranking. We defined the top 5% as the 'high translation efficiency mRNAs' and the bottom 5% as the 'low translation efficiency mRNAs'. We analysed the global transcriptome trends of DMS reactivity of the 5′ UTR, CDS and 3′ UTR for both the high translation efficiency mRNAs and the low translation efficiency mRNAs (Fig. 3a).

Codon periodicity and codon position signature. We assessed the codon periodicity by applying a discrete Fourier transform. We collected the DMS reactivity data from the Fourier-transformed patterns of the 40-nt 5′ UTR, the first 100 nt of the CDS, the last 100 nt of the CDS and the 40-nt 3′ UTR regions (Fig. 3a and Extended Data Fig. 5b). We also computed the average DMS reactivity for each codon position, collected from the entire CDS across 22,721 unique mRNAs (see above for explanation of mRNAs chosen). We applied the Student's *t*-test to assess

the significance of the difference between the average DMS reactivity for different codon positions ($P < 0.01$ as significant) (Extended Data Fig. 5c). The same methodology was applied to the high and low translation efficiency mRNA subsets (Fig. 3a and Extended Data Fig. 5d).

Alternative polyadenylation structural patterns. Alternative polyadenylation sites were defined on the basis of a previous genome-wide study of alternative polyadenylation in *A. thaliana*[22]. First we computed and plotted the nucleotide occurrence 50 nt upstream and 50 nt downstream of the alternative polyadenylation cleavage site for all alternatively polyadenylated mRNAs represented in our RNA structurome (Extended Data Fig. 6a). There were 5,959 mRNAs in our data set with alternative polyadenylation cleavage sites. Then we mapped the average DMS reactivity of these upstream and downstream regions. We applied the Student's *t*-test to analyse the significance of the difference in the average DMS reactivity between the structured region (−15 nt to −2 nt uracil- and adenine- rich region upstream of the alternative polyadenylation cleavage sites) and the average DMS reactivity of the whole 100 nt (Fig. 3b). We also did the same analysis for the significance of the difference in the average DMS reactivity between the unstructured region (−1 nt to 5 nt adenine-rich region of the alternative polyadenylation cleavage sites) and the average DMS reactivity of the whole 100 nt (Fig. 3b).

Structure across alternative splice sites. On the basis of a previous study of genome-wide alternative splicing in *Arabidopsis* seedlings[26], we identified, for each mRNA in our data set, whether all introns were spliced out or whether alternative splicing (including exon skipping and intron retention) occurred. This yielded a data set of 15,441 mRNAs with alternative splicing events. We then examined average DMS reactivity of the 100 nt at the 3′ end of the 5′ exon and compared this parameter in unspliced versus spliced events (Fig. 3c). For the unspliced events, we applied the Student's *t*-test to analyse the significance of the difference in the average DMS reactivity between the 40 nt upstream of the 5′ splice site and the remaining 60 nt of the 100-nt region upstream of the 5′ splice site. The same analysis was performed for the spliced events. As a nucleotide composition control for the unspliced events, the identical nucleotide composition of the 40 nt upstream of the 5′ splice site in the unspliced events was shuffled and remapped to find regions on the mRNAs in the TAIR *Arabidopsis* cDNA library that were not located at 5′ splice site junctions. For all of the resulting regions that were also present in our data set, the average DMS reactivity for each nucleotide along the 40-nt regions plus the additional 60 nt upstream of these regions was collected as a total 100-nt control, and the resulting average DMS reactivity was compared to that of the unspliced events (Fig. 3c). The above set of analyses was also applied to the 100-nt regions of the 3′ splice site except that the nucleotide composition control was performed with a 100-nt shuffle (Extended Data Fig. 6d).

All global structure trends in mRNA regions and periodicity, alternative polyadenylation and alternative splicing that we describe (Fig. 3 and Extended Data Figs 5 and 6) remained significant when global analyses were redone on the smaller, 10,623 mRNA subset with ≥ 1 average reverse transcriptase stop per (A+C) nucleotide.

Comparison between *in vivo* constrained RNA structures and *in silico* predicted RNA structures. *In vivo* DMS-constrained RNA structuromes were graphed with nucleotide resolution. A total of 10,623 mRNAs with ≥ 1 average reverse transcriptase stop per (A+C) nucleotide were analysed (see below). We used the criterion of ≥ 1 average reverse transcriptase stop per (A+C) nucleotide because PPV and Gene Ontology analyses rely on nucleotide resolution throughout the entire mRNA. All 10,623 mRNAs (including all splice variants) were aligned by their start codon. Colour scales were applied to indicate the DMS reactivity. Each row in Fig. 4a represents the DMS-guided RNA structurome information of one mRNA. mRNAs were organized by transcript length. The figure was constructed using Python matplotlib module (http://matplotlib.org/).

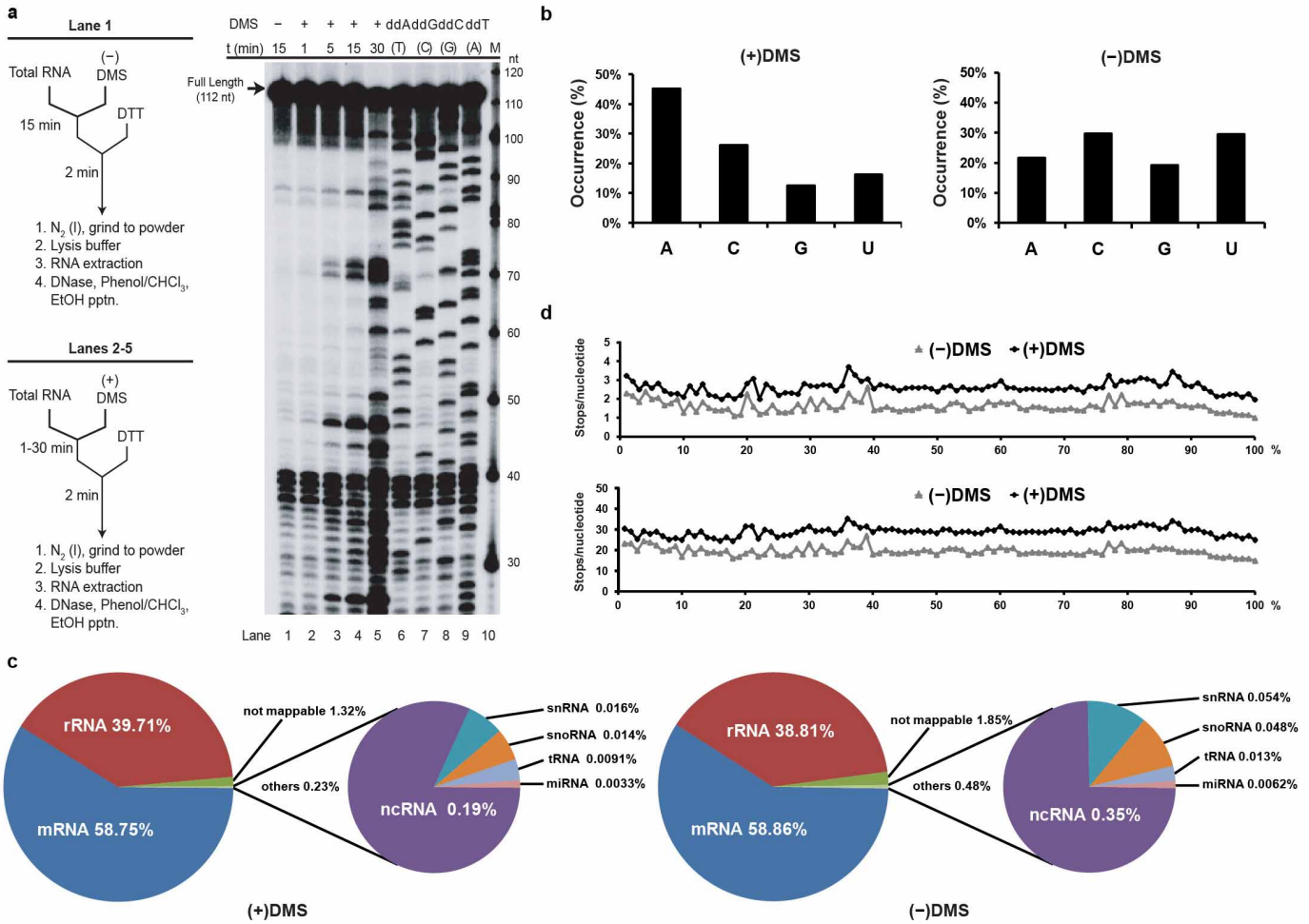
To obtain predicted RNA structures, we folded each of the 10,623 *A. thaliana* mRNAs with ≥ 1 average reverse transcriptase stop per (A+C) nucleotide using the program RNAstructure[27] (http://rna.urmc.rochester.edu/RNAstructure.html) with slope (1.8) and intercept (−0.6) for the pseudo-free energy function and either with or without our *in vivo* DMS constraints. (After testing on several protein-free regions of 18S rRNA, we concluded that for the pseudo-free energy function used by RNAstructure[27] the intercept and slope as defined in ref. 33 were adequate.) We compared *in vivo* DMS-constrained RNA structure with *in silico*-predicted RNA structure (that is, without constraints) for each mRNA by examining the PPV and sensitivity of base pairs[28]. Simply, when comparing two structures, PPV implies the proportion of base pairs in the *in vivo* DMS-constrained RNA structure that also appear in the *in silico*-predicted RNA structure[28]. The sensitivity indicates the proportion of base pair coverage *in silico* that also appears *in vivo*[28]. These criteria indicate the extent of divergence of *in vivo* constrained and *in silico* structures[28]. In our data, the PPV and sensitivity for the mRNA population are highly correlated (PCC = 0.99), thus we use PPV to represent the difference between the *in vivo* and *in silico* structures (Fig. 4b). Negative predictive value (NPV)[34] implies the proportion of single-stranded nucleotides common to both structures. The PPV and NPV are also highly correlated in our data set (PCC = 0.90), and so PPV was used for subsequent analyses. We plotted the PPV values for each transcript across the 10,623 mRNAs (Fig. 4b). We then took the mRNAs with PPV values in the top 5% and those with PPV values in the bottom 5% and performed Gene Ontology annotation analysis[29] for these two groups using the hypergeometric test ($P < 0.01$ as significant) (Fig. 4c). For Gene Ontology analysis of mRNAs with splice variants, we defined the PPV value as the average of the PPV values of all the splice variants of that mRNA present in our data set. Structure prediction with inclusion of pseudoknot prediction was performed for the top 1% and bottom 1% of mRNAs in the PPV distribution using RNAstructure (ShapeKnots command)[33,35].

Comparison of high and low PPV mRNAs. To better understand the underlying mechanisms causing the variation of PPV among mRNAs, we selected the mRNAs in the top 5% and bottom 5% of the PPV distribution and performed two sample *t*-tests to assess whether there was significant difference between the two groups for several RNA structural features for both *in silico* structures and *in vivo* structures: single-strand percentage, maximum loop length and free energy per nucleotide within an mRNA, and DMS reactivity per nucleotide. We similarly compared the prevalence of pseudoknots (pseudoknots per 100 nt of structure) in the top and bottom 1% of the PPV distribution.

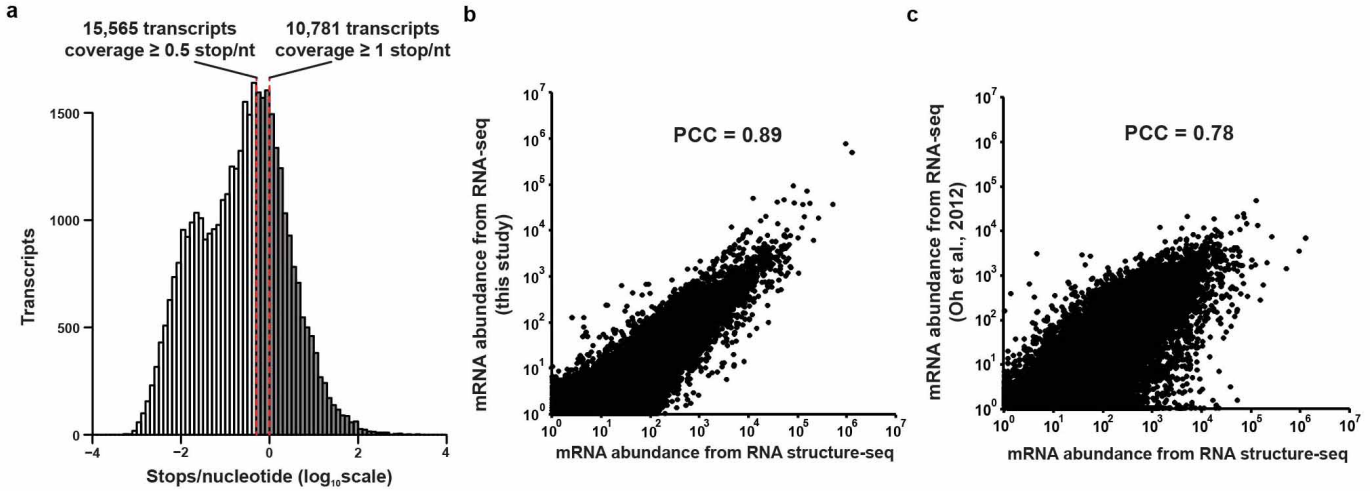
Gel-based method data collection and quantification. The gel-based method of structure probing used the same *in vivo* total RNA pools from the same (+)DMS and (−)DMS plant material as for high-throughput RNA structure-seq. To accomplish gel-based structure probing, reverse transcription was performed using gene-specific ^{32}P-radiolabelled DNA primers (18S reverse primer for region 1 for gel-based method, 5′- AACTGATTTAATGAGCCATTCGCAG-3′; 18S reverse primer for region 2 for gel-based method, 5′-GAGCCCGCGTCGACCTTTTATC-3′; 18S reverse primer for region 3 for gel-based method, 5′-GGTAATTTGCGCG CCTGCT-3′; *CAB1* mRNA (At1g29930) reverse outer primer for gel-based method, 5′-TTCCAAGGACTTCAGATGCC-3′; *CAB1* mRNA (At1g29930) reverse inner primer for gel-based method, 5′-GGAAAGCTTGACGGCCTTAC-3′; ssDNA adaptor for gel-based method, 5′-pNNNCTGCTGATCACCGACTGCCCATAG AG-3′−Spacer; adaptor forward primer for gel-based method, 5′-CTCTATGGG CAGTCGGTGAT-3′). The cDNA samples were then size fractionated on 8.3 M urea 8% polyacrylamide gels for DNA size separation. The power was maintained at 90–100 W throughout the 1.5–2 h run, and the surface temperature was ~55–65 °C, which helps to ensure denaturation of the DNA. Each gel was dried and exposed using a PhosphorImager (Molecular Dynamics) cassette.

Gel images were collected with a Typhoon PhosphorImager 9410, and bands were quantified using ImageQuant 5.2. The differences in band intensity between (+)DMS and (−)DMS samples were calculated. The most intense peak was normalized as 100% intensity[5]. As DMS specifically targets the Watson–Crick position of A and C nucleotides, the G and U nucleotides were not included during signal processing.

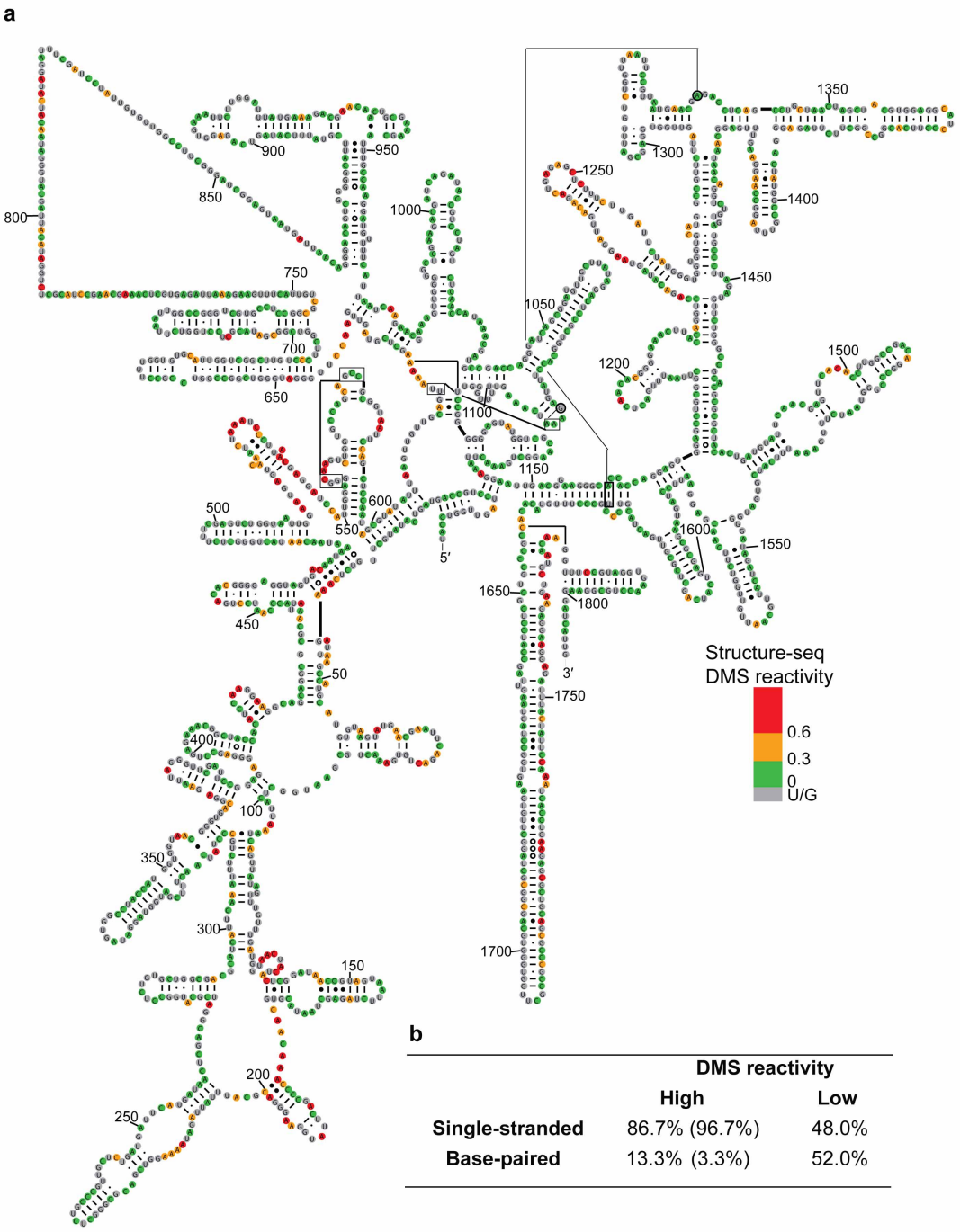
31. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl Acad. Sci. USA* **108,** 11063–11068 (2011).
32. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
33. Hajdin, C. E. *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl Acad. Sci. USA* **110,** 5498–5503 (2013).
34. Smith, C. J. Diagnostic tests (2) – positive and negative predictive values. *Phlebology* **27,** 305–306 (2012).
35. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11,** 129 (2010).
36. Lawley, P. D. & Brookes, P. Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem. J.* **89,** 127–138 (1963).
37. Weeks, K. M. & Crothers, D. M. RNA recognition by Tat-derived peptides: interaction in the major groove? *Cell* **66,** 577–588 (1991).



Extended Data Figure 1 | Time course of DMS modification and overview of structure-seq libraries. **a**, Time course of *in vivo* DMS modification of 18S rRNA in *Arabidopsis* etiolated seedlings. Five-day-old *Arabidopsis* etiolated seedlings were DMS-treated for different durations (1 min, 5 min, 15 min and 30 min; lanes 2–5, respectively). In all cases the final DMS concentration was 0.75% (~75 mM). The 18S rRNA DMS modification read-out was assessed by gel-based probing, which was done here near the 5′ end to afford a view of the full-length RNA band. The 15-min time point is the optimal duration for DMS modification, as it is the longest time point for which single-hit kinetics still occur as revealed by the intense full-length band. The 30-min time point is too long, as revealed by significant loss of the full-length band and increase of shorter length bands. Lanes 6–9 show the dideoxy sequencing of 18S rRNA. Lane 1 is the (−)DMS control. Lane 10 is a DNA marker (M) that was size fractionated to confirm the size of the full-length band (112 nt). **b**, DMS modification is RNA nucleotide specific. Nucleotide occurrence of RNA bases one nucleotide upstream of the position of reverse transcriptase stalling on the (+)DMS library and (−)DMS library, respectively. The (+)DMS library shows higher occurrence of A and C than of U and G (A is more than 1 standard deviation higher compared to C, G and U, and C is more than 1 standard deviation higher compared to G and U if leaving out A), consistent with the properties of DMS modification of nucleobases[36]. The percentages of each RNA base in the (−)DMS library are also indicated and are found to be similar (within 1 standard deviation). This figure combines results from both biological replicates. **c**, The total number of reads was classified into different classes of RNAs on a percentage basis from a total number of 121,258,873 reads for the (+)DMS library and 85,371,519 reads for the (−)DMS library. This figure combines results from both biological replicates. **d**, Structure-seq reads coverage. RNA structure information from structure-seq is distributed evenly across transcripts, with no 3′ bias. Each of the 37,558 transcripts (all transcripts with ≥ 1 internal reverse transcriptase stop and length ≥ 100 nt) was divided into 100 bins to normalize the transcript length. The reverse transcriptase stops per each A and C nucleotide (top) and the reverse transcriptase stops per each A and C nucleotide with ≥ 1 reverse transcriptase stop (bottom) from both the (+)DMS library (black diamonds) and the (−)DMS library (grey triangles) were averaged within each bin and plotted. The reverse transcriptase stops are well distributed over the entire transcript length.

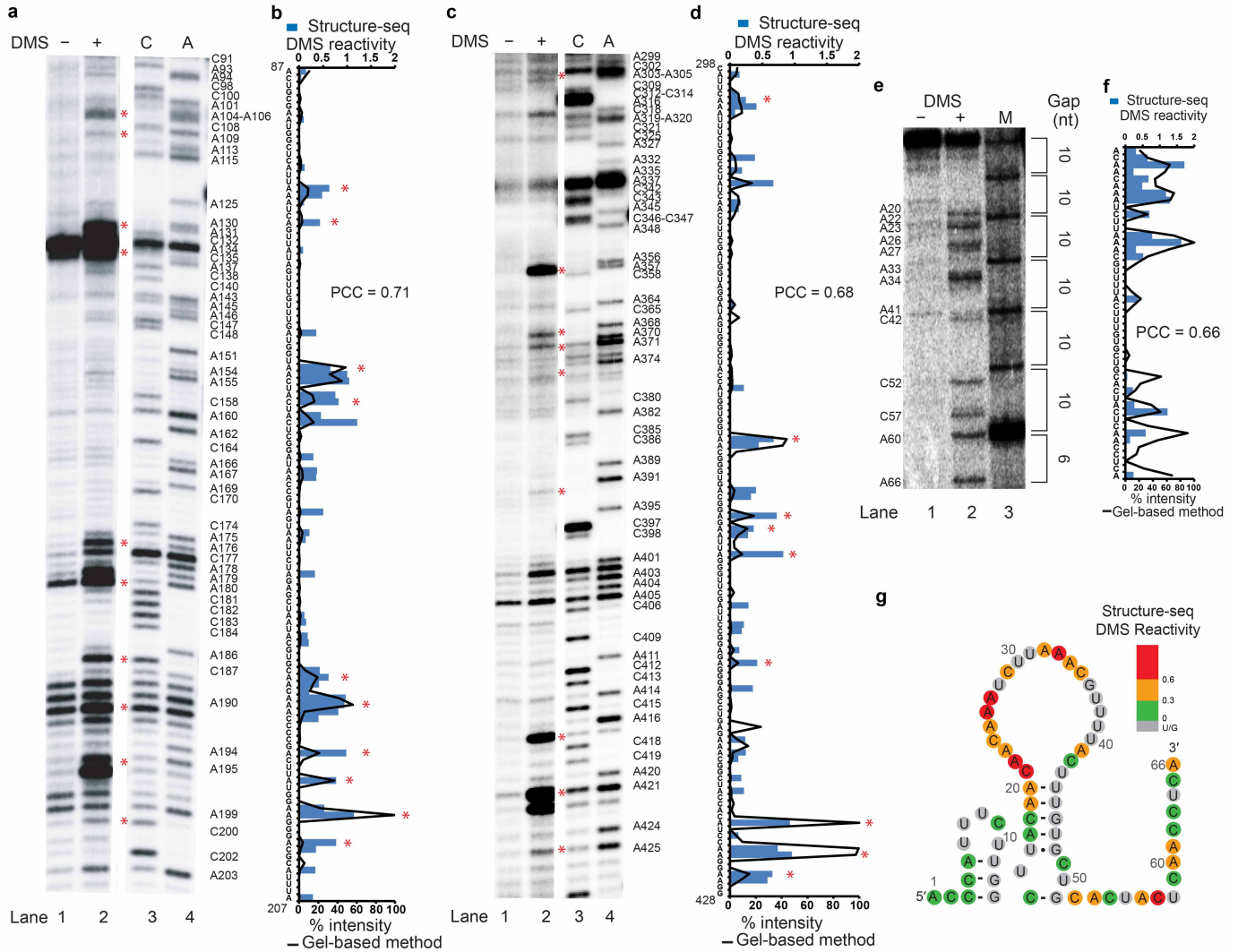


Extended Data Figure 2 | Structure-seq reveals *in vivo* RNA secondary structures for over 10,000 transcripts and correlates with mRNA abundance. **a**, Structure-seq reveals *in vivo* RNA secondary structures for over 10,000 transcripts. The histogram shows the number of transcripts as a function of the average reverse transcriptase stops associated with A + C nucleotides of a transcript, divided by the total number of the A + C nucleotides of that transcript, calculated for all individual transcripts in our data set. (Note that it is expected that not all As and Cs of a transcript will be DMS-modified and associated with an reverse transcriptase stop, because some As and Cs will be protected, for example, by base-pairing, tertiary structure or protein binding.) There are 10,781 transcripts with ≥ 1 average read per A + C nucleotides (dark shading and to the right of the right-most dashed red line). With a threshold of 0.5 average reads per A + C nucleotides, there are 15,565 transcripts (to the right of the left-most dashed red line). It is of interest to compare structure-seq, which provides the first high-throughput *in vivo* RNA structurome, with previous high-throughput studies of RNA structures conducted *in vitro*[5–8]. We have coverage with ≥ 1 average reverse transcriptase stop per nucleotide across 10,623 mRNAs, which compares favourably with ~3,000 mRNAs with load (number of reads per nucleotide) > 1 from an *in vitro* study of yeast[5]. In comparison with 3.9×10^5 reads (0.0078 RNase One cleavages per nucleotide on average) on mRNAs in the single-stranded RNA-seq library of an *in vitro* study of RNA structure in *Arabidopsis*[6], we have much improved coverage with 7.1×10^7 reads (1.4 reverse transcriptase stops per nucleotide on average) on mRNAs in our (+)DMS *in vivo* library. **b**, **c**, Structure-seq queries *in vivo* RNA structures in proportion to their abundance in the transcriptome. mRNA abundance within our structure-seq data set is highly correlated with mRNA abundance from RNA-seq analysis in this study (**b**) and with RNA-seq analysis from a previous study (**c**)[14]. Correlation of mRNA abundance is based on average sequencing reads per mRNA between structure-seq and RNA-seq. The RNA-seq data set in our study was generated in parallel with the structure-seq data set from seedlings under the identical growth conditions but without DMS; that is, the RNA-seq data are extracted from the (−)DMS library. The RNA-seq data set from ref. 14 was generated from five-day-old etiolated seedlings. The PCCs of 0.89 and 0.78, respectively, indicate that more abundant mRNAs are more likely to have sufficient coverage available for structure-seq analysis.

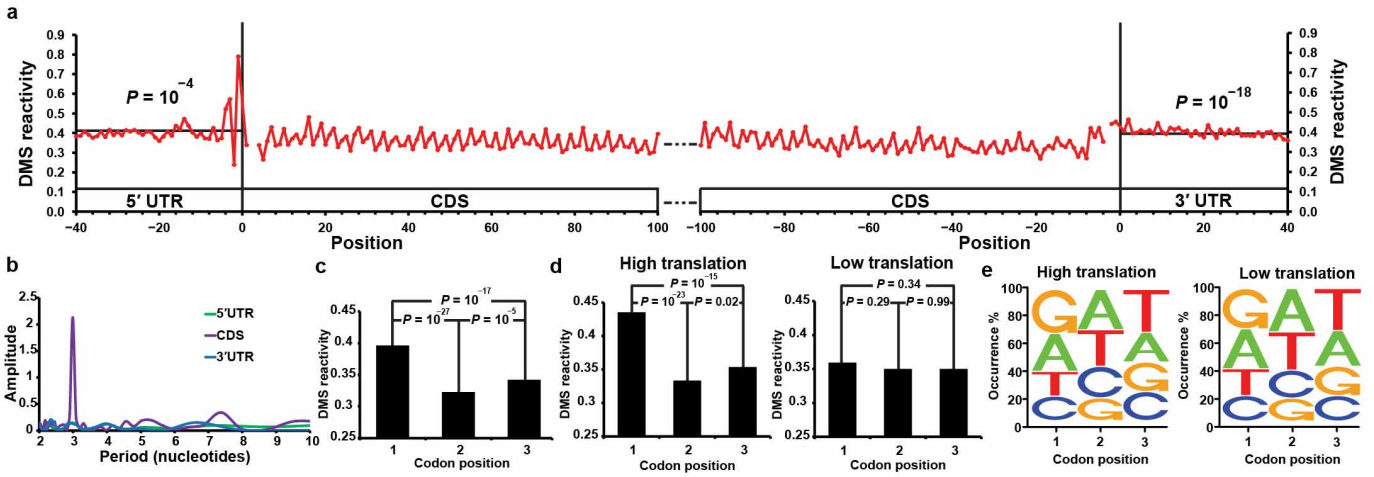


	DMS reactivity	
	High	Low
Single-stranded	86.7% (96.7%)	48.0%
Base-paired	13.3% (3.3%)	52.0%

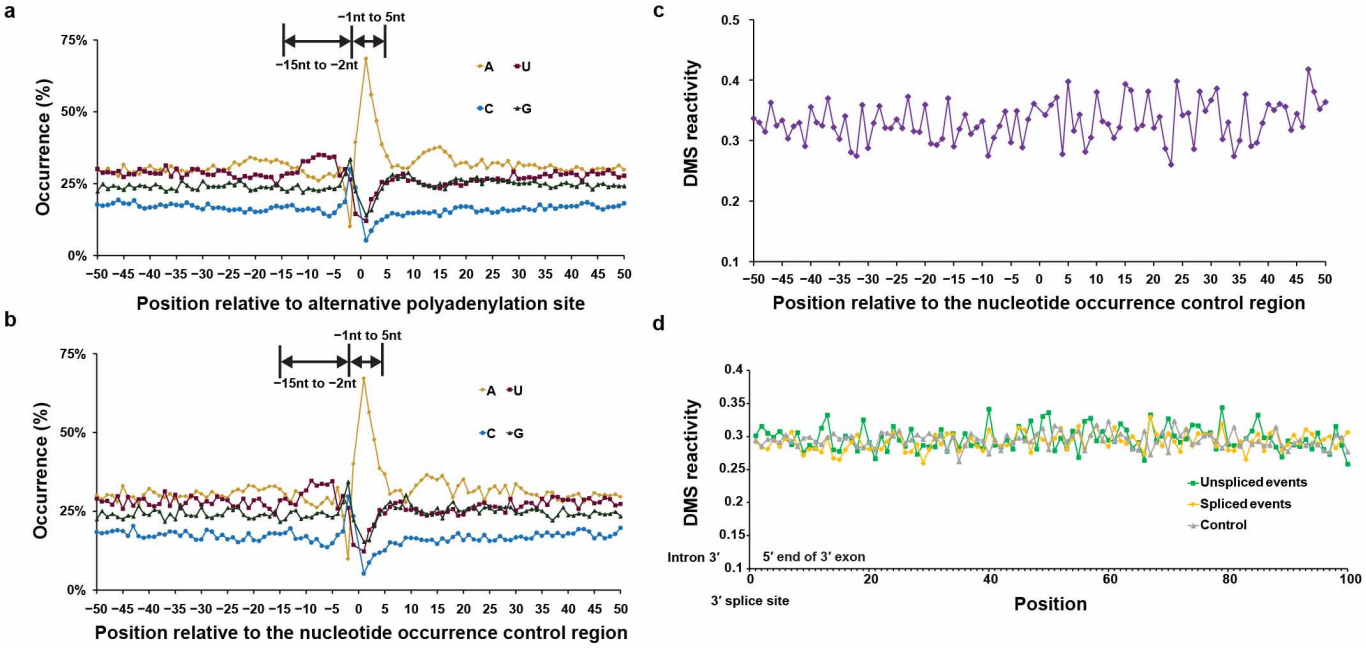
Extended Data Figure 3 | Structure-seq provides the complete map of the 18S rRNA *in vivo* structure at nucleotide resolution. **a**, Structure-seq provides the complete map of the 18S rRNA *in vivo* structure at nucleotide resolution. The complete 18S rRNA phylogenetic structure[16] is colour-coded according to the DMS reactivity generated from structure-seq (DMS reactivity ≥0.6 marked in red; DMS reactivity 0.3–0.6 marked in yellow; DMS reactivity 0–0.3 marked in green; and U/G bases marked in grey). **b**, High correlation between structure-seq and 18S rRNA phylogenetic structure. In the entire 18S rRNA (length = 1,808 nt), 86.7% (true positive) of the As and Cs that show high *in vivo* DMS reactivity (defined as ≥0.6) in our data set correspond to single-stranded regions in the phylogenetic structure[16], whereas 52.0% (true negative) of the As and Cs that show low *in vivo* DMS reactivity (defined as ≤ 0.3) in our data set correspond to base-paired regions in the phylogenetic structure. The 48.0% (false negative) of the As and Cs that show low *in vivo* DMS reactivity in our data set but correspond to single-stranded regions in the phylogenetic structure presumably are protected by either ribosomal proteins or non-base-pairing tertiary RNA structure. Of the 13.3% (false positive) reactive nucleotides (defined as ≥ 0.6 from structure-seq) that are annotated as base-paired in the phylogenetic structure, 75% of these nucleotides are positioned either at the end of a helix or adjacent to a helical defect such as a bulge or loop, locations that are known to lead to flexibility[37]. Values in parentheses, corrected for this positioning, show higher true positive and lower false positive percentages.



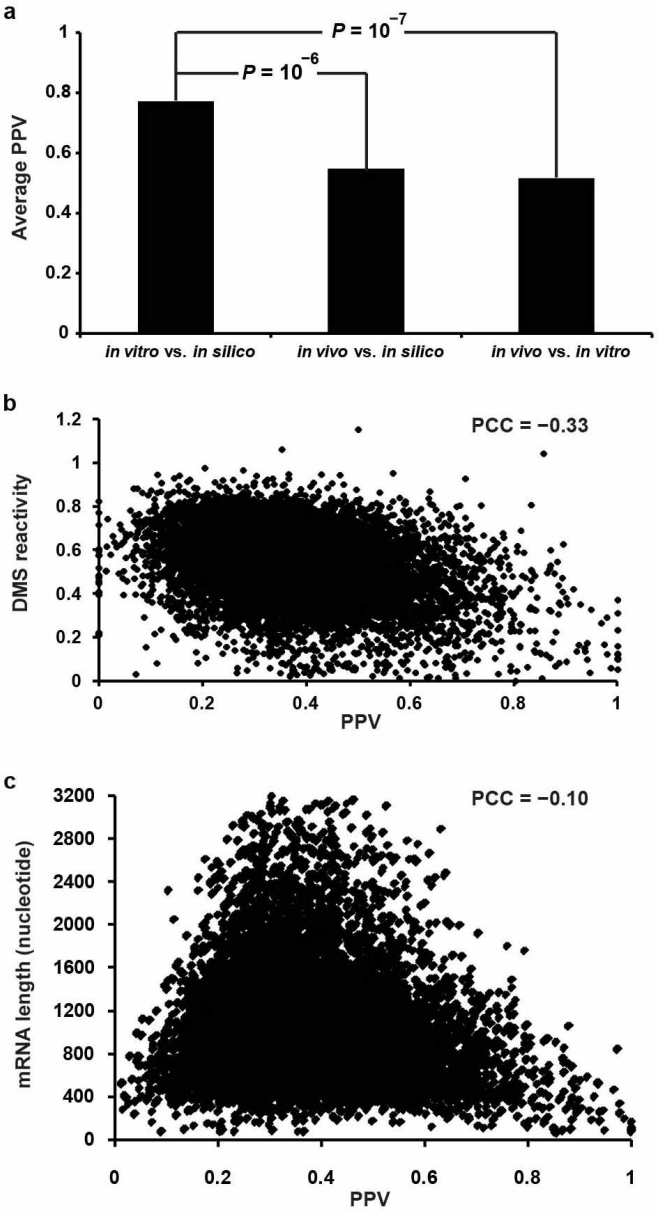
Extended Data Figure 4 | Structure-seq results are strongly correlated with results from the conventional gel-based RNA structure probing method. **a**, Nucleotides 87–207 of 18S rRNA were probed by the conventional gel-based method. Lanes 1–2 show the (−)DMS and (+)DMS results on the region of interest. Lanes 3–4 show C and A dideoxy sequencing. For both this panel and structure-seq, the starting material was the same total population of *in vivo* DMS-modified RNA. **b**, The results from structure-seq (blue bars) are compared to results from the conventional gel-based method, presented as normalized band intensity (black lines), with the highest intensity normalized to 100%[5]. The red asterisks indicate nucleotides that have significant DMS modifications from both methods, and are also shown in panel **a**. Structure-seq results are strongly correlated with results from the conventional gel-based RNA structure probing method: the PCC between the two methods is 0.71. **c**, **d**, Nucleotides 298–428 of 18S rRNA as probed by structure-seq and also analysed by the conventional gel-based method. The PCC is 0.68. **e**–**g**, Structure-seq results are also strongly correlated with results from the conventional gel-based RNA structure probing method for an individual mRNA, *CAB1* (At1g29930). The 5′ UTR of *CAB1* was probed by structure-seq and analysed by the gel-based method; in both cases, the starting material was the same total population of *in vivo* DMS-modified RNA. **e**, Lanes 1–2 show the (−)DMS and (+)DMS results on the region of interest as analysed by the conventional gel-based method. A 10-nt marker (M) was size fractionated (lane 3) to allow nucleotide assignment based on spacing. **f**, DMS reactivity from structure-seq is plotted with nucleotide resolution (blue bars). Results from the gel-based RNA structure probing method are presented as normalized quantified band intensity (black lines), with the highest intensity normalized to 100%[5]. For the gel-based method, the nucleotides near the 5′ end cannot be confidently quantified and assigned due to band compression at the top of the gel and proximity to the full-length band. The PCC between the two methods is 0.66. **g**, The secondary structure of the 5′ UTR of *CAB1* mRNA (At1g29930) was determined using the i*n vivo* DMS constraints obtained from structure-seq. (DMS reactivity ≥ 0.6 marked in red; DMS reactivity 0.3–0.6 marked in yellow; DMS reactivity 0–0.3 marked in green; and U/G bases marked in grey).



Extended Data Figure 5 | Structure-seq reveals global trends in mRNA secondary structure *in vivo* that correlate with translation efficiency. **a**, Average DMS reactivity on an A + C nucleotide basis in selected regions of 22,721 mRNAs (including all splice variants) that have 5′ and 3′ UTR regions longer than 40 nt: 5′ UTR region (40 nt upstream of the start codon); CDS initial region (100 nt downstream of the start codon); CDS final region (100 nt upstream of the stop codon); and 3′ UTR region (40 nt downstream of the stop codon) are depicted. The transcripts were aligned by their start codon and stop codon (vertical lines). (Us and Gs in the start codon and the stop codon were not counted, marked by a break in the red line.) The 40-nt 5′ UTR and 3′ UTR regions show significantly higher average DMS reactivity than the flanking 100 nt of the CDS region, with *P* values of 10^{-4} and 10^{-18}, respectively (Student's *t*-tests). The first 5 nt immediately upstream of the start codon show significantly higher reactivity than the average DMS reactivity across the first 100 nt of the CDS with *P* value of 10^{-112} (Student's *t*-test). **b**, Discrete Fourier transform of average DMS reactivity on a nucleotide basis was performed on the 40-nt 5′ UTR (green line), the first 100 nt of the CDS (purple line) and the 40-nt 3′ UTR (blue line) regions. Only the CDS shows the periodic signal. For the analysis, the 40-nt 5′ UTRs and 3′ UTRs were compared to the first 100 nt of the CDS regions. **c**, The average DMS reactivity of the three positions in each codon was computed from the entire CDS regions of all 22,721 mRNAs. The first position of each codon shows significantly higher average DMS reactivity compared with the second position of each codon ($P = 10^{-27}$). The third position of each codon shows significantly higher average DMS reactivity compared with the second position ($P = 10^{-5}$) but significantly lower average DMS reactivity compared with the first position of each codon ($P = 10^{-17}$) (Student's *t*-tests). **d**, Structure-seq reveals significantly stronger periodic signal in the coding regions of high translation efficiency mRNAs (1,136 mRNAs) as compared to low translation efficiency mRNAs (1,136 mRNAs). We analysed the polyribosome-associated mRNA populations defined in a previous study[21], ranking the mRNAs according to their polyribosome-associated mRNA abundance[21]. We defined the top 5% ($n = 1{,}136$ mRNAs) as the 'high translation efficiency mRNAs' and the bottom 5% ($n = 1{,}136$ mRNAs) as the 'low translation efficiency mRNAs'. The average DMS reactivity of the three positions of each codon was computed along the entire CDS for the high translation efficiency mRNAs and the low translation efficiency mRNAs. The difference in average DMS reactivity between the three nucleotides is significantly greater in the high translation efficiency transcripts (nt 1–2, $P = 10^{-23}$; nt 2–3, $P = 0.02$; nt 1–3, $P = 10^{-15}$) than in the low translation efficiency transcripts (nt 1–2, $P = 0.29$; nt 2–3, $P = 0.99$; nt 1–3, $P = 0.34$) (Student's *t*-tests). **e**, No nucleotide or codon bias in high versus low translation efficiency mRNAs occurs in any of the three positions of the codon. There is no difference between high translation efficiency mRNAs (1,136 mRNAs) and low translation efficiency mRNAs (1,136 mRNAs) in the frequency of nucleotide occurrence at each codon position. The correlation between the codon usage of the high translation efficiency mRNAs and low translation efficiency mRNAs is very high (PCC = 0.90).

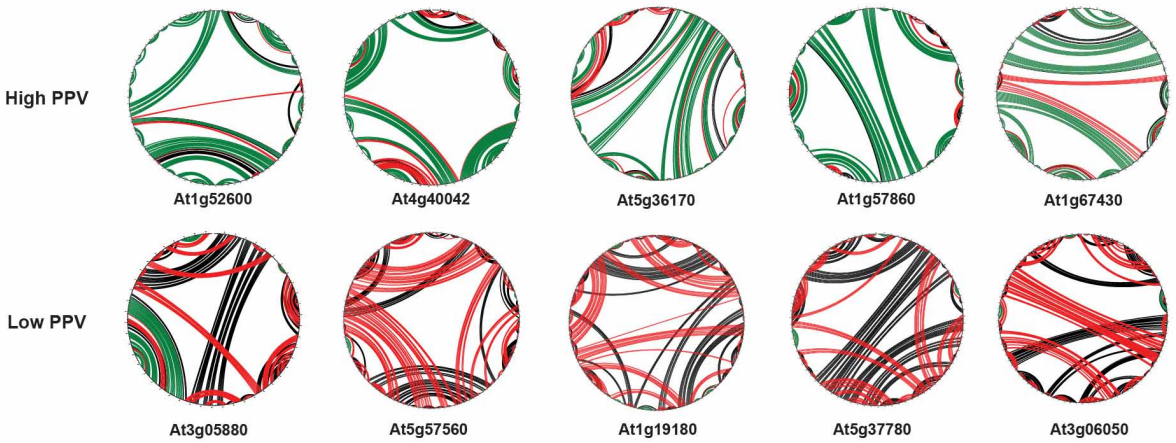


Extended Data Figure 6 | Control analyses for alternative polyadenylation and alternative splicing. **a**, The percentages of nucleotide occurrence around the site of alternative polyadenylation show a U/A rich region from -15 nt to -2 nt ($P = 10^{-16}$ Student's *t*-test), and the region from 1 nt upstream to 5 nt downstream (nt -1 to 5) of the cleavage site is A-rich ($P = 10^{-5}$ Student's *t*-test). This pattern is not unlike that reported for a combined data set of all polyadenylation sites[23]. The percentages of nucleotide occurrence are plotted relative to the alternative polyadenylation site position collected from a previous study[22], indicated by 0: (A (orange diamonds); U (dark red squares); C (blue circles); and G (green triangles)). **b**-**c**, Nucleotide composition and sequence alone cannot account for the RNA structural pattern of the alternative polyadenylation site. **b**, We identified 20 nt regions in our structure-seq mRNA data set that are not alternative polyadenylation cleavage sites but contain the same exact nucleotide sequence as the region 15 nt upstream and 5 nt downstream of each alternative polyadenylation cleavage site that we analysed. The percentages of nucleotide occurrence are plotted relative to the position corresponding to where the alternative polyadenylation site (designated as position zero) would be situated: (A (orange diamonds); U (dark red squares); C (blue circles); and G (green triangles)). **c**, For the selected control region from panel **b**, DMS reactivity of these selected 20 nt control regions as well as the regions upstream (35 nt) and downstream (45 nt) was averaged on a nucleotide basis and plotted, revealing absence of any structural features (violet line). **d**, Extensive RNA secondary structure was not apparent at the 3′ splice site. A previous genome-wide study of alternative splicing (AS) in *Arabidopsis* seedlings[26] was used to identify for each mRNA in our data set, whether all introns were spliced out or whether AS (including exon skipping and intron retention) occurred. DMS reactivity along 100 nt in the exons upstream of the 3′ splice site was averaged on a nucleotide basis from the unspliced events, including both exon skipping and intron retention (green lines), and the spliced events (yellow lines). The same nucleotide composition of the 100 nt in the unspliced AS events was shuffled and remapped to regions in our structure-seq mRNA data set that were not located at the junction of a 3′ splice site. The averaged DMS reactivity collected from the control regions with the same nucleotide composition served as the control (grey lines).



Extended Data Figure 7 | *In vitro* structures differ from *in vivo* structures; PPV does not correlate with average DMS reactivity or with mRNA length. **a**, *In vitro* structures differ from *in vivo* structures, and *in vitro* structures are more similar to *in silico* structures than are *in vivo* structures. The 61 *Arabidopsis* mRNAs with coverage ≥ 0.5 cleavages per nucleotide from Li *et al.*'s *in vitro* data were compared among the *in silico* structure (from RNAstructure), the *in vitro* structure (*in silico* structures from RNAstructure constrained by Li *et al.*'s *in vitro* data)[6], and the *in vivo* structure (*in silico* structures from RNAstructure constrained by our *in vivo* data). PPV (the base pairs in one structure that are also present in another structure, as a proportion) was averaged across these 61 mRNAs. The PPV between *in vitro* structures and *in silico* structures is 0.77, which is significantly higher than the PPV between *in vivo* structures and *in silico* structures and is also significantly higher than the PPV between *in vivo* and *in vitro* structures, according to two sample *t*-tests with *P* values as shown in the figure. *In vivo* structures are different from both *in vitro* structures (PPV = 0.51) and *in silico* structures (PPV = 0.55). **b**, PPV does not correlate with average DMS reactivity per nucleotide. For each of 10,623 mRNAs in our structure-seq data set, the corresponding PPV of each mRNA was plotted, revealing an absence of correlation between PPV and average DMS reactivity per nucleotide (PCC = −0.33). **c**, PPV does not correlate with mRNA length. For each of 10,623 mRNAs, the corresponding PPV of each mRNA was plotted as a function of mRNA length, revealing an absence of correlation between these two variables (PCC = −0.10).

a



b

High PPV

ID	Annotation	PPV	Single strandedness (*in silico*)	Single strandedness (*in vivo*)	Max. loop length (*in silico*)	Max. loop length (*in vivo*)	Free energy per nucleotide (*in silico*)	Free energy per nucleotide (*in vivo*)
At1g52600	Peptidase S24/S26	0.87	0.35	0.35	11	11	−0.26	−0.46
At4g40042	SPC12 Microsomal signal peptidase 12 kDa subunit	0.78	0.35	0.34	12	12	−0.26	−0.51
At5g36170	*ATPRFB* Required for normal processing of polycistronic plastidial transcripts	0.78	0.37	0.34	15	13	−0.28	−0.53
At1g57860	Translation protein SH3-like family protein	0.77	0.37	0.40	8	7	−0.28	−0.45
At1g67430	Ribosomal protein L22p/L17e	0.78	0.39	0.39	10	10	−0.27	−0.39

Low PPV

ID	Annotation	PPV	Single strandedness (*in silico*)	Single strandedness (*in vivo*)	Max. loop length (*in silico*)	Max. loop length (*in vivo*)	Free energy per nucleotide (*in silico*)	Free energy per nucleotide (*in vivo*)
At3g05880	*RCI2A* (*RARE COLD-INDUCIBLE 2A*)	0.19	0.37	0.42	11	12	−0.21	−0.31
At5g57560	*TCH4* (*TOUCH 4*), endotransglucosylase/ hydrolase rapidly upregulated in response to environmental stimuli	0.07	0.39	0.43	11	39	−0.25	−0.40
At1g19180	*JAZ1* (*JASMONATE-ZIM-DOMAIN PROTEIN 1*), involved in jasmonate signaling	0.15	0.38	0.41	17	21	−0.22	−0.39
At5g37780	*CAM1* (*CALMODULIN 1*), detection of calcium ion	0.09	0.38	0.43	9	11	−0.24	−0.33
At3g06050	*ATPRXIIF*, *PEROXIREDOXIN IIF* involved in redox homeostasis under oxidative stress	0.09	0.37	0.41	13	19	−0.29	−0.41

Extended Data Figure 8 | Examples for *in vivo* and *in silico* structural feature comparison of high and low PPV mRNAs. **a**, Ten examples for *in vivo* and *in silico* structural comparison of high and low PPV mRNAs. Five examples from the high PPV mRNA group (top) and five examples from the low PPV mRNA group (bottom). At1g52600 and At3g05880 mRNA structures were given in Fig. 4d. Base pair predictions are indicated with coloured lines: red, uniquely *in vivo* base pair; black, uniquely *in silico* base pair; green, base pair present in both the *in vivo* and the *in silico* structure. Plots were generated using the CircleCompare program in the RNAstructure package[35]. Low PPV mRNAs show more extensive differences between *in vivo* and *in silico* structures than do high PPV mRNAs. **b**, Characteristics of *in vivo* and *in silico* structural features in the ten high and low PPV mRNAs. The same five examples from both high PPV and low PPV mRNAs as in **a** were assessed for RNA structural features in both *in silico*-predicted (without *in vivo* constraints) and *in vivo* (*in silico* prediction with constraints from our *in vivo* structure-seq data) structures. *In vivo* structures of low PPV mRNAs show more single stranded regions, longer maximum loop length, and higher (that is, less favourable) free energy per nucleotide as compared to high PPV mRNAs. By contrast, *in silico*-predicted structures do not show such major differences between low and high PPV mRNAs.

Extended Data Table 1 | Statistical analysis of structure-seq libraries

a

	Between Two Biological Replicates		Biological Replicate I	Biological Replicate II
Library	(+)DMS/(+)DMS	(−)DMS/(−)DMS	(−)DMS/(+)DMS	(−)DMS/(+)DMS
Correlation	0.91	0.74	0.49	0.61

b

Library	Total reads	uniquely mapped reads	uniquely mapped %	not mappable reads	not mappable %
(−)DMS Biological Replicate 1	3.93×10^7	3.86×10^7	98.24	6.94×10^5	1.76
(+)DMS Biological Replicate 1	3.25×10^7	3.22×10^7	98.88	3.65×10^5	1.12
(−)DMS Biological Replicate 2	4.60×10^7	4.52×10^7	98.07	8.87×10^5	1.93
(+)DMS Biological Replicate 2	8.87×10^7	8.75×10^7	98.61	1.23×10^6	1.39

c

RNA type	(+)DMS library (reads)	(+)DMS library (% reads)	(−)DMS library (reads)	(−)DMS library (% reads)
mRNA	7.12×10^7	58.75	5.03×10^7	58.86
rRNA	4.81×10^7	39.71	3.31×10^7	38.81
ncRNA	2.29×10^5	0.19	3.02×10^5	0.35
snRNA	1.95×10^4	0.016	4.57×10^4	0.054
tRNA	1.11×10^4	0.0091	1.15×10^4	0.013
miRNA	3.96×10^3	0.0033	5.31×10^3	0.0062
snoRNA	1.68×10^4	0.014	4.12×10^4	0.048
Total	1.21×10^8	100	8.53×10^7	100

a, High correlation (PCC) between biological replicates for (+) and (−)DMS libraries, and low correlation between the (+)DMS and (−)DMS libraries for each biological replicate. **b**, High read number and mappability of our (+)DMS and (−)DMS libraries. **c**, mRNAs and rRNAs predominate among different classes of RNAs in (+)DMS and (−)DMS libraries (combined data from two biological replicates).

Extended Data Table 2 | *In vivo* constraints improve the prediction of structure in 18S rRNA

Row		PPV/Sensitivity
1	*in silico* vs. phylogenetic structure	0.27/0.31
2	*in vivo* vs. phylogenetic structure	0.41/0.45
3	*in vivo* vs. phylogenetic structure, omitting false negatives	0.50/0.52
4	ideal A/C constraint vs. phylogenetic structure	0.63/0.63
5	ideal A/C/U/G constraint vs. phylogenetic structure	0.68/0.65
6	*in vivo* vs. *in silico*	0.48/0.46

We calculated the PPV/sensitivity between *in silico* and phylogenetic structure, *in vivo* and phylogenetic structure, and *in vivo* and *in silico* structure in 18S rRNA. (Sensitivity is defined as the proportion of base pairs occurring *in silico* that also appear *in vivo*.) We also compared the *in vivo* structure with the phylogenetic structure upon omission of false negatives (i.e., we did not apply a pseudo-free energy constraint to the false negative data), because false negatives presumably result from protection by either ribosomal proteins or non-base-pairing tertiary RNA structure rather than base pairing. In addition, we folded the RNAs with the constraints generated from ideal A/C or ideal A/C/U/G base-pairing information (the predicted structure with the A/C or A/C/U/G constraints as generated directly from the phylogenetic structure), and compared the resultant structure predictions with actual phylogenetic structures.

LETTER

doi:10.1038/nature12894

Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*

Silvi Rouskin[1], Meghan Zubradt[1], Stefan Washietl[2,3,4], Manolis Kellis[2,3,4] & Jonathan S. Weissman[1]

RNA has a dual role as an informational molecule and a direct effector of biological tasks. The latter function is enabled by RNA's ability to adopt complex secondary and tertiary folds and thus has motivated extensive computational[1,2] and experimental[3–8] efforts for determining RNA structures. Existing approaches for evaluating RNA structure have been largely limited to *in vitro* systems, yet the thermodynamic forces which drive RNA folding *in vitro* may not be sufficient to predict stable RNA structures *in vivo*[5]. Indeed, the presence of RNA-binding proteins and ATP-dependent helicases can influence which structures are present inside cells. Here we present an approach for globally monitoring RNA structure in native conditions *in vivo* with single-nucleotide precision. This method is based on *in vivo* modification with dimethyl sulphate (DMS), which reacts with unpaired adenine and cytosine residues[9], followed by deep sequencing to monitor modifications. Our data from yeast and mammalian cells are in excellent agreement with known messenger RNA structures and with the high-resolution crystal structure of the *Saccharomyces cerevisiae* ribosome[10]. Comparison between *in vivo* and *in vitro* data reveals that in rapidly dividing cells there are vastly fewer structured mRNA regions *in vivo* than *in vitro*. Even thermostable RNA structures are often denatured in cells, highlighting the importance of cellular processes in regulating RNA structure. Indeed, analysis of mRNA structure under ATP-depleted conditions in yeast shows that energy-dependent processes strongly contribute to the predominantly unfolded state of mRNAs inside cells. Our studies broadly enable the functional analysis of physiological RNA structures and reveal that, in contrast to the Anfinsen view of protein folding whereby the structure formed is the most thermodynamically favourable, thermodynamics have an incomplete role in determining mRNA structure *in vivo*.

A wide range of chemicals and enzymes have been used to monitor RNA structure[7,11]. We focused on DMS as it enters cells rapidly[9,12] and is a well-established tool for the analysis of RNA structure[13]. DMS is highly reactive with solvent-accessible, unpaired residues but reliably unreactive with bases engaged in Watson–Crick interactions, thus nucleotides that are strongly protected or reactive to DMS can be inferred to be base-paired or unpaired, respectively. We coupled DMS treatment to a massively parallel sequencing readout (DMS-seq) by randomly fragmenting the pool of modified RNAs and size-selecting before 3′ ligation with a specific adaptor oligonucleotide (Fig. 1a). Because DMS modifications at adenine and cytosine residues block reverse transcription[14], we used a second size-selection step to collect and sequence only the prematurely terminated complementary DNA fragments. Sequencing of the fragments reveals the precise site of DMS modification, with the number of reads at each position providing a measure of relative reactivity of that site. The results are highly reproducible and robust against changes in the time of modification or concentration of DMS used (Fig. 1b). The sequencing readout allowed global analysis with a high signal-to-noise ratio—in DMS treated samples, >90% of reads end with an adenine and cytosine, corresponding to false positives for A and C of 7% and 17%, respectively (Fig. 1c). For each experiment, we measured RNA structure both *in vivo* and *in vitro* (that is, refolded RNA in the absence of proteins). We also measured DMS reactivity under denaturing conditions (95 °C) as a control for intrinsic biases in reactivity, library

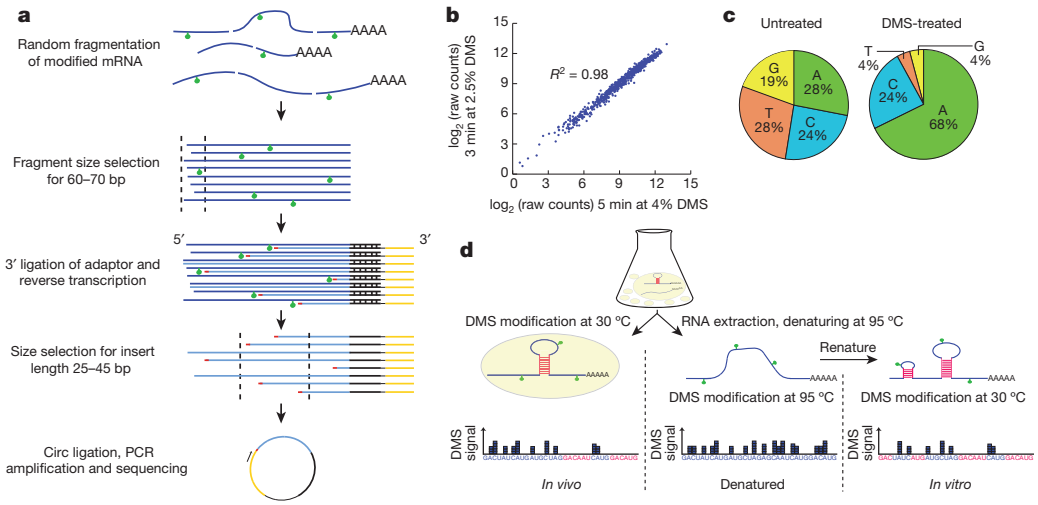


Figure 1 | Using dimethyl sulphate for RNA structure probing by deep sequencing. **a**, Schematic of strategy for library preparation with DMS-modified RNAs. **b**, DMS-seq data are highly reproducible between biological replicates and robust against changes in time and DMS concentration. **c**, *In vivo* DMS treatment markedly enriches for sequencing reads mapping to A/C bases compared to untreated control. **d**, DMS-seq was completed for *in vivo*, denatured and *in vitro* samples. The denatured sample served as an 'unstructured' control.

[1]Department of Cellular and Molecular Pharmacology, California Institute of Quantitative Biology, Center for RNA Systems Biology, Howard Hughes Medical Institute, University of California, San Francisco, California 94158, USA. [2]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]The Broad Institute, Cambridge, Massachusetts 02139, USA.

generation or sequencing, revealing only modest variability compared to that caused by structure-dependent differences in reactivity (Fig. 2c and Extended Data Fig. 1a).

The *in vivo* DMS-seq data are in excellent agreement with known RNA structures. We examined three validated mRNA structures in *S. cerevisiae*: *HAC1*, *RPS28B* and *ASH1*[15–18]. In each case, the DMS-seq pattern qualitatively recapitulates secondary structure with high reactivity constrained to loop regions in both the *in vivo* and the *in vitro* samples, but not in the denatured samples (Fig. 2a, b). Recent determination of a high-resolution yeast 80S ribosome crystal structure[10] allowed us to comprehensively evaluate the DMS-seq data for ribosomal RNAs. Comparison of the 18S (Fig. 2c) and 25S (Extended Data Fig. 1b) rRNA DMS signal *in vivo* versus denatured reveals a large number of strongly protected bases *in vivo*. Based on DMS reactivity, we used a threshold to bin bases into reactive and unreactive groups, then calculated agreement with the crystal structure model as a function of the threshold. True positives were defined as both unpaired and solvent-accessible bases according to the crystal structure, and true negatives defined as paired bases. A receiver operator characteristic (ROC) curve shows a range of thresholds with superb agreement between the *in vivo* DMS-seq data and the crystal structure model (Fig. 2d). For example, at a threshold of 0.2, the true positive rate, false positive rate and accuracy are 90%, 6% and 94%, respectively. Bases that were not reactive at this threshold *in vivo* showed normal reactivity when denatured (Extended Data Fig. 1c). This indicates that the small fraction (~10%) of residues that are designated as accessible, but are nonetheless strongly protected from reacting with DMS, resulted from genuine differences in the *in vivo* conformation of the ribosome and the existing crystal structures. Agreement with the crystal structure was far less good for *in vitro* refolded rRNA (as expected given the absence of ribosomal proteins) and was completely absent for denatured RNA. By contrast, probing of intact purified ribosomes gave a very similar result to that seen *in vivo*, further demonstrating that DMS-seq yields comparable results *in vitro* and *in vivo* when probing the same structure.

Qualitatively, we observed many mRNA regions where structure was apparent *in vitro* but not *in vivo*. For example, computational analysis[19] predicts a stem-loop structure in *RPL33A*. The *in vitro* DMS-seq data strongly supported this predicted structure, whereas this region showed little to no evidence of structure in cells (Fig. 3a). To explore systematically the relationship between mRNA structure *in vivo* and *in vitro*, we quantitated structure in a given region using two metrics: Pearson's correlation coefficient (*r* value), which reports on the degree of similarity of the modification pattern to that of a denatured control, and the Gini index[20], which measures disparity in count distribution as would be seen between an accessible loop versus a protected stem (Fig. 3b). We then applied these metrics to windows containing a total of 50 A/C nucleotides. Globally, mRNAs are much more structured *in vitro* compared to *in vivo*: there is a strong shift towards low *r* values and high Gini indices for the *in vitro* data that is far less pronounced *in vivo* (Fig. 3c). Thus unlike the rRNA, we find little evidence within mRNAs for *in vivo* DMS protection beyond what we observe *in vitro*, indicating that the DMS protection we observe *in vivo* is not due to mRNA–protein interactions. For example, using a cut-off (*r* value < 0.55, Gini index > 0.14) which captured the rRNAs and functionally validated mRNA

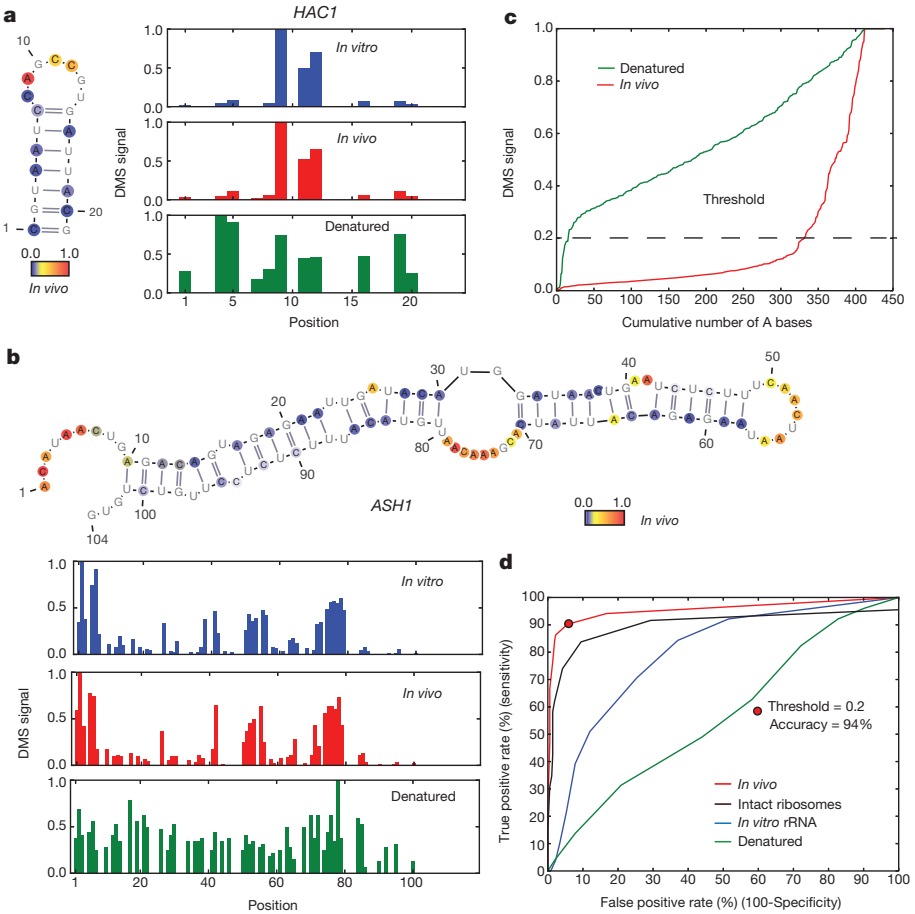


Figure 2 | Comparison of DMS-seq data to known RNA structures. **a**, **b**, DMS signal in *HAC1* (position 1 corresponds to chromosome VI:75828) (**a**) and *ASH1* (position 1 corresponds to chromosome XI:96245) (**b**). Number of reads per position was normalized to the highest number of reads in the inspected region, which is set to 1.0. Also shown are the known secondary structures with nucleotides colour-coded reflecting DMS-seq signal *in vivo*. **c**, DMS signal on 18S rRNA A bases plotted from least to most reactive. **d**, ROC curve on the DMS signal for A/C bases from the 18S rRNA. Threshold at 94% accuracy corresponds to 0.2 for the A bases.

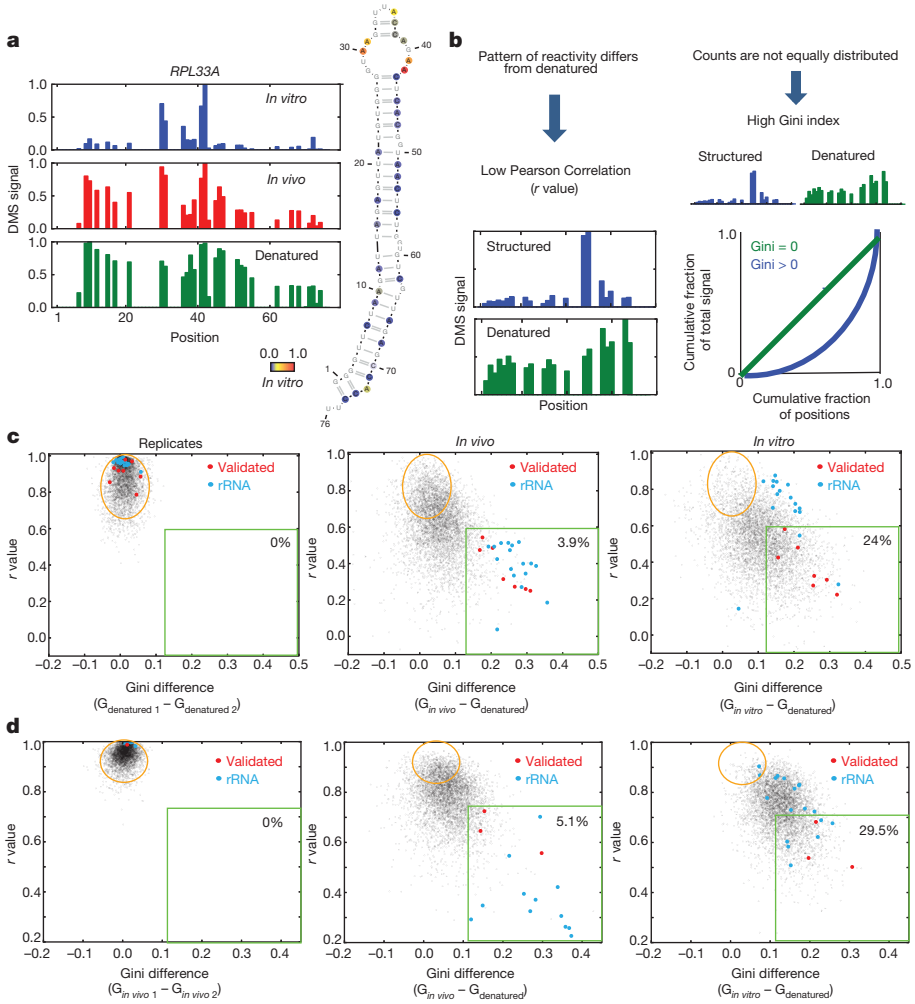


Figure 3 | Identification of structured mRNA regions reveals far less structure *in vivo* than *in vitro*. **a**, DMS signal in *RPL33A* mRNA, position 1 corresponds to chromosome XVI:282824. *In vitro* DMS signal colour-coded proportional to intensity and plotted onto the Mfold structure prediction. **b**, Schematic representation of the two metrics used to define structured regions within mRNAs. **c**, **d**, Scatter plots of Gini index difference versus *r* value from biological replicates or *in vivo* and *in vitro* relative to denatured samples for non-overlapping mRNA regions of 50 A/C nucleotides for yeast (**c**) and K562 cells (**d**). A total of 5,000 randomly selected regions are shown. Red dots represent regions spanning validated mRNA structures and blue dots are regions from rRNA. Evaluated regions have a minimum of 15 reads per A/C on average and their total number for *in vivo* data are 23,412 (**c**) and 17,242 (**d**).

structures, including both previously characterized and newly identified structures (see below), we found that out of 23,412 mRNA regions examined (representing 1,948 transcripts), only 3.9% are structured *in vivo* compared to 24% *in vitro* (Fig. 3c and Extended Data Fig. 2 for similar results obtained with windows of different sizes). In addition, 29% of the regions *in vivo* are indistinguishable from denatured (Fig. 3c, orange circle), whereas *in vitro* only 9% of regions were fully denatured. We also applied DMS-seq to mammalian cells (both K562 cells and human foreskin fibroblasts), which revealed results qualitatively very similar to yeast—a limited number of stable structures *in vivo* compared to *in vitro* (Fig. 3d and Extended Data Figs 3 and 4).

Because the pool of stable structures seen *in vivo* includes previously validated functional mRNA structures, this relatively small subset of mRNA regions provides highly promising candidates for novel functional RNA structures. To explore this, we focused on two structured 5′ untranslated regions (UTRs) from *PMA1* and *SFT2* and on the structured *PRC1* 3′ UTR for more detailed functional analyses. We fused these UTRs upstream or downstream, respectively, of a Venus protein reporter and quantified Venus levels by flow cytometry. Stem loop structures in these UTRs significantly increased (5′*SFT2*) or decreased (5′*PMA1* and 3′*PRC1*) protein levels upon disruption of their predicted base pairing interactions, and Venus protein levels were rescued by compensatory mutations (Extended Data Figs 5 and 6, Extended Data Table 1). Phylogenetic analysis revealed the 5′ UTR *PMA1* stem is under positive evolutionarily selection (Extended Data Fig. 5c), lending additional support for a physiological function. A list of 189 structured regions, along with a model of their secondary structures that are similarly supported by phylogenetic analysis of compensatory mutations, is hosted on an online database (http://weissmanlab.ucsf.edu/yeaststructures/index.html). In addition, we mutated predicted stems in three 3′ UTRs with evidence of strongly ordered structures *in vitro* but not in cells, and these mutations resulted in minimal expression changes (Extended Data Fig. 6d). Nonetheless, it remains possible that transient, heterogeneous or weakly ordered structures *in vivo* have biological roles, especially if they become more ordered under different physiological conditions.

To evaluate what role *in vitro* thermodynamic stability has in driving mRNA folding *in vivo*, we performed genome-wide structure probing experiments *in vitro* at five temperatures (30, 45, 60, 75 and 95 °C). As temperature rises and structure unfolds (Fig. 4a), the DMS signal becomes more even (low Gini index) and the modification pattern resembles that of the 95 °C denatured control (high *r* value). We defined *in vitro* temperature of unfolding (T_{unf}) as the lowest temperature where a region appeared similar to the denatured controls. Remarkably, many regions with little or no detectable structure *in vivo* show similar thermostability to highly structured regions, including structures that are functionally validated (Fig. 4a, b). For example, the regions of *RPL33A*

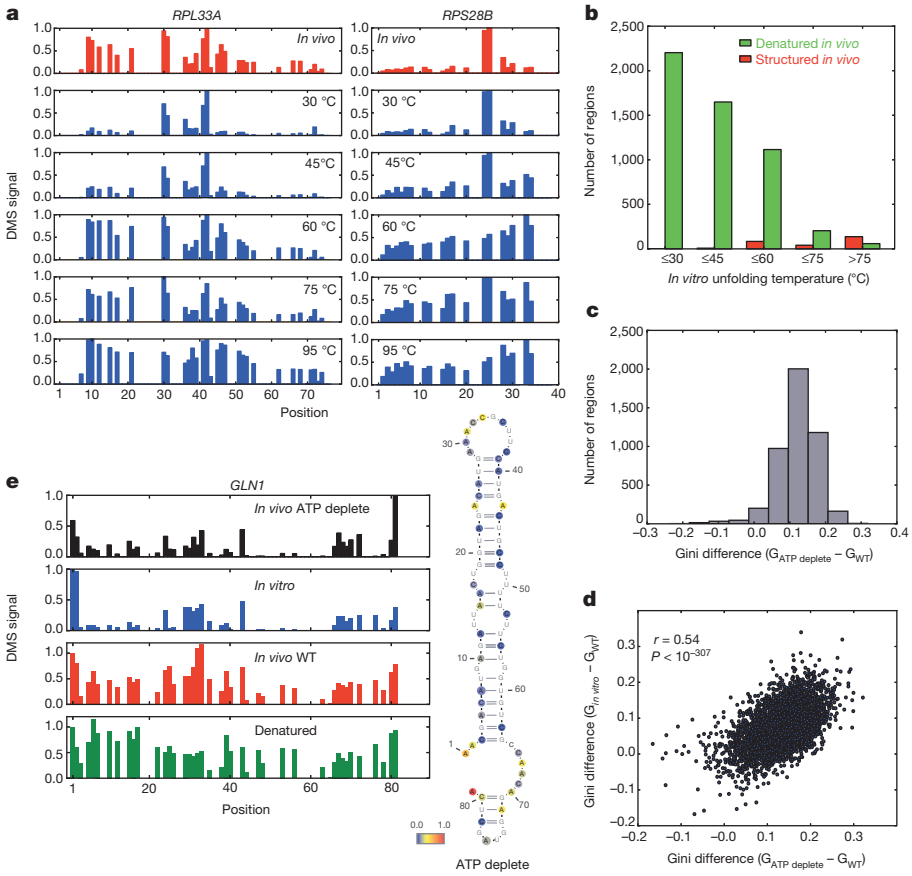


Figure 4 | Factors affecting the difference between mRNA structure *in vivo* and *in vitro*. **a**, Example of DMS signal changes for *RPL33A* and *RPS28B* *in vivo* and *in vitro* with increasing temperature. **b**, Histogram of *in vitro* unfolding temperature (T_{unf}) for denatured (green bars) or structured (red bars) regions *in vivo*. **c**, Histogram of Gini index difference between ATP-depleted and wild-type yeast samples. **d**, Gini index differences in ATP-depleted yeast or *in vitro* refolded mRNAs relative to wild-type yeast, calculated over 50 A/C nucleotides. **e**, Example of *in vivo* structure changes during ATP depletion. Position 1 corresponds to chromosome XVI:643069.

(unfolded *in vivo*) and *RPS28B* (a functionally validated structure *in vivo*) are both highly structured *in vitro* and have $T_{unf} = 60$ °C. Nonetheless, we find that structures present *in vivo* do have a strong propensity for high thermostability (Fig. 4b), consistent with a recent *in vitro* mRNA thermal unfolding study[8]. In addition to the role of thermostability in explaining the disparity of RNA structure between *in vivo* and *in vitro* samples, we tested the effect of Mg^{2+} concentration *in vitro*. We obtained similar structure results with 2–6 mM Mg^{2+}. However, at 1 mM Mg^{2+}, we observe unfolding of most structures, including the functionally validated ones (Extended Data Fig. 7a). The above observations indicate that Mg^{2+} concentration and thermodynamic stability have an important but incomplete role in determining mRNA structure *in vivo*.

A central question is what accounts for the differences between *in vivo* and *in vitro* mRNA structure. Although translation by ribosomes has a role in unwinding structure, this is unlikely to be the dominant force for unfolding *in vivo* because the average *in vivo* structure for coding regions was not distinguishable from 5′ and 3′ UTRs (Extended Data Fig. 7b). Moreover, within coding regions, high ribosome occupancy of an mRNA as measured by ribosome profiling[21] was not generally associated with lower structure (Extended Data Fig. 7c). It is likely that both active mechanisms (for example, RNA helicases) and passive mechanisms (for example, single-stranded-RNA binding proteins) counteract mRNA's intrinsic propensity to form the stable structures[22] seen with *in vitro* studies[3,23] and computational approaches[19]. To investigate how energy-dependent processes contribute to unfolding mRNA *in vivo*, we performed DMS-seq on yeast depleted of ATP[24]. We observed a marked increase in mRNA structure *in vivo* following ATP depletion (Fig. 4c). Moreover, the structural changes seen upon ATP depletion are strongly correlated ($r = 0.54$, $P < 10^{-307}$) to the changes between *in vivo* and *in vitro* samples (Fig. 4d, e and Extended Data Fig. 8). We also observed a large increase in mRNA structure at 10 °C *in vivo* (Extended Data Fig. 9a), but these changes are not as strongly correlated with those seen upon ATP depletion (Extended Data Fig. 9b). Thus the mRNA structures present in a cell are affected by a range of factors, underscoring the value of DMS-seq in defining the RNA structures present in a specific physiological condition or perturbation.

In summary, DMS-seq provides the first comprehensive exploration of RNA structure in a cellular environment and reveals that in rapidly dividing cells, mRNAs *in vivo* are far less structured than *in vitro*. This scarcity of structure is well suited for the primary role of mRNA as an informational molecule providing a uniform substrate for translating ribosomes. Nonetheless, we identify hundreds of specific mRNA regions that are highly structured *in vivo*, and we show for three examples that these structures affect protein expression. Our studies provide an excellent set of candidate regions, among the truly enormous number of structured regions seen *in vitro*, for exploring the regulatory role of structured mRNAs. The DMS-seq approach is readily extendable to other organisms, including human-derived samples as we show here, and to the analysis of the wide range of functional RNA molecules present in a cell. Thus DMS-seq broadly enables the analysis of structure-function relationships for both informational and functional RNAs. Among the many potential applications, attractive candidates include the analysis of long noncoding RNAs[25,26], the relationship between mRNA structure and microRNA/RNA interference targeting[27], and functional identification and analysis of ribozymes[28], riboswitches[29] and thermal sensors[30].

METHODS SUMMARY

DMS modification. For *in vivo* DMS modification, 15 ml of exponentially growing yeast (strain BY4741) at 30 °C were incubated with 300–600 µl DMS for 2–4 min (which results in multiple modifications per mRNA molecule). DMS was quenched with the addition of 30 ml stop solution (30% beta-mercaptoethanol (BME), 25% isoamyl alcohol). Total RNA was purified using hot acid phenol (Ambion). PolyA(+) mRNA was obtained using magnetic poly(A)$^+$ Dynabeads (Invitrogen).

Library generation. Sequencing libraries were prepared as outlined in Fig. 1. Specifically, DMS-treated mRNA samples were denatured at 95 °C and fragmented in 1× RNA fragmentation buffer (Ambion). Fragments of 60–70 nucleotides were gel-purified and ligated to microRNA cloning linker-1 (IDT) and reverse transcribed

using Superscript III (Invitrogen). Truncated reverse transcription products were gel-purified and circularized using CircLigase (Epicentre). Illumina sequencing adapters were introduced by 8–10 cycles of PCR.

Sequencing and sequence alignment. Raw sequences obtained from Hiseq2000 (Illumina) were aligned against *Saccharomyces cerevisiae* assembly R62 (UCSC: sacCer2). Aligned reads were filtered so that no mismatches were allowed and alignments were required to be unique.

Online resources. For secondary structure models that are supported by DMS-seq and have evidence for phylogenetic conservation, visit http://weissmanlab.ucsf.edu/yeaststructures/index.html.

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 25 March; accepted 25 November 2013.

Published online 15 December 2013.

1. Gruber, A. R., Neuböck, R., Hofacker, I. L. & Washietl, S. The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. *Nucleic Acids Res.* **35,** W335–W338 (2007).
2. Ouyang, Z., Snyder, M. P. & Chang, H. Y. SeqFold: Genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* **23,** 377–387 (2013).
3. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467,** 103–107 (2010).
4. Underwood, J. G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nature Methods* **7,** 995–1001 (2010).
5. Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nature Chem. Biol.* **9,** 18–20 (2013).
6. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl Acad. Sci. USA* **108,** 11063–11068 (2011).
7. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl Acad. Sci. USA* **106,** 97–102 (2009).
8. Wan, Y. *et al.* Genome-wide measurement of RNA folding energies. *Mol. Cell* **48,** 169–181 (2012).
9. Wells, S. E., Hughes, J. M., Igel, A. H. & Ares, M. Jr. Use of dimethyl sulfate to probe RNA structure *in vivo*. *Methods Enzymol.* **318,** 479–493 (2000).
10. Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334,** 1524–1529 (2011).
11. Ziehler, W. A. & Engelke, D. R. Probing RNA structure with chemical reagents and enzymes. *Curr. Protoc. Nucleic Acid Chem.* **6,** Unit 6.1 (2001).
12. Zaug, A. J. & Cech, T. R. Analysis of the structure of *Tetrahymena* nuclear RNAs *in vivo*: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA* **1,** 363–374 (1995).
13. Cordero, P., Kladwang, W., VanLang, C. C. & Das, R. Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry* **51,** 7037–7039 (2012).
14. Inoue, T. & Cech, T. R. Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc. Natl Acad. Sci. USA* **82,** 648–652 (1985).
15. Gonzalez, T. N., Sidrauski, C., Dörfler, S. & Walter, P. Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. *EMBO J.* **18,** 3119–3132 (1999).
16. Badis, G., Saveanu, C., Fromont-Racine, M. & Jacquier, A. Targeted mRNA degradation by deadenylation-independent decapping. *Mol. Cell* **15,** 5–15 (2004).
17. Chartrand, P., Meng, X. H., Singer, R. H. & Long, R. M. Structural elements required for the localization of *ASH1* mRNA and of a green fluorescent protein reporter particle *in vivo*. *Curr. Biol.* **9,** 333–338 (1999).
18. Rüegsegger, U., Leber, J. H. & Walter, P. Block of *HAC1* mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. *Cell* **107,** 103–114 (2001).
19. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–3415 (2003).
20. Wittebolle, L. *et al.* Initial community evenness favours functionality under selective stress. *Nature* **458,** 623–626 (2009).
21. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–223 (2009).
22. Herschlag, D. RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **270,** 20871–20874 (1995).
23. Li, F. *et al.* Global analysis of RNA secondary structure in two metazoans. *Cell Rep* **1,** 69–82 (2012).
24. Stade, K. *et al.* Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90,** 1041–1050 (1997).
25. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493,** 231–235 (2013).
26. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* (2013).
27. Tan, X. *et al.* Tiling genomes of pathogenic viruses identifies potent antiviral shRNAs and reveals a role for secondary structure in shRNA efficacy. *Proc. Natl Acad. Sci. USA* **109,** 869–874 (2012).
28. Tang, J. & Breaker, R. R. Structural diversity of self-cleaving ribozymes. *Proc. Natl Acad. Sci. USA* **97,** 5784–5789 (2000).
29. Li, S. & Breaker, R. R. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Res.* **41,** 3022–3031 (2013).
30. Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E. & Vilardell, J. Deciphering 3′ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. *Mol. Cell* **43,** 1033–1039 (2011).

Acknowledgements We thank R. Andino, M. Bassik, J. Doudna, J. Dunn, T. Faust, N. Stern-Ginossar, C. Gross, C. Guthrie, N. Ingolia, C. Jan, M. Kampmann, D. Koller, G.W. Li, S. Mortimer, E. Oh, C. Pop and members of the Weissman laboratory for discussions; J. Stewart-Ornstein and O. Brandman for plasmids; C. Chu, N. Ingolia and J. Lund for sequencing help. This research was supported by the Center for RNA Systems Biology (J.S.W.), the Howard Hughes Medical Institute (J.S.W.), and the National Science Foundation (M.Z.).

Author Contributions S.R., M.Z. and J.S.W. designed the experiments. S.R. and M.Z. performed the experiments, and S.R. analysed the data. S.W. and M.K. completed the phylogenetic analysis. S.R., M.Z. and J.S.W. drafted and revised the manuscript.

Author Information All data are deposited in Gene Expression Omnibus (accession number GSE45803). 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.S.W. (weissman@cmp.ucsf.edu).

LETTER

doi:10.1038/nature12946

Landscape and variation of RNA secondary structure across the human transcriptome

Yue Wan[1,2]*, Kun Qu[1]*, Qiangfeng Cliff Zhang[1], Ryan A. Flynn[1], Ohad Manor[3], Zhengqing Ouyang[1]†, Jiajing Zhang[1], Robert C. Spitale[1], Michael P. Snyder[4], Eran Segal[3] & Howard Y. Chang[1]

In parallel to the genetic code for protein synthesis, a second layer of information is embedded in all RNA transcripts in the form of RNA structure. RNA structure influences practically every step in the gene expression program[1]. However, the nature of most RNA structures or effects of sequence variation on structure are not known. Here we report the initial landscape and variation of RNA secondary structures (RSSs) in a human family trio (mother, father and their child). This provides a comprehensive RSS map of human coding and non-coding RNAs. We identify unique RSS signatures that demarcate open reading frames and splicing junctions, and define authentic microRNA-binding sites. Comparison of native deproteinized RNA isolated from cells versus refolded purified RNA suggests that the majority of the RSS information is encoded within RNA sequence. Over 1,900 transcribed single nucleotide variants (approximately 15% of all transcribed single nucleotide variants) alter local RNA structure. We discover simple sequence and spacing rules that determine the ability of point mutations to impact RSSs. Selective depletion of 'riboSNitches' versus structurally synonymous variants at precise locations suggests selection for specific RNA shapes at thousands of sites, including 3′ untranslated regions, binding sites of microRNAs and RNA-binding proteins genome-wide. These results highlight the potentially broad contribution of RNA structure and its variation to gene regulation.

We performed parallel analysis of RNA structure[2] (PARS) on RNA isolated from lymphoblastoid cells of a family trio (Fig. 1a). Deep sequencing of RNA fragments generated by RNase V1 or S1 nuclease (Extended Data Fig. 1a) determined the double or single-stranded regions, respectively, across the human transcriptome. We obtained over 160-million mapped reads for each individual. Transcript abundance and structure profiles are highly correlated among the individuals (Extended Data Fig. 2a, b). Summation of PARS data from the trio produced structural information for >20,000 transcripts with at least 1 read per base (load ≥ 1, Fig. 1b), and accurately identified known RSSs in RNAs (Fig. 1c and Extended Data Fig. 1b, c). We also developed methods for RNA extraction, deproteinization, and PARS under native conditions (native deproteinized samples) that accurately captured structures with known RSS, and revealed RSS for 6,524 transcripts (Extended Data Fig. 3a–d).

PARS data for thousands of transcripts afforded a genome-wide view of the structural landscape of human messenger RNAs. Metagene analysis shows that, on average, the coding region (CDS) is demarcated by focally accessible regions near the translational start site and stop codon. Contrary to yeast, human CDS is slightly more single-stranded than the untranslated regions (UTRs) (Fig. 1d), similar to previous trends in other metazoans[3]. A three-nucleotide structure periodicity is present in the CDS and absent in UTRs, consistent with prior computational prediction[4]. Both renatured and native mRNAs showed similar RSS features, suggesting that RNA sequence is a strong determinant of RSS. However, RNA structures also deviate from sequence content. In particular, human 3′ UTR has low GC content but is highly structured (Fig. 1d). We also identified 583 (5.7%) consistently different regions between native deproteinized and renatured structure profiles, providing candidate sites for regulation of RNA structure *in vivo* (Supplementary Table 1). Highly structured RNAs have fewer structure differences as compared to mRNAs (Extended Data Fig. 3e), suggesting stronger evolutionary selection for functional conformations. We note that 3.7% of bases (residing in 9.7% of transcripts) have both strong V1 and S1 reads, indicating the existence of multiple mRNA conformations.

We detected unique signatures of RSSs at sites of post-transcriptional regulation. RNA structure is believed to be important in regulating distinct splicing signals on exons and introns of pre-messenger RNAs[5]. We observed a unique asymmetric RSS signature at the exon–exon junction in both renatured and native deproteinized transcripts that is not simply explained by GC content. The terminal AG dinucleotide at the end of the 5′ exon tends to be more accessible, whereas the first nucleotides of the 3′ exon are more structured (Fig. 2a and Extended Data Fig. 3f). Hence, a specific RSS signature may contribute to RNA splicing.

Regulation of mRNAs by microRNAs (miRNAs) is an important post-transcriptional process that causes translation repression and/or mRNA degradation[6]. However the extent to which structural accessibility drives productive miRNA targeting is still unclear. Analysis of RSS from renatured RNA around predicted miRNA targets revealed that true Argonaute (AGO)-bound target sites[7] show strong structural accessibility from −1 to 3 nucleotides upstream of the miRNA-target site compared to predicted targets not bound by AGO ($P < 10^{-10}$, Wilcoxon rank-sum test; Fig. 2b, orange window, and Extended Data Fig. 4a). AGO-bound sites are also more accessible at bases 4 to 6 of the miRNA-target site ($P = 0.004$, Wilcoxon rank-sum test), agreeing with prior computational predictions[8]. To test whether our identified 5′ accessibility neighbourhood (−1 to 3 nucleotides) is truly important for AGO binding, we performed AGO individual nucleotide-resolution crosslinking and immunoprecipitation (iCLIP) on each member of the trio. Separating the predicted target sites according to average 5′ structural accessibility showed that single-stranded targets are more likely to be AGO-bound than double-stranded targets (Fig. 2c and Extended Data Fig. 4b). The most significant difference in AGO binding occurs close to our identified accessible region ($P = 0.01$, Fig. 2d). Separating predicted targets into five accessibility quantiles also demonstrated that the most accessible 20% of predicted targets are most AGO bound ($P < 10^{-19}$, Fig. 2e). Furthermore, ectopic expression of miR142 or miR148 in HeLa cells[9] resulted in greater repression of mRNAs with the 100 most accessible sites as compared to mRNAs with the 100 least accessible sites ($P < 0.005$, Wilcoxon rank-sum test; Fig. 2f and Extended Data Fig. 4c, d). This indicates that mRNAs with accessible miRNA sites are more likely to be true targets, and upstream accessibility is important for miRNA targeting.

[1]Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Stem Cell and Development, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672. [3]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovet 76100, Israel. [4]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. †Present address: The Jackson Laboratory for Genomic Medicine, 263 Farmington Avenue, ASB Call Box 901 Farmington, Connecticut 06030, USA. *These authors contributed equally to this work.

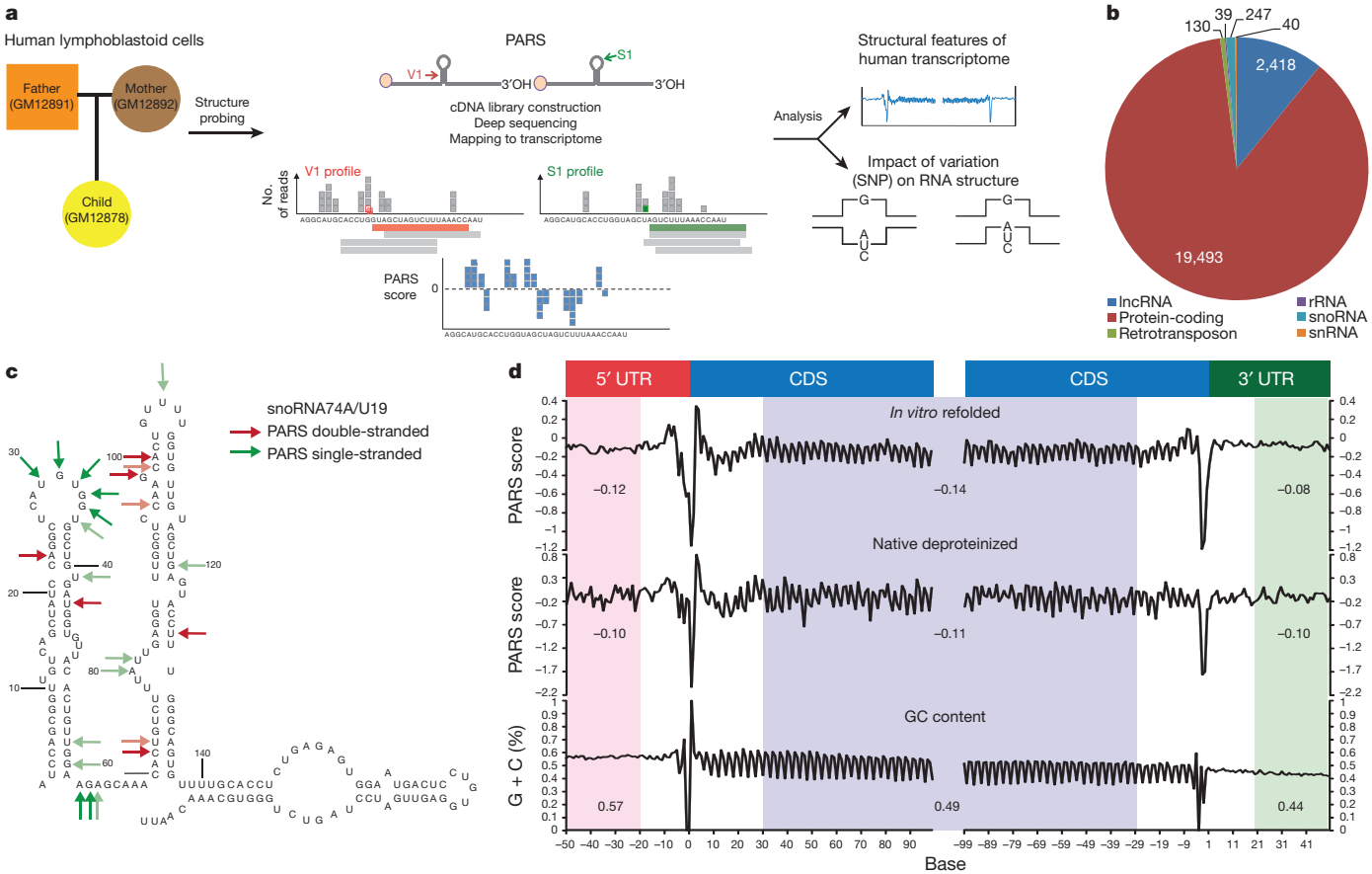


Figure 1 | PARS reveals the landscape of human RNA structure. **a**, Experimental overview. Circles represent females, squares represent males. **b**, Pie chart showing the distribution of structure-probed RNAs with a coverage of at least one read per base. **c**, High (red arrows) and low (green arrows) PARS scores were mapped onto the secondary structure of small nucleolar RNA snoRNA74A. Red (positive PARS score), double-stranded regions by PARS score; green (negative PARS score), single-stranded regions by PARS score. The colour intensity reflects the magnitude of the PARS scores. Darker red and darker green, reflect more positive and more negative PARS scores (double- and single-stranded regions), respectively. **d**, PARS score (top, renatured transcripts; middle, native deproteinized transcripts) and GC content (bottom) across the 5′ UTR, the coding region, and the 3′ UTR, averaged across all transcripts, aligned by translational start and stop sites. Averaged regions are shaded in pink, blue and green for 5′ UTR, CDS and 3′ UTR, respectively.

Comparison of RNA structural landscapes between individuals revealed the impact of diverse sequence variants on RNA structure. As a class, local PARS score differences at single nucleotide variants (SNVs) were significantly greater than biological replicates of an invariant doped in RNA ($P < 0.001$ Kolmogorov–Smirnov test; Extended Data Fig. 5a). SNVs that alter RNA structure, known as 'riboSNitches', also exhibit threefold greater local structure change than replicates of the same sequence in different individuals (Extended Data Fig. 5b). At a gene level, transcripts with SNVs are significantly more disrupted, calculated using the experimental structure disruption coefficient (eSDC)[10], than transcripts without SNVs ($P = 1.3 \times 10^{-4}$, Kolmogorov–Smirnov Test; Extended Data Fig. 5c, d). Furthermore, 78.2% of all structure changing bases lie in transcripts that contain either SNVs or indels, suggesting that sequence variation is important in shaping RSS variation in the human transcriptome (Extended Data Fig. 5e). The list of the top 2,000 disrupted transcripts is shown in Supplementary Table 2.

To pinpoint riboSNitches[11], we calculated structure changes between each pair of individuals (Fig. 3a) and selected SNVs that had large PARS score differences, low false discovery rate (FDR), significant P value, and high local read coverage (Methods). Permutation analysis across genotypes and along transcripts confirmed that riboSNitches are significantly detected over random noise (Methods). We experimentally validated nine riboSNitches using independent structure probing methods such as nucleases, selective 2′ hydroxyl acylation and primer extension (SHAPE) or dimethyl sulphate (DMS), and confirmed the ability of PARS to discover riboSNitches (Extended Data Figs 6–9). The SeqFold program is used to visualize structure changes caused by riboSNitches[12] (Fig. 3b, c and Extended Data Fig. 7g, h).

We found that 1,907 out of 12,233 (15%) SNVs switched RNA structure in the trio (Fig. 3d, Extended Data Fig. 5e and Supplementary Table 3). As riboSNitches are expected to cause RSS changes in a heritable and allele-specific fashion, we performed allele-specific PARS in the cell line derived from the child by mapping uniquely across each of the two alleles for SNVs that are homozygous and different in the parents (for example, father AA and mother GG, with child AG when he or she inherits one copy from each parent) (Methods and Extended Data Fig. 6e). Out of 172 parental homozygous riboSNitches, 117 (68%) were validated by allele-specific mapping in the child. As only reads upstream of the riboSNitch can be uniquely mapped and detected, this is likely to be an underestimate. We also observed a validation rate of 61% in native deproteinized samples of the child, indicating that the structural changes are biologically relevant *in vivo* (Extended Data Fig. 9b).

The large numbers of riboSNitches identified raised the possibility that riboSNitches may have greater influence on gene regulation and human diseases than previously appreciated. Intersection with expression quantitative trait loci (eQTL) identified 211 riboSNitches that are associated with changes in gene expression (Supplementary Table 4). Overlapping riboSNitches with the NHGRI catalogue of genome-wide association studies identified 22 unique riboSNitches that are associated with diverse human diseases and phenotypes, including multiple

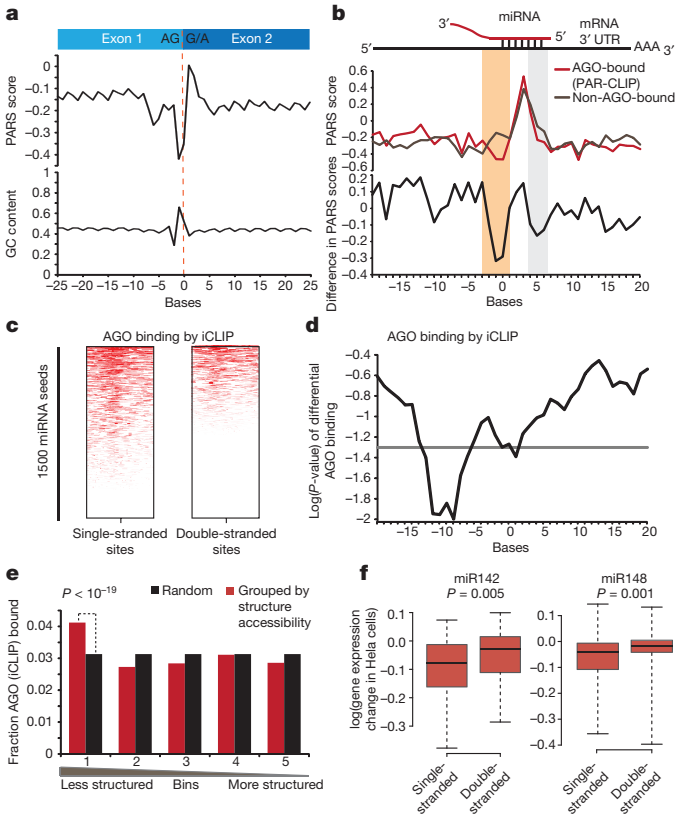


Figure 2 | RSS signatures of post-transcriptional regulation. **a**, Average PARS score and GC content across transcript exon–exon junctions. **b**, Average PARS score (top) and PARS score difference (bottom) across miRNA sites for AGO-bound (red) versus non-AGO-bound sites (grey). Structurally different regions are in orange and light grey. **c**, AGO-iCLIP binding for single- versus double-stranded miRNA target sites. **d**, *P* value for differential AGO-iCLIP binding (*t*-test, $P = 0.05$ in grey). **e**, Observed versus expected AGO binding (*P* value, chi-squared test). **f**, Expression changes of mRNAs with accessible and inaccessible miR142 (left) or miR148 (right) sites, upon miRNA overexpression (Wilcoxon rank-sum test).

sclerosis, asthma and Parkinson's disease (Supplementary Table 5). Hence, many non-coding changes in the transcriptome may alter gene function by altering RNA structure.

We also observed sequence and context rules in riboSNitches. First, riboSNitches that lie in double- or single-stranded regions tend to become more single- or double-stranded, respectively, after nucleotide change (Fig. 3e). Second, the nucleotide content of the riboSNitch is instructive of the direction of RSS change. Bases that undergo G/C to A/T changes tend to become more single-stranded, whereas bases that change from A/T to G/C tend to become more paired (Fig. 3f). This effect is stronger for homozygous riboSNitches than heterozygous riboSNitches, and typically disrupts 10 bases centred on the mutation. Third, the structural context flanking SNVs influence their transition to become more single- or double-stranded (Extended Data Fig. 10a–c). Fourth, riboSNitches have fewer SNVs around them as compared to non-structure changing SNVs, suggesting that co-variation of some SNVs may help to maintain functional RNA structures (Extended Data Fig. 10d).

The distribution of extant riboSNitches provides insights into regions of the transcriptome that require specific RNA shape. If an RSS is functionally important, a riboSNitch that disrupts the structure will be evolutionarily selected against, whereas a non-structure-changing SNV will not (Fig. 4a)[13]. We tested whether such selection occurs in the human transcriptome, and found that riboSNitches are significantly depleted at 3′ UTRs compared to control SNVs ($P < 10^{-20}$, chi-squared test; Fig. 4b). This depletion is even stronger for larger disruptions which would be expected to be less tolerated (Extended Data Fig. 10e). Additional genomic features associated with riboSNitches are also found (Extended Data Fig. 10f, Supplementary Table 6). RiboSNitches are also significantly depleted around predicted miRNA target sites ($P < 10^{-5}$, chi-squared test; Fig. 4c) and RNA binding protein (RBP) binding sites ($P = 0.004$, chi-squared test). However, depletion of riboSNitches varies for each individual RBP (Fig. 4d), suggesting that different RBPs may have different RSS requirements for binding. RiboSNitches may also influence gene regulation through splicing. Indeed, riboSNitches near splice junctions are associated with greater alternative splicing changes (defined as percentage spliced in (PSI)[14,15]; Fig. 4e), suggesting that RNA structures could regulate splicing.

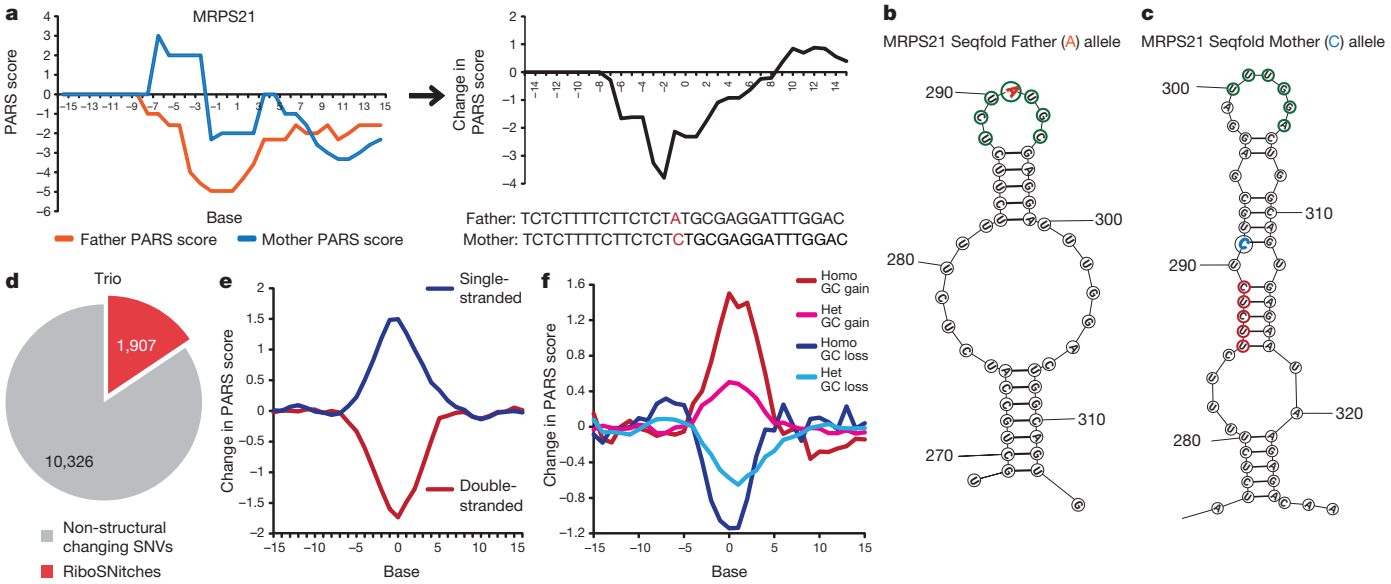


Figure 3 | PARS identifies riboSNitches genome-wide. **a**, PARS score (left) and PARS-score difference (right) of *MRPS21* father's and mother's alleles. **b**, **c**, SeqFold models of *MRPS21* A and C alleles (single-and double-stranded bases circled in green and red, respectively). **d**, Number of SNVs identified as riboSNitches in the trio. **e**, **f**, Average PARS score changes of riboSNitches that (**e**) originally reside in double-stranded (red) or single-stranded regions (blue); or (**f**) undergo nucleotide changes from A/T to G/C (red, pink) or from G/C to A/T (dark and light blue). 0 indicates the position of SNV on the *x* axis.

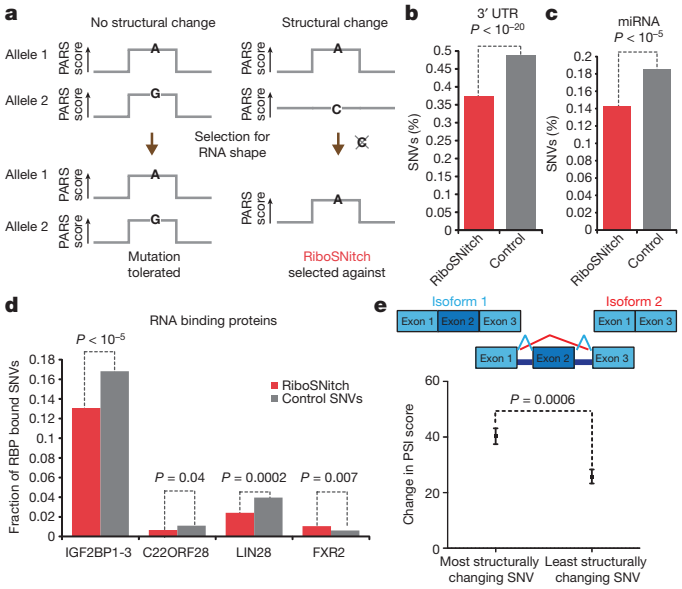


Figure 4 | Genetic evidence for functional RSS elements in the transcriptome. **a**, Schematic of RSS selection test: mutations that do not change the shape of an important RNA structure may be tolerated and accumulates (left), but a riboSNitch that changes RNA shape will be evolutionarily selected against and removed. Brown arrows, alleles that were present before and after selection for RNA shape. **b–d**, Selective depletion of riboSNiches versus structurally synonymous SNVs at 3′ UTRs (**b**); predicted miRNA target sites (**c**); specific RBP binding sites (**d**). *P* value is calculated using chi-square test. **e**, RiboSNitches impact splicing. PSI score is calculated to be the ratio of alternatively spliced isoform versus total isoforms (Methods), $P = 0.0006$, Student's *t*-test. Error bars show mean ± s.e.m.

In summary, the landscape and variation of RSS across human transcriptomes suggest important roles of RNA structure in many aspects of gene regulation. We provide the experimental and analytical frameworks to evaluate SNVs that change RSSs, and demonstrate potentially much broader roles for riboSNitches in multiple steps of post-transcriptional regulation. In the future, use of high resolution, *in vivo* probes of RSSs[16] and studies of many individuals of diverse genetic backgrounds may allow systematic determination of functional RSSs across the transcriptome.

METHODS SUMMARY

Sample preparation and structure probing for human renatured RNAs. Human lymphoblastoid cell lines GM12878, GM12891 and GM12892 were obtained from Coriell. Total RNA was isolated using TRIzol reagent (Invitrogen) and polyA selected as described previously[2]. Two micrograms of Poly(A)$^+$ RNA was structure probed at 37 °C using RNase V1 (Life Technologies, final concentration of 10^{-5} units per μl) or S1 nuclease (Fermentas, final concentration of 0.4 units per μl) at 37 °C for 15 min.

Sample preparation and structure probing for human native deproteinized RNAs. GM12878 cells were lysed in lysis buffer (150 mM NaCl, 10 mM $MgCl_2$, 1% NP40, 0.1% SDS, 0.25% Na deoxycholate, Tris pH 7.4) on ice for 30 min. The lysate was deproteinized by phenol chloroform extractions. Total RNA (1 μg per 90 μl) was incubated in 1 × RNA structure buffer at 37 °C for 15 min and structure probed using RNase V1 (final concentration of 2×10^{-5} units per μl) and S1 nuclease (final concentration of 0.2 units per μl) at 37 °C for 15 min.

Library construction and analysis. The structure probed RNA was cloned using Ambion RNA-Seq Library Construction Kit (Life Technologies)[2], and sequenced using Illumina Hi-seq. The reads were trimmed and mapped to UCSC RefSeq and the Gencode v12 databases (hg19 assembly) using the software Bowtie2 (ref. 17). Double (V1) and single-stranded reads (S1) for each sequencing sample were normalized by sequencing depth.

RiboSNitch analysis. Data normalization for each sample was performed by calculating standard deviation (s.d.) for each transcript and dividing the PARS score per base by the s.d. of that transcript. We defined a structure difference of the *i*th base of transcript *j* between conditions *m* and *n* in this formula, where PARS represents the normalized PARS score, abs represents absolute value, and *k* represents the *k*th base of the transcript:

$$\mathrm{StrucDiff}_{i,j,m,n} = \sum_{k=i-2}^{k=i+2} \frac{\mathrm{abs}(\mathrm{PARS}_{k,j,m} - \mathrm{PARS}_{k,j,n})}{5}$$

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 5 April; accepted 16 December 2013.

1. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. *Nature Rev. Genet.* **12,** 641–655 (2011).
2. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467,** 103–107 (2010).
3. Li, F. *et al.* Global analysis of RNA secondary structure in two metazoans. *Cell. Rep.* **1,** 69–82 (2012).
4. Shabalina, S. A., Ogurtsov, A. Y. & Spiridonov, N. A. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* **34,** 2428–2437 (2006).
5. Barash, Y. *et al.* Deciphering the splicing code. *Nature* **465,** 53–59 (2010).
6. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215–233 (2009).
7. Skalsky, R. L. *et al.* The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog.* **8,** e1002484 (2012).
8. Marín, R. M., Voellmy, F., von Erlach, T. & Vanicek, J. Analysis of the accessibility of CLIP bound sites reveals that nucleation of the miRNA:mRNA pairing occurs preferentially at the 3′-end of the seed match. *RNA* **18,** 1760–1770 (2012).
9. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27,** 91–105 (2007).
10. Ritz, J., Martin, J. S. & Laederach, A. Evaluating our ability to predict the structural disruption of RNA by SNPs. *BMC Genomics* **13,** (Suppl. 4) S6, (2012).
11. Halvorsen, M., Martin, J. S., Broadaway, S. & Laederach, A. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet.* **6,** e1001074 (2010).
12. Ouyang, Z., Snyder, M. P. & Chang, H. Y. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* 377–387 (2013).
13. Salari, R., Kimchi-Sarfaty, C., Gottesman, M. M. & Przytycka, T. M. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. *Nucleic Acids Res.* **41,** 44–53 (2013).
14. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7,** 1009–1015 (2010).
15. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338,** 1587–1593 (2012).
16. Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nature Chem. Biol.* **9,** 18–20 (2013).
17. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).

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

Acknowledgements We thank members of the Chang laboratory, S. Rouskin, and J. Weissman, A. Mele and R. Darnell for discussion. This work is supported by NIH R01-HG004361 (H.Y.C. and E.S.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

Author Contributions H.Y.C. conceived the project; Y.W. and H.Y.C. developed the protocol and designed the experiments; Y.W. and R.A.F. performed experiments; Y.W., K.Q., Q.C.Z., O.M., Z.O., J.Z., R.C.S., M.P.S., E.S., and H.Y.C. planned and conducted the data analysis; Y.W., K.Q. and H.Y.C. wrote the paper with contributions from all authors.

Author Information Data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE50676. 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 H.Y.C. (howchang@stanford.edu) and Y.W. (wany@gis.a-star.edu.sg).

METHODS

Sample preparation for renatured RNA structure probing. Human lymphoblastoid cell lines GM12878, GM12891 and GM12892 were obtained from Coriell. Total RNA was isolated from lymphoblastoid cells using TRIzol reagent (Invitrogen). Poly(A)$^+$ RNA was obtained by purifying twice using the MicroPoly(A)Purist kit (Life Technologies). The Tetrahymena ribozyme RNA was *in vitro* transcribed using the T7 RiboMax Large-scale RNA production system (Promega) and added into 2 μg of poly(A)$^+$ RNA (1% by mole) for structure probing and library construction.

Structure probing of renatured poly(A)$^+$ RNA. Two micrograms of Poly(A)$^+$ RNA in 160 μl of nuclease free water was heated at 90 °C for 2 min and snap-cooled on ice for 2 min. Twenty microlitres of 10 × RNA structure buffer (150 mM NaCl, 10 mM $MgCl_2$, Tris, pH 7.4) was added to the RNA and the RNA was slowly warmed up to 37 °C over 20 min. The RNA was then incubated at 37 °C for 15 min and structure probed independently using RNase V1 (Life Technologies, final concentration of 10^{-5} units per μl) or S1 nuclease (Fermentas, final concentration of 0.4 units per μl) at 37 °C for 15 min. The cleavage reactions were inactivated using phenol chloroform extraction.

Structure probing and ribosomal RNA depletion for native deproteinized RNA structure probing. GM12878 cells were lysed in lysis buffer (150 mM NaCl, 10 mM $MgCl_2$, 1% NP40, 0.1% SDS, 0.25% Na deoxycholate, Tris, pH 7.4) on ice for 30 min. The chromatin pellet was removed by centrifugation at 16,000*g* for 10 min at 4 °C. The lysate was deproteinized by passing through two phenol followed by one chloroform extractions. The concentration of RNA in the deproteinized lysate was measured using the Qubit fluorometer (Invitrogen). We diluted the RNA to a concentration of 1 μg per 90 μl using 1 × RNA structure buffer (150 mM NaCl, 10 mM $MgCl_2$, Tris, pH 7.4) and incubated the RNA at 37 °C for 15 min. The native deproteinized RNA was structure probed independently using RNase V1 (final concentration of 2×10^{-5} units per μl) and S1 nuclease (final concentration of 0.2 units per μl) at 37 °C for 15 min.

To compare structural differences between renatured and native deproteinized RNAs, we independently prepared an RNA sample that was similarly lysed and deproteinized. After removal of proteins, we ethanol precipitated the RNA and dissolved it in nuclease free water. We diluted the RNA to a concentration of 1 μg per 80 μl in water and heated the RNA at 90 °C for 2 min before snap-cooling the RNA on ice. We added 10 × RNA structure buffer and renatured the RNA by incubating it at 37 °C for 15 min and performed structure probing similarly as in native deproteinized RNAs.

The cleavage reactions were inactivated using phenol chloroform extraction and DNase treated before undergoing ribosomal RNA depletion using Ribo-Zero Ribosomal RNA removal kit (Epicentre).

Validation of riboSNitches by manual footprinting. We cloned approximately 200 nucleotide fragments of both alleles of *MRPS21, WSB1, HLA-DRB1, HLA-DQA1, hnRNP-AB, HLA-DRA, LDHA, XRCC5* and *FNBP1* from GM12878, GM12891 and GM12892 using a forward-T7-gene-specific primer and a reverse-gene-specific primer. All constructs were confirmed by sequencing using capillary electrophoresis. DNA from each of the different clones was then *in vitro* transcribed into RNA using MegaScript Kit from Ambion, following manufacturer's instructions.

Two picomoles of each RNA is heated at 90 °C for 2 min and chilled on ice for 2 min. 3.33 × RNA folding mix (333 mM HEPES, pH 8.0, 20 mM MgCl2, 333 mM NaCl) was then added to the RNA and the RNA was allowed to fold slowly to 37 °C over 20 min. The RNA was then structure probed with either DMS (final concentration of 100 mM) or 2-methylnicotinic acid imidazolide (NAI) (final concentration of 100 mM)[16] at 37 °C for 20 min or structure probed with S1 nuclease (final concentration of 0.4 units per μl) or RNase V1 (final concentration of 0.0001 units per μl) at 37 °C for 15 min. The DMS structure probed samples were quenched using 2-mercaptoethanol before phenol chloroform extraction. The NAI and nuclease treated samples were phenol chloroform extracted directly after structure probing. The structure probed RNA was then recovered through ethanol precipitation. The RNA structure modification/cleavage sites were then read out using a radiolabelled RT primer by running onto denaturing PAGE gel as described previously[18].

Library construction. The structure-probed RNA was fragmented at 95 °C using alkaline hydrolysis buffer (50 mM Sodium Carbonate, pH 9.2, 1 mM EDTA) for 3.5 min. The fragmented RNA was then ligated to 5′ and 3′ adapters in the Ambion RNA-Seq Library Construction Kit (Life Technologies). The RNA was then treated with Antarctic phosphatase (NEB) to remove 3′ phosphates before re-ligating using adapters in the Ambion RNA-Seq Library Construction Kit (Life Technologies). The RNA was reverse-transcribed using 4 μl of the RT primer provided in the Ambion RNA-Seq Library Construction Kit and polymerase chain reaction (PCR)-amplified following the manufacturer's instructions. We performed 18 cycles of PCR to generate the complementary DNA library.

Illumina sequencing and mapping. We performed paired end sequencing on Illumina's Hi-Seq sequencer and obtained approximately 400-million reads for each paired end lane in an RNase V1 or S1 nuclease library. Obtained raw reads were truncated to 50 bases, (51 bases from the 3′ end were trimmed). Trimmed reads were mapped to the human transcriptome, which consists of non-redundant transcripts from UCSC RefSeq and the Gencode v12 databases (hg19 assembly), using the software Bowtie2 (ref. 17). We allowed up to one mismatch per seed during alignment, and only included reads with perfect mapping or with Bowtie2 reported mismatches on positions annotated as SNVs in genetically modified cells. We obtained 166- to 212-million mapped reads for an RNase V1 or S1 nuclease sample.

PARS-score calculation. After the raw reads were mapped to the transcriptome, we calculated the number of double-stranded reads and single-stranded reads that initiated on each base on an RNA. The number of double (V1) and single stranded reads (S1) for each sequencing sample were then normalized by sequencing depth. For a transcript with N bases in total, the PARS score of its ith base was defined by the following formula where V1 and S1 are normalized V1 and S1 scores, respectively. A small number 5 was added to reduce the potential over-estimating of structural signals of bases with low coverage:

$$\mathrm{PARS}_{i=1\ldots N} = \log_2(\mathrm{V1}_i + 5) - \log_2(\mathrm{S1}_i + 5)$$

To identify structural changes caused by SNVs, we applied a 5-base average on the normalized V1 and S1 scores to smoothing the nearby bases' structural signals; therefore, the PARS score is defined as:

$$\mathrm{PARS}_{i=1\ldots N} = \log_2\left(\sum_{j=i-2}^{j=i+2} \frac{\mathrm{V1}_j + 5}{5}\right) - \log_2\left(\sum_{j=i-2}^{j=i+2} \frac{\mathrm{S1}_j + 5}{5}\right)$$

Bases with both high V1 and S1 scores, and transcripts with multiple conformations. Bases with both strong single- and double-strand signals are potentially present in multiple conformations. We first normalized all bases with detectable S1 or V1 counts by their sequencing depth. We then calculated an S1 ratio and a V1 ratio by normalizing S1 (and V1) counts to the transcript abundance. S1 and V1 ratios indicate the relative strength of single and double signals respectively. We then ranked all the bases by their S1 ratio and V1 ratio independently, and used the top one-million S1 ratio bases and the top one-million V1 ratio bases as high S1 ratio bases and high V1 ratio bases, respectively. We defined a base as being in multiple conformations if the base has both high S1 and high V1 ratios. If a transcript contains more than five multi-confirmation bases, this transcript is defined as a multi-confirmation transcript.

V1 replicates correlation analysis. Pearson correlation of RNase V1 replicates on GM12878 was performed using a parsV1 score (a value that uses the V1 score only to represent secondary structure) defined as:

$$\mathrm{parsV1}_{i=1..N} = \log_2(\mathrm{V1}_i + 5)$$

Structure differences between AGO PAR-CLIP bound and not bound transcripts. Predicted conserved and non-conserved miRNA target sites of conserved miRNA families were obtained from TargetScan[19]. AGO PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) data set in Epstein–Barr virus (EBV)-transformed lymphoblastoid cells was obtained from ref. 7. For 11 of the most abundant miRNAs that were expressed in the 4 lines of EBV transformed lymphoblastoid cells, we asked whether the predicted target site fell within the AGO CLIP clusters. Predicted target sites that resided within the PAR-CLIP clusters were considered as AGO-bound, whereas the rest were considered as non-AGO-bound. The non-AGO-bound transcripts are further controlled to fall within 25 and 75% of 3′ UTR length, mRNA abundance and CpG dinucleotide content of the AGO-bound transcripts. The PARS scores for AGO-bound and non-bound transcripts were aligned to the start (either −7 or −8 position of the miRNA) of the miRNA:target binding site and averaged. *P* values of structural changes were calculated using the Wilcoxon rank-sum test.

AGO-iCLIP library generation. AGO iCLIP was performed as described previously[20] with the following modifications: 2×10^7 genetically modified cells (per biological replicate) were collected under log-phase growth and washed once in ice-cold 1 × PBS. The pellet was resuspended in 10 × pellet volumes of ice-cold 1 × PBS and plated out on 10-cm tissue-culture dishes. Cells were crosslinked with ultraviolet radiation at 254nm for 0.3 J cm^{-2}, collected in ice-cold PBS and cell pellets were frozen on dry ice. Lysate preparation, RNaseA, and immunoprecipitation of AGO were performed as described previously[21] using the anti-AGO antibody (clone 2A8, Millipore). To produce iCLIP libraries, on-bead enzymatic steps and off-bead final-library preparation was performed as described previously[21]. AGO-iCLIP libraries were produced in biological duplicates for each individual (GM12891, GM12892 and GM12878), barcoded, and pooled for sequencing. Samples were single-end-sequenced for 75 bases on an Illumina HiSeq2500 machine.

Processing of AGO-iCLIP data. Raw sequencing reads were preprocessed using FASTX-Tookit before alignment was performed. Sequencing adaptor was trimmed off using fastx_clipper and low-quality reads were filtered using fastq_quality_filter. PCR duplicates were further removed using the program fastq_collapser. Preprocessed reads were aligned to hg19 genome assembly using Bowtie[22], and AGO–RNA crosslinking positions were obtained through self-generated script passing through the sequence alignment/map (SAM) file. AGO-RNA binding signal was smoothed by extending ± 10 bases around the crosslinking position, and signals from both replicates were normalized by sequencing depth. AGO–RNA per-base enrichment was defined as the minimum signal of the replicates divided by the corresponding RNA abundance.

To identify miRNA predicted sites for miRNAs that are expressed in GM12878 cells, we downloaded the small RNA sequencing data from the ENCODE consortium (GEO accession number GSM605625), and aligned the raw reads to the human miRNA database using Blastn. We estimated the amount of miRNA expression by counting the Blastn perfect matches for each miRNA. Predicted miRNA target sites from the top 100 highest expressed miRNA were then aligned to the miRNA–target binding sites and were separated into two groups: 0 predicted sites with an average PARS score of less than −1 (from −3 to 1 of the miRNA–target pair) were classified as single-stranded sites, whereas those with an average PARS score of greater than 1 (from −3 to 1 of the miRNA–target pair) were classified as double-stranded sites. We then calculated the average AGO-iCLIP enrichment score for the two groups of miRNA binding sites (from −25 to 25 bases), and estimated the significance of their difference using the Student's *t*-test.

miRNA-target downregulation in Hela cells. Average gene expression changes upon expression of miR142 or miR148 in HeLa cells were obtained from Grimson *et al.* by averaging the gene expression changes induced by the miRNA at 12 h and 24 h of overexpression[9]. For the miR142 or miR148 Targetscan predicted miRNA sites, we calculated the average PARS score across −3 to +1 (from the start of the miRNA–target pair) and sorted the predicted sites according to their structural accessibility. The *P* value for difference in downregulation of transcripts that contain the top 100 accessible sites versus transcripts that contain the bottom 100 accessible sites was calculated using Wilcoxon rank-sum test.

RiboSNitch analysis. RNAs with known secondary structures were doped into the initial RNA pool as positive controls to estimate the baseline changes in RNA structure in PARS. We calculated the PARS scores for all the bases in the transcripts and performed data normalization in order to compare directly secondary structures between different individuals. To normalize the data, we calculated the standard deviation (s.d.) for each transcript and divided the PARS score per base by the s.d. of that transcript. This resulted in a normal distribution of PARS scores for each transcript in each individual and enabled us to calculate the change in PARS scores due to SNVs by subtraction of PARS scores between the individuals. Since a true structure change is likely to extend beyond a single base, we define a structure difference of the *i*th base of transcript *j* between conditions *m* and *n* in this formula, where *PARS* represents the normalized PARS score:

$$\mathrm{StrucDiff}_{i,j,m,n} = \sum_{k=i-2}^{k=i+2} \frac{\mathrm{abs}(\mathrm{PARS}_{k,j,m} - \mathrm{PARS}_{k,j,n})}{5}$$

We calculated the StrucDiff for all the bases in all the transcripts between each pair of individuals: GM12891 and GM12892, GM12891 and GM12878, GM12892 and GM12878. To identify riboSNitches, we downloaded SNV annotations from HapMap project[23], and then converted SNV annotations from hg18 assembly to hg19 assembly using UCSC executable LiftOver. We then overlaid the hg19 SNV coordinates with our transcriptome annotation, a non-redundant combination of RefSeq and Gencode v12 transcriptome assembly, to identify the positions in the transcriptome that have SNVs. For highly confident detection of structural changes, we require that the sequencing coverage around SNV is dense, such that: first, the SNV is located on a transcript whose average coverage is greater than 1 (on average one read per base); and second, the average coverage in a 5-base window centred around the SNV is greater than 10 (average S1 + V1 ≥ 5). We exclude bases that fall within 100 nucleotides from the 3′ end of all the transcripts due to the blind tail of 100 nucleotides.

To identify SNVs with statistically significant changes in structure, we estimated a global baseline of structural change by calculating the fold differences between the doping control and SNV cumulative frequencies. We calculated a *z*-score for each detected SNV: $z = (\mathrm{StrucDiff} - \mathrm{mean})/(\text{s.d. of doped in controls})$. We used the Tetrahymena ribozyme as the doped in control. We noticed that a StrucDiff ≥ 1 is equivalent to a *z*-score ≥ 4.5 and a 100-fold difference between the SNV and doping control cumulative frequencies. To calculate the *P* value for the structural change at each detected SNV, we performed 1,000 permutations on the absolute values of the non-zero δ PARS scores within each transcript that contains SNV. This *P* value is an estimate of the likelihood that a 5-base average of the permutated PARS structural change is greater than the 5-base average of the SNV base's structural change. The false discovery rate (FDR) of the significance of the structural change at the SNV site is estimated by a multi-hypothesis testing performed using the p.adjust function in R. A SNV is defined as a riboSNitch if: first, its StrucDiff is greater than 1 (equivalent to *z*-score ≥ 4.5 and 100-fold cumulative frequency difference); second, its *P* value is less than 0.05 and FDR less than 0.1; and third, local read coverage greater than 10 and at least 3 out of 11 bases contain S1 or V1 signals in an 11-base sliding window centred by the SNV site. We also permutated the structural changes between the trio by shuffling the StrucDiffs within every transcript. After structural PARS scores were permutated, we identified only 16 riboSNitches based on the exact same aforementioned methods and thresholds. This number is less than 1% of the original number of riboSNitches found, indicating that most of the discovered riboSNitches are not random noise.

RiboSNitch noise and signal estimation. We estimated the amount of structural change between two replicates with the same sequence and compared it to the change in two replicates with differing sequences. For example, the father may have heterozygous alleles A and C at a particular locus, whereas the mother has the alleles C and C and the child has alleles A and C at the same locus. As the local genotype of the father is the same as that of the child, we can calculate the amount of structure change between that of the father and child (δ_1, noise). If this SNP was predicted to be a riboSNitch, then the local structural change between the father and mother (δ_2, signal) should be significantly greater than the noise. We took all the heterozygous riboSNitches we predicted that satisfy the above-mentioned pattern (861, 558 and 519 SNVs between three pairs of individuals in the trio), and calculated the absolute structure change in a 21-nucleotide window centred on the riboSNitch. Plotting signal (δ_2) and noise (δ_1) windows across these riboSNitches demonstrated that on average, the signal plot has threefold greater structure changes than that of the noise plot ($P = 7.94 \times 10^{-177}$, Student's *t*-test), indicating that the riboSNitches that we identified clearly distinguishes from the biological noise.

As a further control, we generated two additional biological replicates of PARS with RNase V1 from refolded RNA of the child, and obtained 70–110-million mapped reads for each sample. As expected, biological replicates of the same individual are better correlated than between individuals. No difference in variance was detected at riboSNitch neighbourhoods versus other sites, or when 5′ UTRs and CDSs were compared against 3′ UTRs. These results indicate that riboSNitches are not simply passenger mutations residing in structurally flexible or poorly measured regions.

Estimation of structural disruption at the gene level. The extent of structural disruption of a transcript is estimated by an eSDC (experimental structural disruption coeffiency) score that is defined as:

$$\mathrm{eSDC} = (1 - \mathrm{cc}) \times \sqrt{l}$$

where cc is a Pearson correlation of the transcript between two samples, and *l* is the length of that transcript[10]. The greater the eSDC is, the more disrupted the transcript is.

RiboSNitch allele-specific cross-validation. We first generated an allele-specific sequence reference for the lymphoblastoid cells by compiling 150-base sequence fragments (50 bases upstream and 100 bases downstream of the SNV) of both wild-type and mutant alleles. We then built Bowtie indexes using this reference, and mapped trimmed raw reads from GM12878 (child) to the indexes. We only accepted reads with perfect match to the wild-type or mutant sequences and calculated S1, V1 and PARS score as described above. We examined riboSNitches that were homozygous in both GM12891 (father) and GM12892 (mother), and that had both alleles detected as expressed in GM12878 (child). A riboSNitch is considered as cross-validated if the structural change between the two detected alleles in the child follows the same direction as the structural changes between the two alleles in the parents. Out of 184 homozygous riboSNitches in the parents, 117 of these riboSNitches can be cross-validated in the child (63.6%). Allele-specific cross-validation using the child's native deproteinized data was also performed as above.

RiboSNitch and microRNA RBP and splicing. Predicted miRNA-target sites (both conserved and nonconserved targets of conserved miRNA families) were downloaded from Targetscan. RBP clip data sets were downloaded from the doRiNA database[24]. In addition, CLIP sequencing data sets for *LIN28* were from ref. 25, and for *DGCR8* were from ref. 26.

RiboSNitch and splicing analysis. We defined a percent inclusion (percentage spliced in, PSI) value similarly to a previous paper[15]. We considered every internal exon in each annotated transcript as a potential 'cassette' exon. Each cassette alternative-splicing event is defined by three exons (C1, A and C2, where A is the alternative exon, C1 is the 5′ constitutive exon and C2 is the 3′ constitutive exon); two constitutive junctions (C1A (connecting exons C1 and A) and AC2 (connecting exons A and C2)); and one alternative (or 'skipped') junction (C1C2 (connecting

exons C1 and C2)). First, we constructed a reference library containing unique, non-redundant constitutive and alternative junction sequences that are based on exon annotations and their RNA sequences. These junction sequences were constructed such that there is a minimum five-nucleotide overlap between the mapped reads and each of the two exons involved. Each junction sequence was annotated with a gene name and exon indexes for downstream analysis. As we trimmed the sequencing raw reads to 50 bases, we created a junction sequence library, indexed using Bowtie-build[22], using junction sequences of 90 bases. We downloaded independent RNA sequencing data from the ENCODE consortium (GM12878, GM12891 and GM12892) to estimate the PSI differences between samples. Raw reads were trimmed to 50 bases and then aligned to the non-redundant junction sequences using Bowtie[22], with unique mapping (the -m option in Bowtie = 1) and allowing a maximum of two mismatches. The number of reads that were uniquely mapped to a junction sequence, corresponding to the junction's effective number of mappable reads, was calculated by an in-house generated script. We then counted the number of reads that were uniquely mapped to each junction C1A, AC2 and C1C2, respectively. The PSI value for each internal exon was defined as:

$$\mathrm{PSI} = 100 \times \frac{\mathrm{average(C1A,AC2)}}{\mathrm{C1C2} + \mathrm{average(C1A,AC2)}}$$

where C1A, AC2 and C1C2 are the normalized read counts for the associated junctions.

We calculated PSIs for all of the internal exons in the samples GM12891, GM12892 and GM12878 and calculated the change in PSI between each pair of samples. Out of 12,233 transcribed SNVs, 498 SNVs were found in internal exons with PSI differences in the trio, and 169 SNVs were located within 20 nucleotides of the splicing sites. We ranked these 169 SNVs by the degree of their structural changes (StrucDiff score), and found that the exons containing SNVs with higher StrucDiff scores (StrucDiff > 1) show greater PSI differences than those exons containing SNVs with lower StrucDiff scores (StrucDiff < 1).

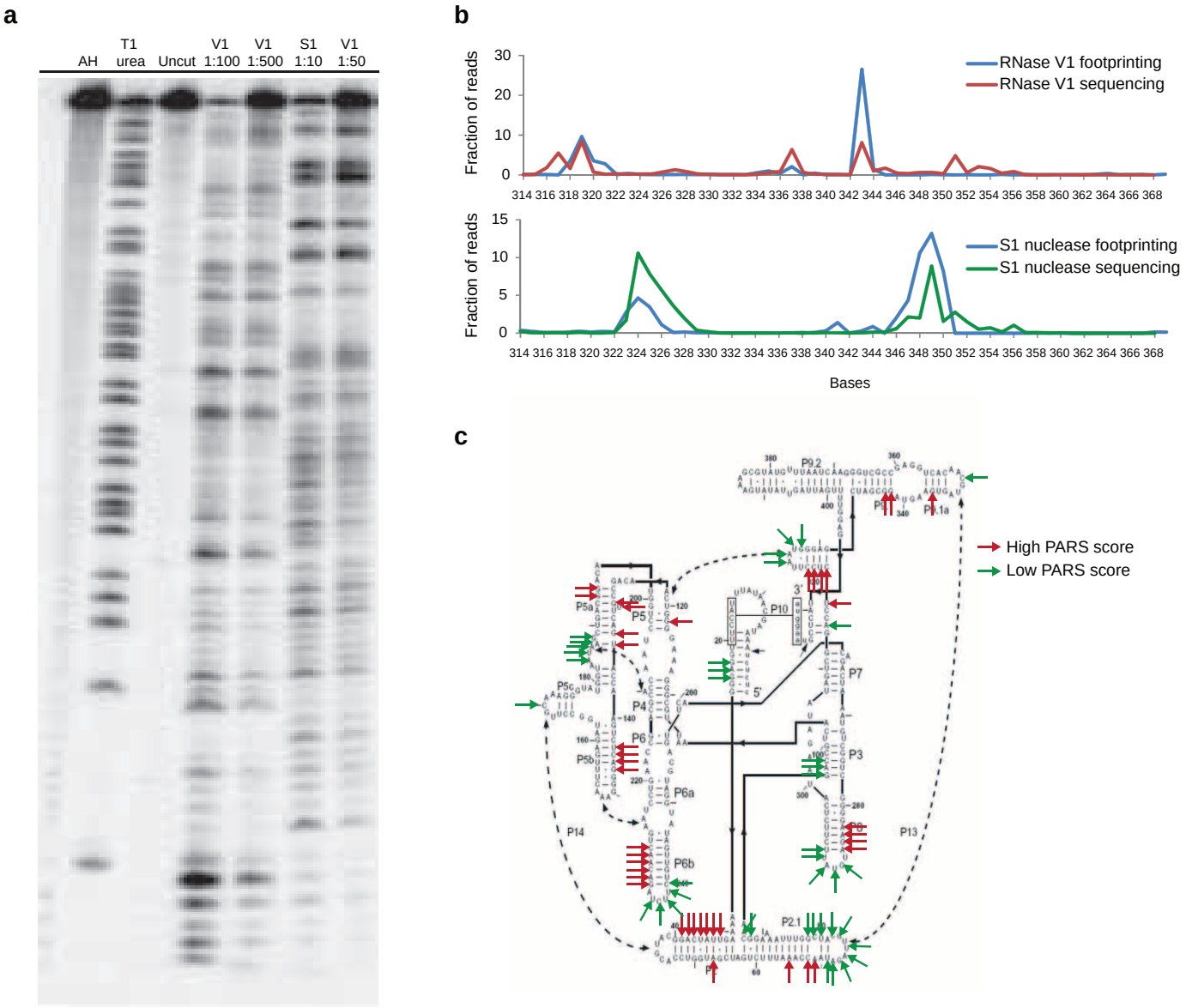
RiboSNitch and local structure environments. We defined bases of PARS scores greater than 1 as double-stranded (D), PARS scores of less than -1 as single stranded (S), and PARS scores between -1 and 1 as poised region (.). Using these cutoffs, we classified local structures around a SNV site into different categories (for example, S.D, DDD), and the average PARS-score changes for riboSNitches under different local structure categories were analysed.

RiboSNitch and SNV densities in flanking regions. We calculated the average number of SNVs within a certain distance to a riboSNitches using SNV annotation from the 1000 Genome Project. We also made the same calculation on 2,450 non-structural changing SNV sites as negative control. We used the Kolmogorov–Smirnov test to determine whether the two distributions are significantly different.

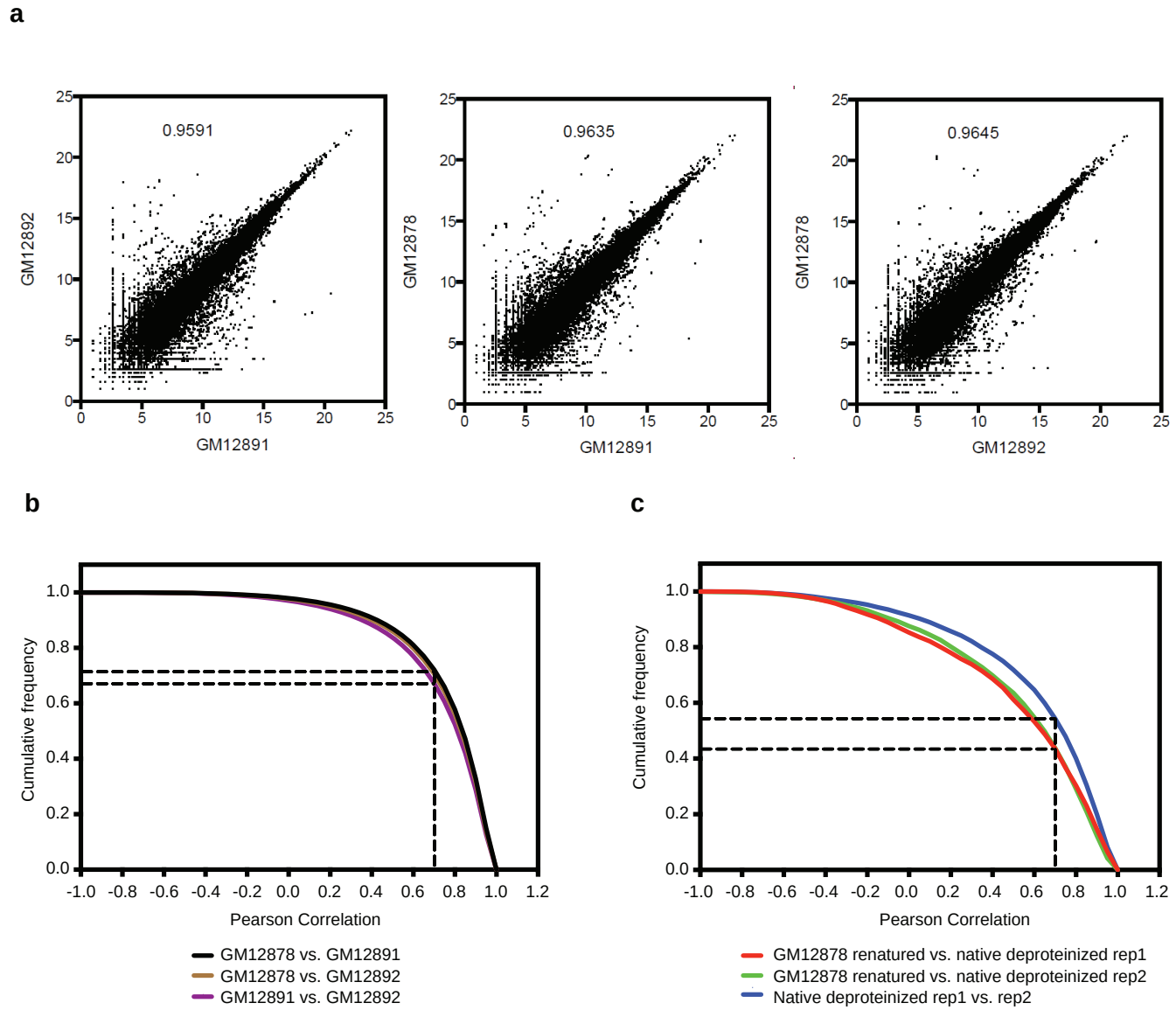
RiboSNitches predicted by SeqFold using PARS scores. For each SNV we used SeqFold to predict RNA secondary structure for a transcript fragment of 151 nucleotides (50 nucleotides upstream to 100 nucleotides downstream of the SNV sites). We used the PARS scores from allele-specific mapping as input to SeqFold. We then compared the SeqFold predicted structures for the different alleles at the SNV site. Green and red circles indicate bases with PARS scores ≤ -1 and ≥ 1, respectively.

Enrichment of SNVs in genomic features. We compared different genomic features or annotations of 993 unique riboSNitches to 1,009 control SNVs. For each genomic annotation, the fraction of riboSNitches that are inside the genomic region covered by the annotation (for example, histone mark) was compared to the fraction of control SNVs by Student's *t*-test. The different genomic annotations were downloaded and compiled from various online resources (Supplementary Table 5). A cutoff value of $P = 0.05$ was used.

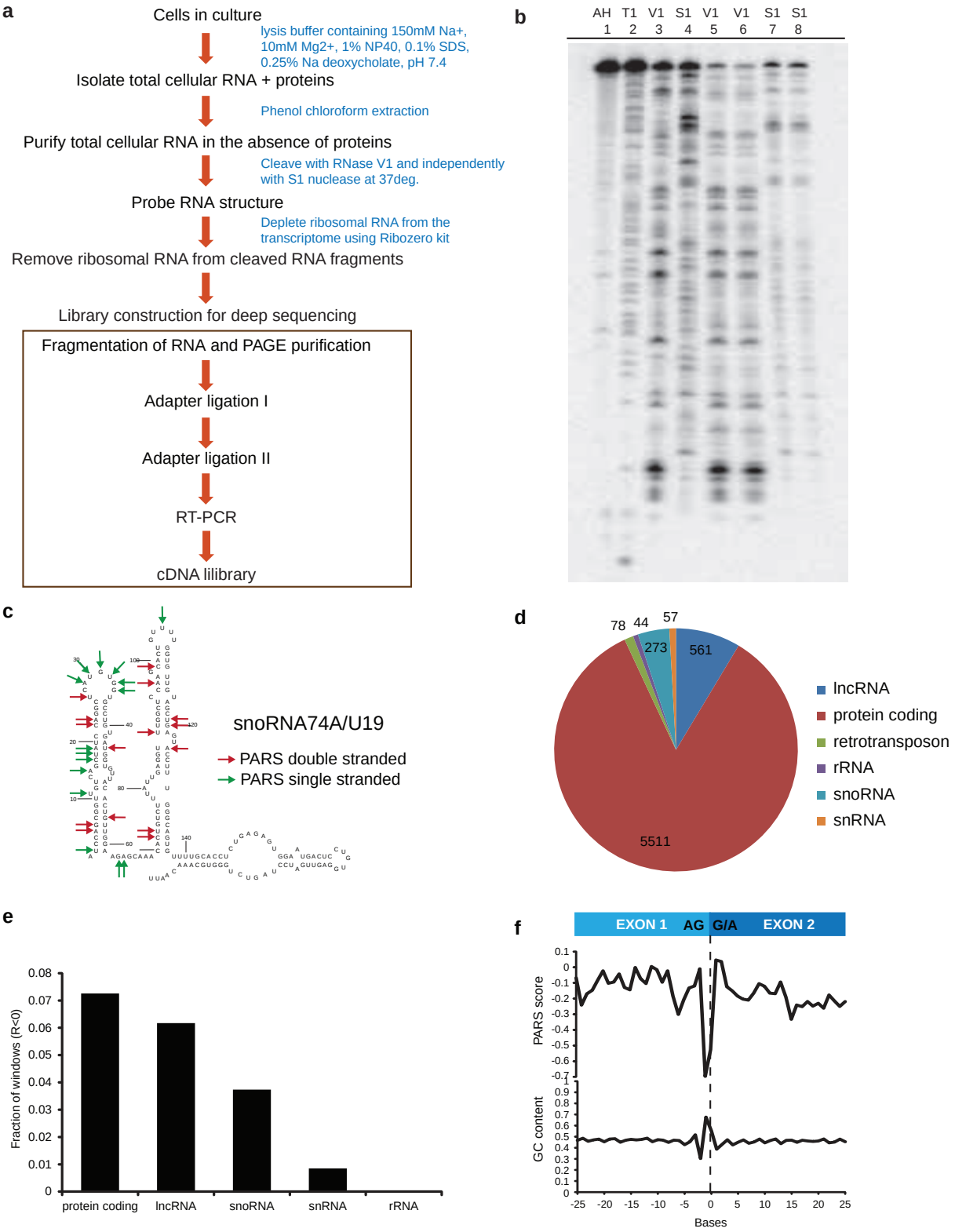
18. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature Protocols* **1,** 1610–1616 (2006).
19. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).
20. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* **460,** 479–486 (2009).
21. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Struct. Mol. Biol.* **17,** 909–915 (2010).
22. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
23. International HapMap Consortium. The International HapMap Project. *Nature* **426,** 789–796 (2003).
24. Anders, G. *et al.* doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* **40,** D180–D186 (2012).
25. Wilbert, M. L. *et al.* LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol. Cell* **48,** 195–206 (2012).
26. Macias, S. *et al.* DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nature Struct. Mol. Biol.* **19,** 760–766 (2012).



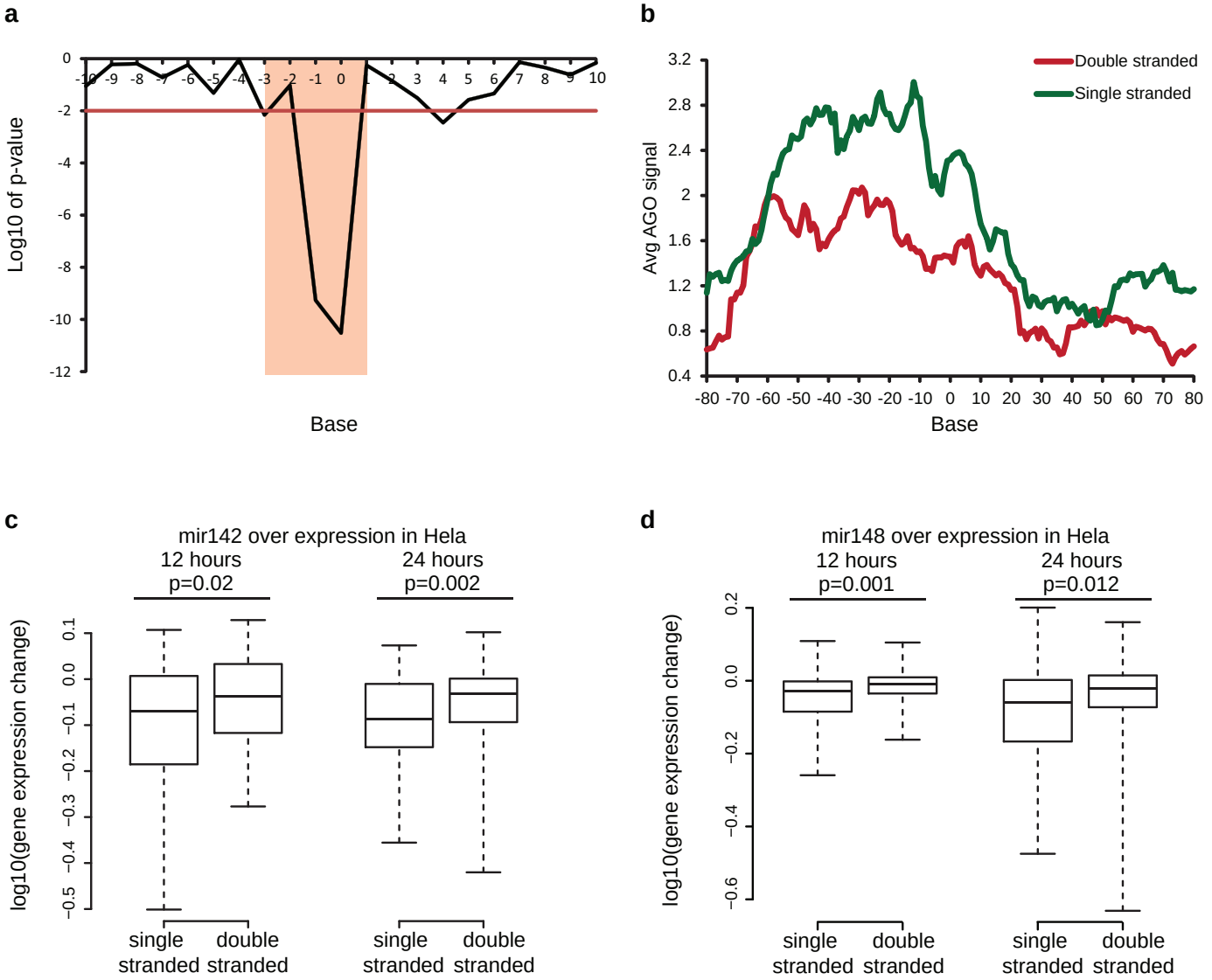
Extended Data Figure 1 | PARS data accurately maps to known structures. **a**, RNase V1 and S1 nucleases were titrated to single hit kinetics in structure probing. Gel analysis of structure probing of yeast RNA in the presence of 1 μg of total human RNA using different dilutions of RNase V1 (lanes 4, 5), and S1 nuclease (lanes 6, 7), cleaved at 37 °C for 15 min. In addition, RNase T1 ladder (lane 2), alkaline hydrolysis (lane 1), and no nuclease treatment (lane 3) are shown. Dilution of V1 nuclease by 1:500 and S1 nuclease by 1:50 results in mostly intact RNA. **b**, PARS signal obtained for the P9–9.2 domain of *Tetrahymena* ribozyme using the double-strand enzyme RNase V1 (red line) or the single-strand enzyme S1 nuclease (green line) accurately matches the signals obtained by traditional footprinting (blue lines). **c**, Top 10th percentile of PARS scores (double-stranded, red arrows) and bottom 10th percentile of PARS score (single-stranded, green arrows) were mapped to the secondary structure of the *Tetrahymena* ribozyme.



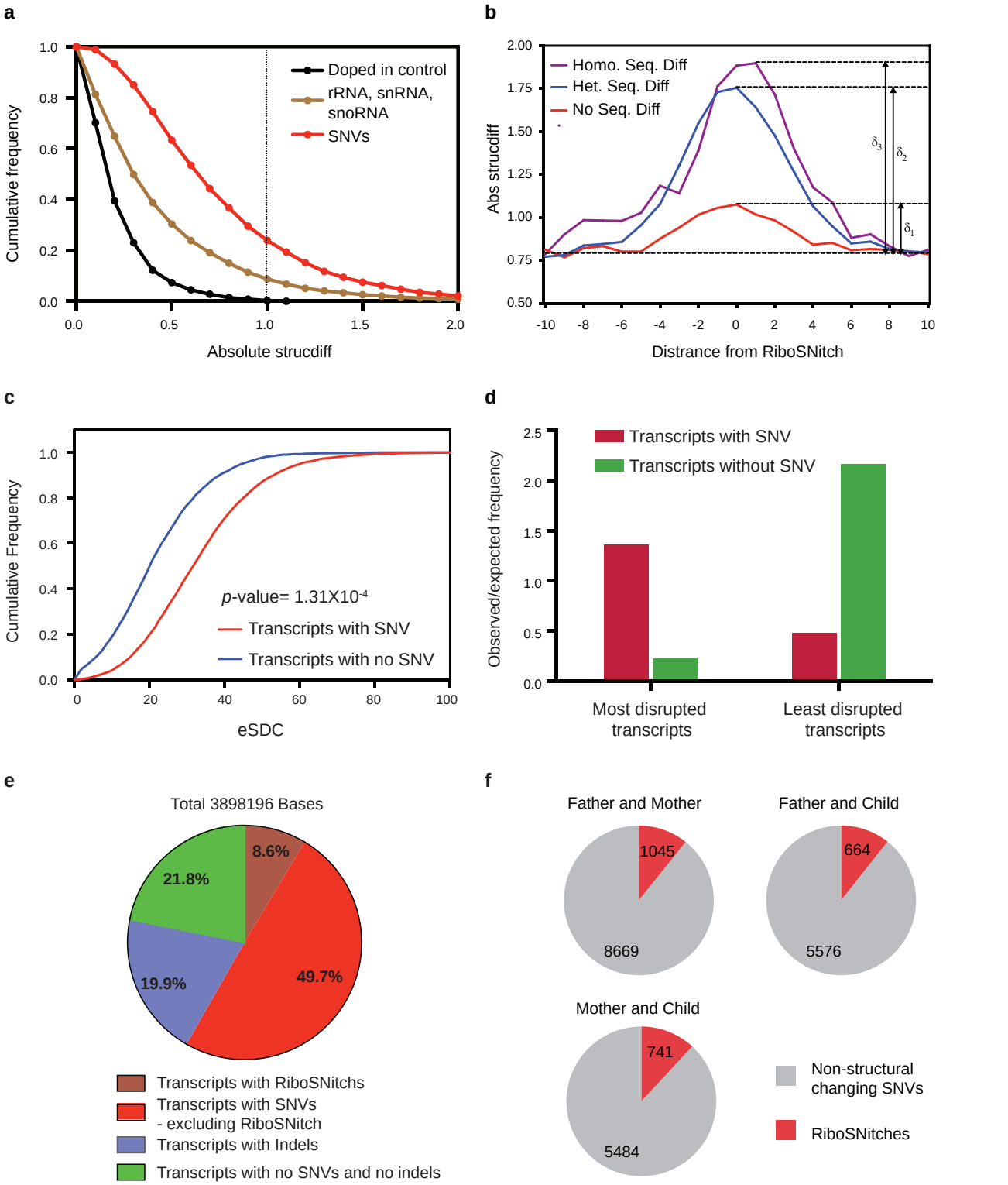
Extended Data Figure 2 | PARS data are reproducible between biological replicates. **a**, Scatter plot of mRNA abundance between the cell lines GM12878, GM12891 and GM12892 indicates that gene expression between the cells is highly correlated ($R > 0.9$). **b**, Cumulative frequency distribution of the Pearson correlation of PARS scores in 20 nucleotide windows, with a coverage of at least 10 reads per base, in transcripts between the cells GM12878 versus GM12891, GM12878 versus GM12892 and GM12891 versus GM12892. The black dotted lines indicate the fraction of windows that are positively correlated. **c**, Cumulative frequency distribution of the Pearson correlation of PARS scores in 20 nucleotide windows, with a coverage of at least 10 reads per base, between GM12878 refolded transcripts versus biological replicate 1 of GM12878 native deproteinized transcripts, GM12878 refolded transcripts versus biological replicate 2 of GM12878 native deproteinized transcripts, as well as between the two biological replicates of native deproteinized transcripts.



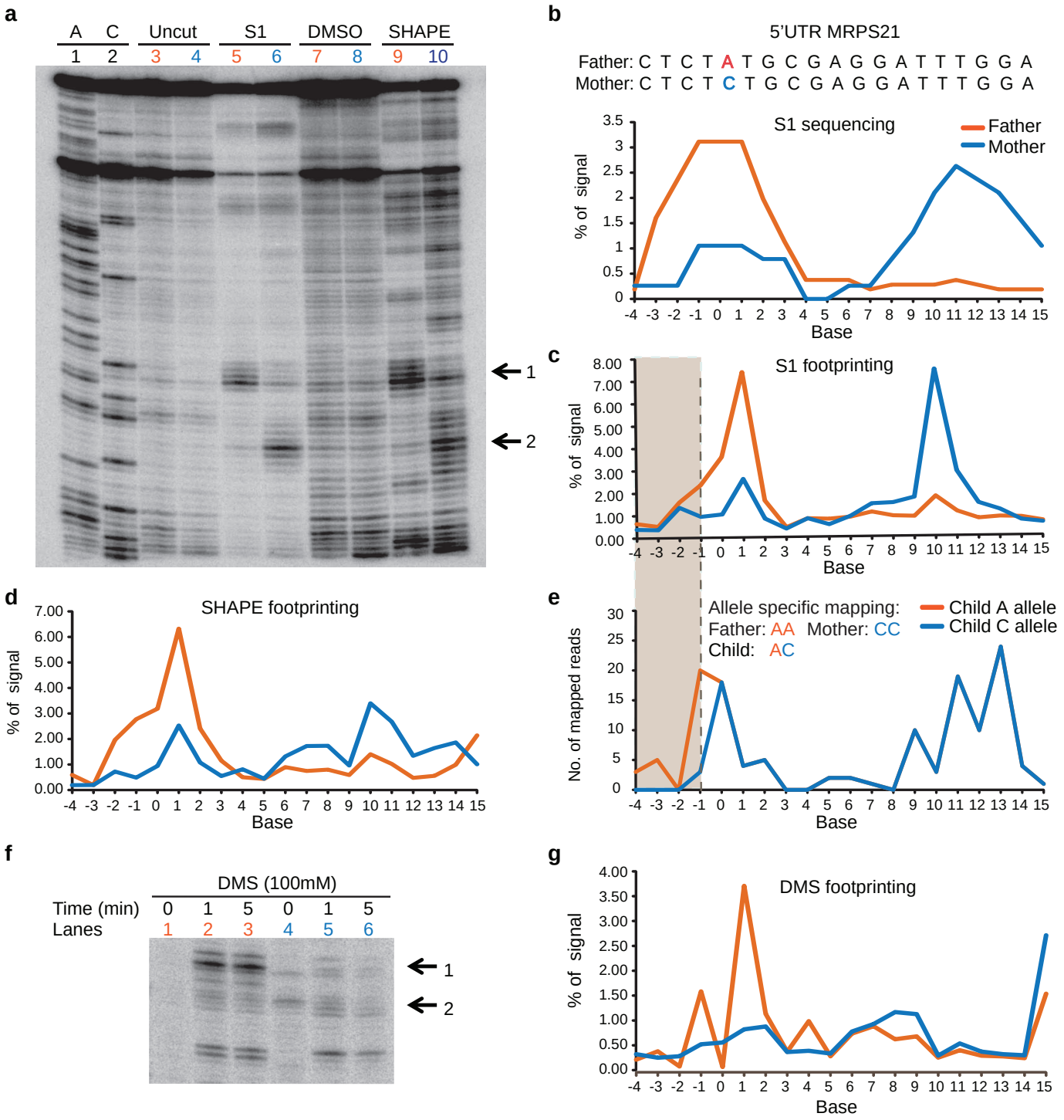
Extended Data Figure 3 | PARS can be applied to native deproteinized RNAs. **a**, Schematic of PARS on native deproteinized transcripts. **b**, Gel analysis of structure probing of yeast RNA using RNase V1 in RNA structure buffer (lane 3), RNase V1 in lysis buffer containing 1% NP40, 0.1% SDS and 0.25% Na deoxycholate (lanes 5 and 6), S1 nuclease in RNA structure buffer (lane 4) and S1 nuclease in lysis buffer (lanes 7 and 8). In addition, RNase T1 ladder (lane 2) and alkaline hydrolysis (lane 1) are shown. The enzymes appear to cleave similarly in lysis buffer and in structure buffer. **c**, Structure probing of native deproteinized snoRNA74A. Top 10th percentile of PARS scores (high, red arrows) and bottom 10th percentile of PARS score (low, green arrows) were mapped onto the secondary structure model of snoRNA74A. **d**, Deep sequencing and mapping of PARS reads on native deproteinized transcripts provided structural information for thousands of transcripts, including coding and non-coding RNAs. **e**, We compared Pearson correlations of 20 nucleotide windows with a coverage of at least 100 reads (coverage ≥ 5) between transcripts that were refolded and native deproteinized. The *y* axis indicates the fraction of negatively correlated windows ($R < 0$) over the total number of windows for each RNA class. **f**, PARS scores across exon–exon junctions, averaged across all native deproteinized transcripts (load ≥ 1). Percentage of nucleotide C plus G was averaged across the transcripts.



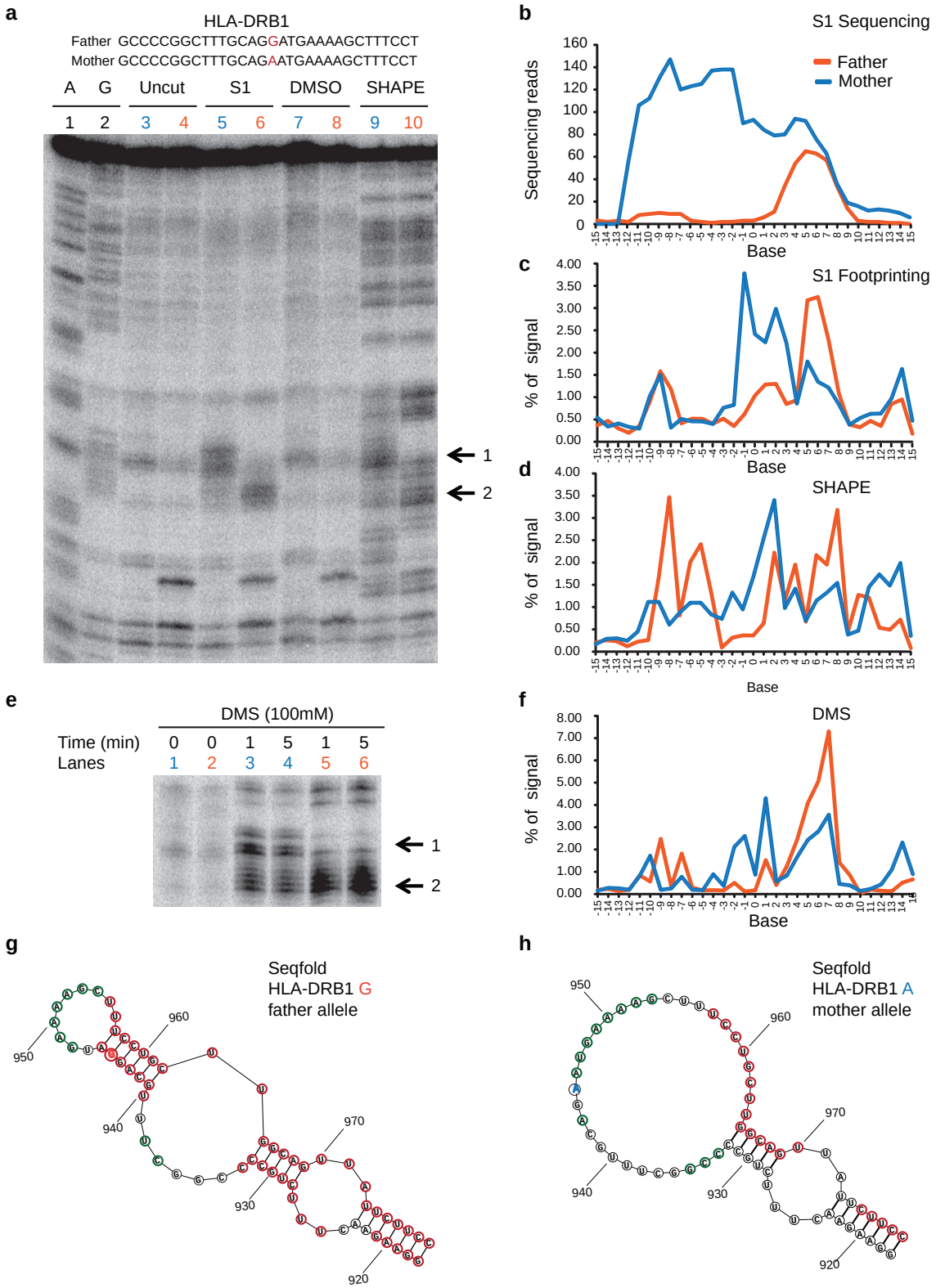
Extended Data Figure 4 | Increased accessibility 5′ of miRNA-target sites influences AGO binding. **a**, Bases that show significantly different PARS scores between AGO bound and non-bound sites in PAR-CLIP. Base 0 is the most 5′ position of the mRNA that directly base-pairs with the miRNA seed region. The *y* axis indicates $\log_{10}$ of the *P* value, calculated by the Wilcoxon rank-sum test. **b**, Metagene analysis of the average AGO-bound reads using iCLIP in predicted miRNA-target sites that are single-stranded (green) or double-stranded (red) from bases −3 to 1. **c**, **d**, Average PARS score is calculated for bases −3 to 1 for each Targetscan-predicted site. Change in gene expression is plotted for genes with most accessible (100) and least accessible (100) sites, upon overexpression of miRNA142 (**c**) and miRNA148 (**d**). *P* value is calculated using the Wilcoxon rank-sum test. Whiskers of box plots indicate extreme values.



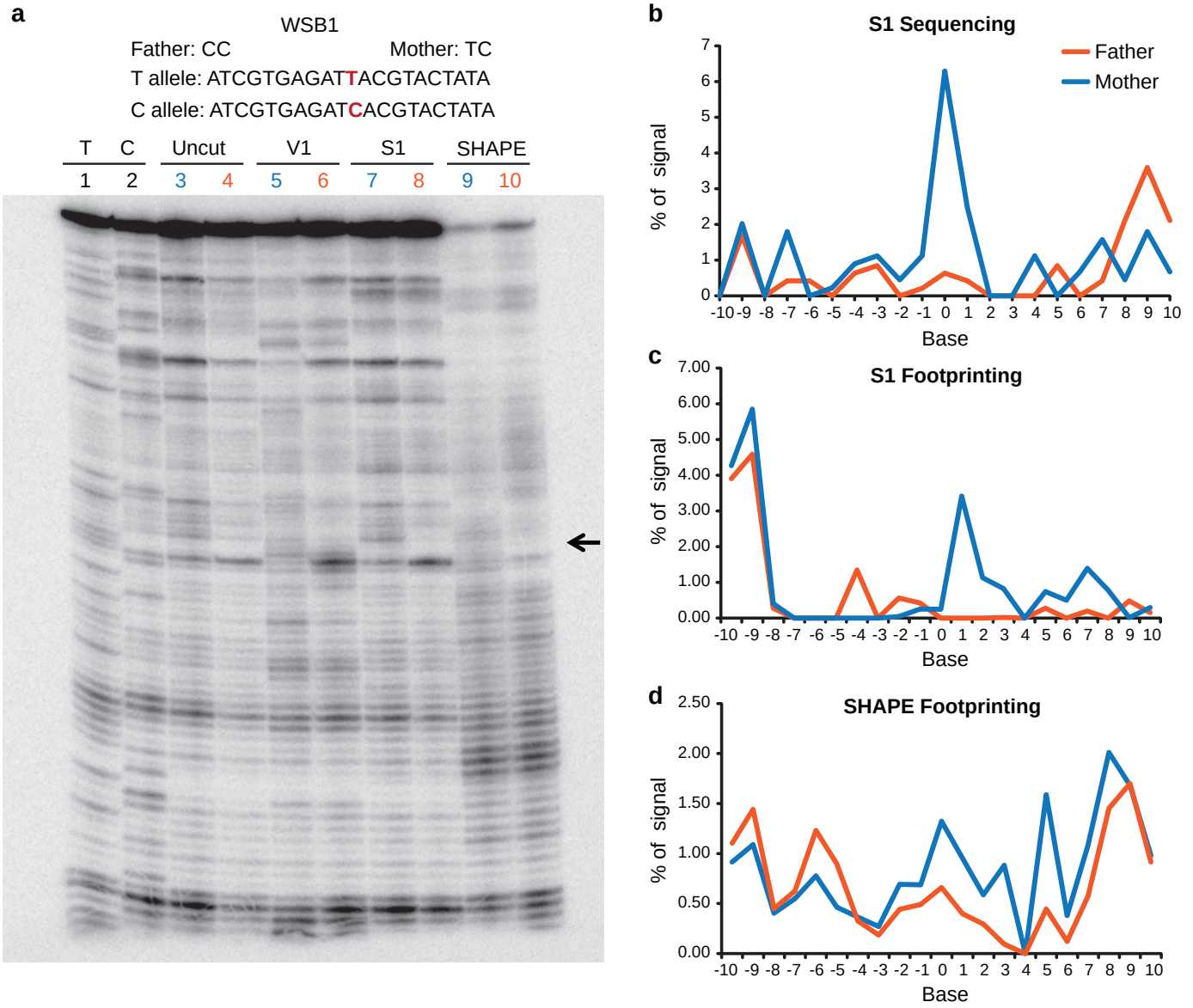
Extended Data Figure 5 | PARS identified riboSNitches in the human transcriptome. **a**, Cumulative frequency plot of PARS score differences between SNVs (GM12891 versus GM12892), doped in controls and structured RNAs including ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs). Dotted black line indicates the threshold beyond which we call an SNV a riboSNitch. The *x* axis indicates the absolute change in PARS score between GM12891 and GM12892. **b**, Absolute change in PARS score around heterozygous, homozygous riboSNitches and biological noise. The red line indicates the change in PARS score between sequences that are the same (noise) across individuals. The blue line indicates the change in PARS score between two sequences that have a riboSNitch. The purple line indicates the change in PARS score between homozygous riboSNitches. **c**, Cumulative frequency plot of the eSDC for transcripts that contain or do not contain SNVs eSDC = (1− Pearson correlation) × sqrt(transcript length). **d**, Transcripts are ranked according to eSDC score and classified into the top 2,000 most and least structurally disrupted transcripts. The most structurally disrupted transcripts are more likely to contain SNVs, whereas the least structurally disrupted transcripts are less likely to contain SNVs. **e**, Pie chart showing the distribution of structurally changing bases ($P = 0.05$, FDR = 0.1) in transcripts with SNVs, riboSNitches, indels and no SNVs and no indels. 78.2% of these bases reside in transcripts with either SNVs or indels, indicating that nucleotide sequence is important for RNA structure. **f**, Number of riboSNitches identified by PARS between each pair of individuals in the trio. Grey indicates non-structurally-changing SNVs, red indicates riboSNitches.



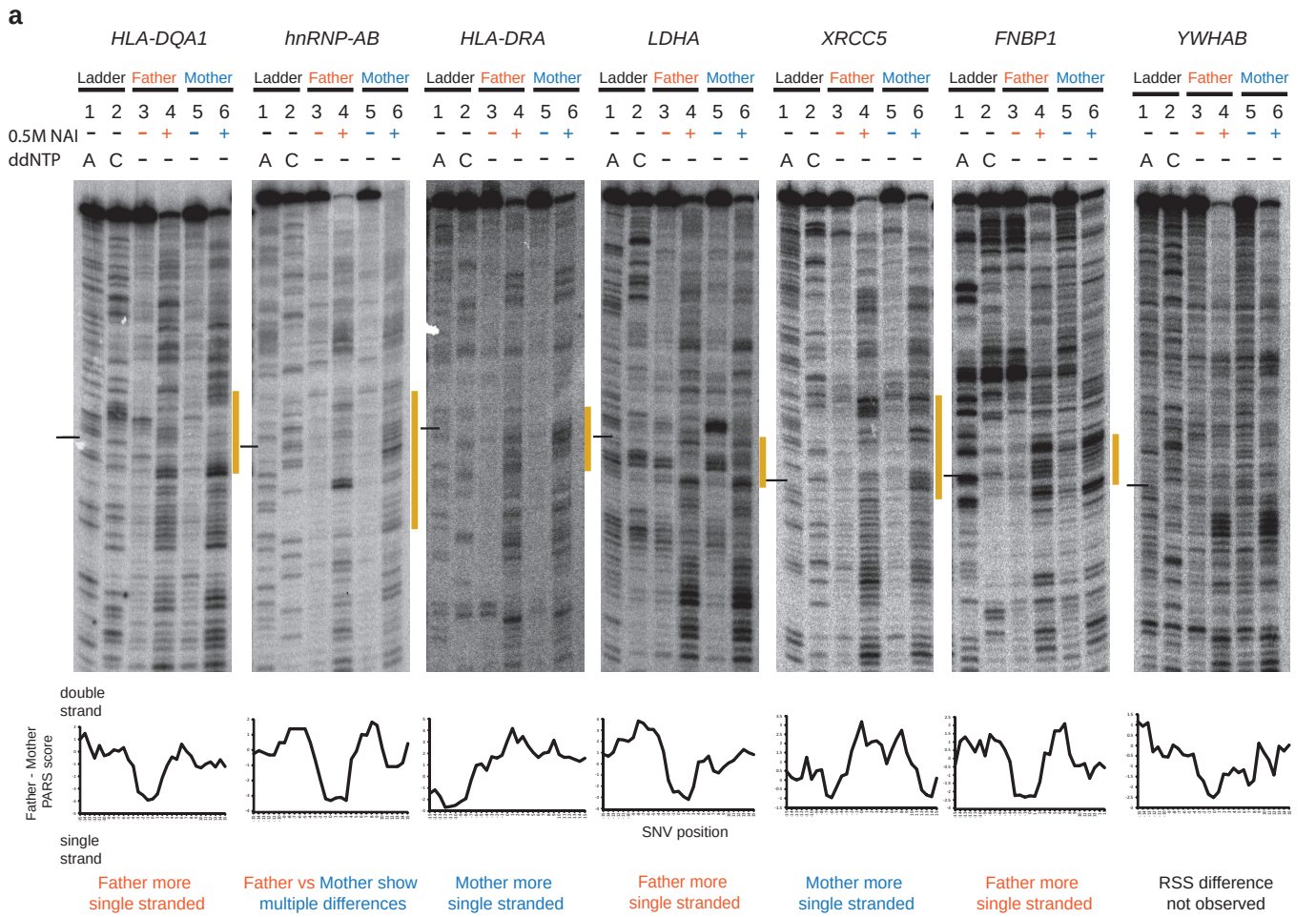
Extended Data Figure 6 | Footprinting validation of a riboSNitch in 5′ UTRs of *MRPS21* identified by PARS. **a**, Gel analysis of 150mer fragments of *MRPS21* RNA using S1 nuclease (lanes 5 (father), 6 (mother)), and SHAPE probing ((lanes 9 (father), 10 (mother)). In addition, sequencing lanes (lanes 1, 2), uncut (lane 3 (father), lane 4 (mother), and DMSO-treated lanes (lane 7 (father), lane 8, (mother)) are also shown. Black arrows indicate the change in structure between the father's and mother's alleles. **b**, Top, the sequence of a portion of the transcript containing the riboSNitch was shown. The riboSNitch is in red. Bottom, single-strand profile by S1 sequencing of the father's and mother's alleles. The *y* axis indicates the percentage of signal at each base over the total signal in the region. **c**, **d**, Semi-automated footprinting analysis (SAFA) quantification of manual structure probing of both *MRPS21* alleles using S1 nuclease (**c**) and SHAPE (**d**). **e**, S1 sequencing reads are mapped uniquely to either the A or C allele in the child. The grey box indicates the bases that show structural differences by allele-specific mapping in the child. **f**, Gel analysis of 150-mer fragments of *MRPS21* RNA using DMS footprinting (lanes 1, 2 and 3 (father), 4, 5 and 6 (mother)). Black arrows indicate the change in structure between father's and mother's alleles. **g**, Quantification of DMS footprinting of both *MRPS21* alleles using SAFA.

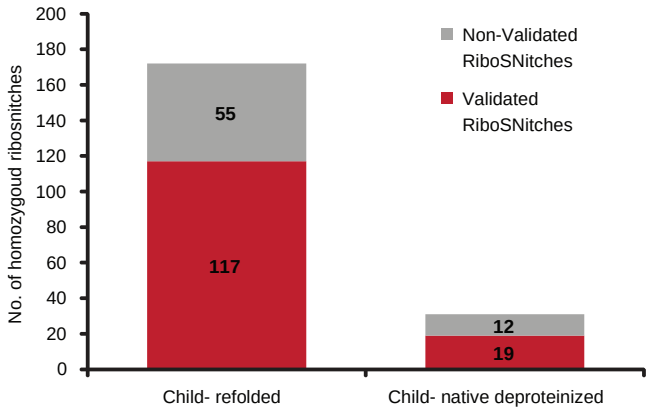


Extended Data Figure 7 | Footprinting validation of a riboSNitch in HLA-DRB1 transcript identified by PARS. **a**, The sequence of a portion of the transcript containing the riboSNitch was shown. The riboSNitch is in red. Gel analysis of two fragments of *HLA-DRB1* RNA A and G alleles using S1 nuclease (lanes 5 (mother), 6 (father)), and SHAPE probing ((lanes 9 (mother), 10 (father)). In addition, sequencing lanes (lanes 1, 2), uncut lanes (lane 3 (mother), lane 4 (father)), and DMSO treated lanes (lane 7 (mother), lane 8, (father)) are also shown. Black arrows indicate the change in structure between the father's and mother's alleles. **b**, S1 sequencing reads across the riboSNitch for both father and mother. **c**, **d**, SAFA quantification of the RNA footprinting of both alleles using S1 nuclease (**c**) and SHAPE (**d**). **e**, Gel analysis of two fragments of *HLA-DRB1* RNA A and G alleles using DMS (lanes 1, 3 and 4 (mother), 2, 5 and 6 (father)). Black arrows indicate the change in structure between father's and mother's alleles. **f**, Quantification of DMS footprinting of both *HLA-DRB1* alleles using SAFA. **g**,**h**, Secondary structure models of the G alelle (**g**) and A allele (**h**) of *HLA-DRB1*, using SeqFold guided by PARS data. The two alleles of the riboSNitch are shown in orange and blue respectively. The red and green circles indicate bases with PARS scores ≥ 1 and ≤ -1, respectively.

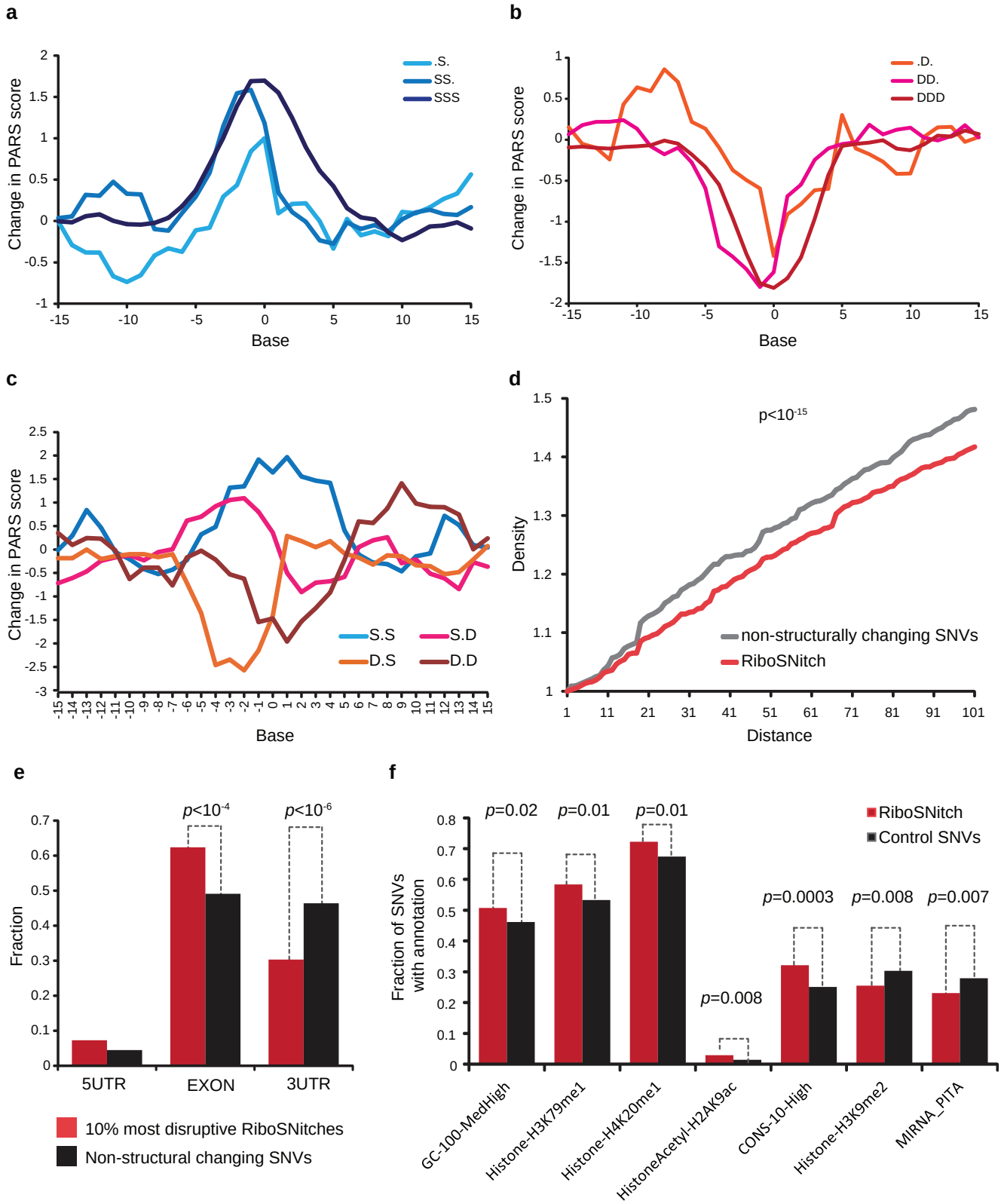


Extended Data Figure 8 | Footprinting validation of a riboSNitch in WSB1 transcript identified by PARS. **a**, The sequence of a portion of the WSB1 transcript containing the riboSNitch was shown. The riboSNitch is in red. Gel analysis of two fragments of WSB1 RNA T and C alleles using RNase V1 (lanes 5 (mother), 6 (father)), S1 nuclease (lanes 7 (mother), 8 (father)), and SHAPE probing ((lanes 9 (mother), 10 (father)). In addition, sequencing lanes (lanes 1, 2), DMSO uncut lanes (lane 3 (mother), lane 4 (father)) are also shown. Black arrow indicates the change in structure between the father's and mother's alleles. **b**, Fraction of S1 sequencing reads over total S1 sequencing reads in the region, across the riboSNitch for both father and mother. **c**, **d**, SAFA quantification of the RNA footprinting of both alleles using S1 nuclease (**c**) and SHAPE (**d**).





Extended Data Figure 9 | Additional footprinting validation of riboSNitches. **a**, Top, gel analysis of fragments of father's and mother's alleles of *HLA-DQA1*, *hnRNP-AB*, *HLA-DRA*, *LDHA*, *XRCC5*, *FNBP1* and *YWHAB* using SHAPE (lanes 4 (father), 6 (mother)). In addition, DMSO controls (lanes 3 (father), 5 (mother)) and ladder lanes (lanes 1 (T ladder), 2 (G ladder)) are also shown. The black line indicates the position of the SNV. The yellow bar along the side of the gel indicates the region that is changing between the father's and mother's alleles. Bottom, difference in PARS signal between father (GM12891) and mother (GM12892), centred at the riboSNitch. Positive PARS score indicates double stranded RNA, and should correspond to lower SHAPE signal. Negative PARS score indicates unpaired RNA with correspondingly higher SHAPE signal. Six out of seven cloned RNAs are validated by SHAPE *in vitro*. *hnRNP-AB* showed multiple differences surrounding the SNV; SHAPE data confirmed the riboSNitch and showed the structural rearrangement is more complex than indicated by PARS. SHAPE data of *YWHAB* did not show the predicted RSS difference. **b**, Bar graphs showing the number of homozygous SNVs in parents that are validated (in red) and not validated (grey) in the child by allele specific mapping. Homozygous riboSNitches between the father and mother are mapped to both the renatured child RNA (*in vitro*; child) and the native deproteinized child RNA (native deproteinized; child). As the depth of coverage is lower in native deproteinized samples, we detect fewer (31) SNVs that were homozygously different in the parents.



Extended Data Figure 10 | Properties of riboSNitches. **a**, **b**, Average PARS-score difference around SNVs that originally reside in increasingly single-stranded (**a**) or increasingly double-stranded (**b**) region. **c**, Average PARS-score difference around SNVs that were flanked by both double-stranded bases, both single-stranded bases, or one single- and one double-stranded base on each side. **d**, Density of other SNVs centred around riboSNitches versus a control group of 2,450 non-structure-changing SNVs. *P* value calculated by Kolmogorov–Smirnov test. **e**, Distribution of top 10% most structurally disruptive riboSNitches, calculated by biggest structural difference between the two alleles, versus a control group of 1,855 SNVs that do not change structure in 5′ UTRs, CDS and 3′ UTRs. **f**, Different genomic features or annotations of 993 unique riboSNitches are compared to 1,009 control SNVs. For each genomic annotation, the fraction of riboSNitches that reside in the genomic region covered by the annotation (for example, histone mark) was compared to the fraction of control SNVs by Student's *t*-test. A cutoff value of $P = 0.05$ (*t*-test) was used.

CORRIGENDUM

doi:10.1038/nature12933

Corrigendum: Primary forests are irreplaceable for sustaining tropical biodiversity

Luke Gibson, Tien Ming Lee, Lian Pin Koh, Barry W. Brook, Toby A. Gardner, Jos Barlow, Carlos A. Peres, Corey J. A. Bradshaw, William F. Laurance, Thomas E. Lovejoy & Navjot S. Sodhi

Nature **478,** 378–381 (2011); doi:10.1038/nature10425

We identified a conversion error in some of the biodiversity values used in our meta-analysis. The meta-analysis requires both mean and standard deviation values for each record, and several studies reported 95% confidence intervals instead of standard deviation. We incorrectly converted from these confidence intervals to estimate standard deviation, affecting 164 rows (7.4%) of the 2,220 rows in our database. Instead of dividing $\sqrt{n} \times$ (upper confidence interval minus mean) by 1.96, we multiplied by 1.96. Because both primary forest and disturbed forest categories were affected, there was no bias towards or against any category. After correcting for this conversion error, our results did not alter substantially (see Table 1 of this Corrigendum). All statements in our Letter remain fully supported, and most effect sizes increased, further strengthening our conclusions. N.S.S. is deceased.

Table 1 | Corrected effect sizes from Supplementary Table 1

		Original effect size	Corrected effect size
Overall		0.51	0.56
Continent	Africa	0.34	0.46
	Asia	0.95	1.07
	Central America	0.10	0.09
	South America	0.44	0.46
Taxon	Arthropods	0.64	0.69
	Birds	0.72	0.81
	Mammals	−0.12	−0.10
	Plants	0.58	0.64
Metric	Abundance	0.19	0.20
	Community structure and function	0.41	0.41
	Demographics	0.00	−0.01
	Forest structure	0.75	0.75
	Richness	0.83	0.98
Disturbance	Abandoned agriculture	1.05	1.05
	Agriculture	1.04	1.24
	Agroforestry	0.65	0.76
	Burned	0.87	0.90
	Clear-cut	2.31	2.31
	Hunted and disturbed	0.01	0.01
	Other extraction	0.59	0.59
	Pastures	0.48	0.49
	Plantations	0.50	0.65
	Secondary	0.41	0.45
	Selectively logged	0.11	0.12
	Shaded plantations	0.65	0.65

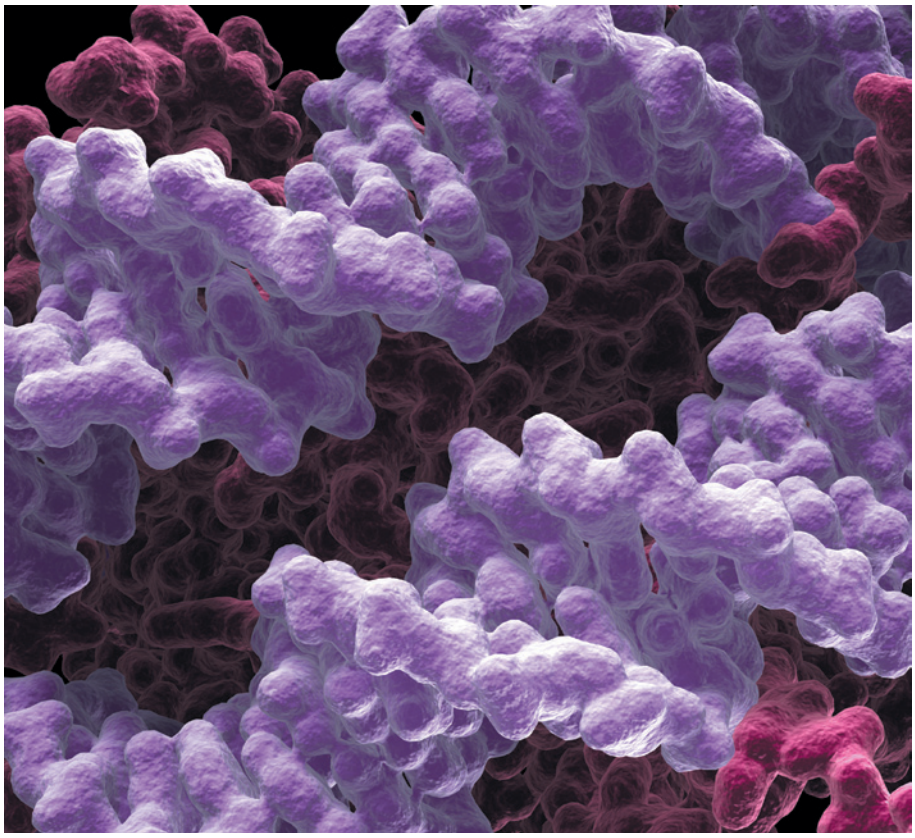
CAREERS

TURNING POINT Insight into decision-making helps neuroscientist advise on policy **p.713**

NATUREJOBS BLOG The latest on science-careers news and tips **go.nature.com/ielkkf**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

KENNETH EWARD/BIOGRAFX/SCIENCE PHOTO LIBRARY



A digital model of a nucleosome, drawn with the use of X-ray crystallography data.

STRUCTURAL BIOLOGY

More than a crystallographer

Researchers trained in X-ray crystallography are still in demand, but must diversify their skill sets to be competitive.

BY LAURA CASSIDAY

Karolin Luger was bitten by the crystallography bug during a biophysics lecture in 1986. "One person gave a talk on X-ray crystallography," she recalls. "The lecture was not that good, but the diffraction patterns were so beautiful that I thought, 'I really want to learn how to do this.'" She learned. As a postdoc, she was first author of a paper that reported the crystal structure of a DNA–protein complex called the nucleosome (see K. Luger *et al. Nature* **389,** 251–260; 1997).

Now a Howard Hughes Medical Institute investigator at Colorado State University in Fort Collins, Luger still uses X-ray crystallography to study chromatin, the DNA–protein complex that packages genomes tightly inside cells. But like most in her field in recent years, she has expanded her toolkit to include other methods.

Twenty years ago, many academic labs existed just for X-ray crystallography. Collaborators would send in samples of their molecules of interest, and labs would crystallize them and solve their structures. Nowadays, labs are much more focused on specific scientific questions, and X-ray crystallography is just one of a suite of tools that they use. Technology has improved so much that the procedure is usually no longer a full-time scientific pursuit. As 'pure' crystallography jobs dwindle, people who are trained in the technique must broaden their expertise to encompass skills such as protein expression and purification, biochemical assays and cell biology.

In fact, many crystallographers now refer to themselves as structural biologists, reflecting the variety of techniques that they use to probe molecular structure. They may have PhDs in biophysics, biochemistry, bioinformatics or computational biology, and find work in academia or industry. But they are united by a desire to 'see' the invisible molecules that make up cells. Those structures, often breathtaking in their beauty and intricacy, provide important clues about functions or sites that might serve as drug targets.

CRYSTALLIZING THE HISTORY

X-ray crystallography has been around for about a century, since scientists realized that atoms in a crystal could diffract X-rays, producing a pattern of spots on a detector. The angles and intensities of the diffracted beams reveal the structure of molecules.

Until recent decades, only specialists with years of training and expensive equipment could perform X-ray crystallography. But in the 1990s, the technique became much more accessible. As synchrotrons — large, ring-shaped particle accelerators that produce powerful X-rays — spread across the globe, researchers could take or send their crystals to the synchrotron facilities, where resident experts guided them in collecting data and interpreting results. The automation of crystallization, improvements in methods for solving structures and a boost in computing power greatly sped up the process, giving researchers time for other scientific pursuits.

Increased competition for research grants also forced crystallography labs to become ▶

more well rounded. Instead of just solving one structure after another, researchers must now link the structure of a molecule to its function through biochemistry and cell-biology experiments. "It's no longer enough to conjecture about the function of a particular protein. You have to test it," says Wayne Hendrickson, who specializes in biochemistry and molecular biophysics at Columbia University in New York.

The story of major crystallography projects such as the Protein Structure Initiative (PSI), supported by the US National Institute of General Medical Sciences (NIGMS), encapsulates the evolution of the field. The PSI has solved more than 5,300 distinct protein structures and spurred innovations in crystallographic methods. Last year, however, NIGMS director Jon Lorsch, acting on the counsel of an advisory panel, decided that the project had run its course, and it will terminate on 30 June 2015 (see *Nature* **503,** 173–174; 2013).

Critics argued that many of the structures that the PSI has solved have little relevance to important biological and medical problems, and that PSI scientists did not adequately poll the biological community to select interesting targets. In addition, such 'big science' programmes consume precious funds that, in the minds of some, would be better spent on individual researcher grants.

Despite the PSI's closure, Hendrickson, whose lab specializes in membrane proteins and was part of the initiative, says that it is too early to gauge the impact on crystallography job prospects. "It will depend on whether PSI centres like ours are able to gain alternative means of support to keep things going," he says. His centre, the New York Consortium on Membrane Protein Structure, is applying to other research organizations and foundations for grants.

TRIAL AND ERROR

Crystallography work increasingly requires a good scientific question rather than just solving structures — something Sheena D'Arcy knows well. As a graduate student, she worked in a crystallography-only lab. "For my postdoc, I wanted a lab that was a bit more driven by scientific questions," she says. She is now working with Luger, using crystallography — and other methods — to study how DNA is packaged into chromatin.

Early in her postdoc, D'Arcy recognized the value of approaching a problem with multiple techniques. She wanted to obtain a crystal structure of nucleosome assembly protein 1 (Nap1), which helps to package DNA in the cell. But she could not get the protein complex to crystallize. And so, while still working on crystallization on the side, she tried an alternative technique — hydrogen–deuterium exchange mass spectrometry. That provided important insights into the structure, and D'Arcy published a paper on it (S. D'Arcy *et al. Mol. Cell* **51,** 662–677; 2013). She says that anyone who is interested in structural biology should consider learning this technique, as well as nuclear magnetic resonance (NMR) spectroscopy.

FRESH APPROACHES

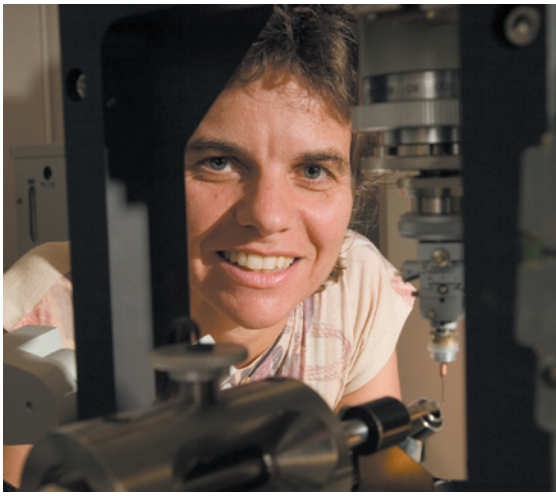
Now that synchrotrons are widespread, crystallography labs no longer need their own expensive X-ray facilities. Luger's lab does retain an X-ray generator for quickly screening crystals and training students; the device is powerful enough to collect publication-quality data from well-ordered crystals that diffract well, but non-ideal crystals or those that are quickly degraded by X-rays are sent to a synchrotron, says D'Arcy. The team has access to a beamline — a path of X-rays coming off the accelerator — at the Advanced Light Source synchrotron at Lawrence Berkeley National Laboratory in Berkeley, California.

The crystallography purist who prefers not to dabble in other techniques might consider a career as a beamline scientist, loading crystals for researchers and overseeing them as they collect data. As well as permanent positions, many synchrotrons offer training programmes in crystallography. They also offer summer programmes and internships for students, postdocs and other researchers who want to learn the technique but lack their own X-ray facilities.

The European Synchrotron Radiation Facility (ESRF) in Grenoble, France, offers a six-week Summer Bachelor Programme for undergraduates, which includes lectures, tutorials, lab work and site visits. The Cheiron School at the SPring-8 synchrotron in Harima, Japan, has ten-day training sessions for graduate students, postdocs and young scientists who wish to pursue careers in fields that involve synchrotron radiation. And the Advanced Photon Source in Argonne, Illinois, presents an annual two-week National School on Neutron and X-ray Scattering, in which graduate students attend lectures and tutorials and conduct short experiments.

"Taking the time to sit down and teach yourself the theory and computer programs is going to pay in the long run."
Sheena D'Arcy

Alexei Bosak began working at the ESRF as a postdoc and is now a beamline scientist. His duties are split between his own research interests in materials science (he has beam time reserved for his own experiments) and the research of ESRF users. "The people come, and we have to make them happy running the experiments," Bosak says. "Sometimes we are less involved, and sometimes we are more involved. But quite frequently a collaboration results."



Karolin Luger, a researcher in X-ray crystallography.

JOHN EISELE, COLORADO STATE UNIVERSITY

NEXT GENERATION

Structural biologists are developing methods to expand the capabilities of conventional X-ray crystallography, with potential implications for future practitioners. In November 2013, the US National Science Foundation (NSF) awarded a US$25-million Science and Technology Center Grant to the University at Buffalo in New York and seven partner institutions to fund the BioXFEL research centre. The centre will further the use of recently developed tools called X-ray free-electron lasers (XFELs) that produce much shorter and more intense pulses of X-rays than synchrotrons (see page 604).

According to Eaton Lattman, a structural biologist at Buffalo and director of the BioXFEL, XFELs can analyse crystals that are 1,000 times smaller than those required for conventional X-ray crystallography. "This opens up a whole new universe of protein molecules for crystallography that we couldn't do before because we couldn't grow big enough crystals," he says. The intense X-ray pulses can also capture frozen images of molecular motion, opening the door for dynamic studies and molecular movies.

The BioXFEL centre will make use of an existing facility at the SLAC National Accelerator Laboratory in Menlo Park, California, among other facilities. A smaller XFEL facility began operating in Harima, near the SPring-8 synchrotron, in 2011. And a larger one is scheduled to open in Hamburg, Germany, in 2015.

Lattman anticipates that the NSF grant will result in a "modest number" of new jobs at member institutions. "Right now, we're really limited by the amount of beam time that is available," he says. "If we start to see more countries around the world building XFEL facilities, then I think we'll see growth in the field comparable to what we saw for traditional crystallography in the 1990s." For now, the field of XFELs

needs technical improvements, such as better data-processing software and specimen delivery systems.

EXPERTS NEEDED

Ironically, the very diversification in skills now required to obtain an academic job has arguably turned many structural biologists into jacks of all trades, masters of none. Today's researchers are accustomed to sending crystals to synchrotrons for analysis, and computer programs perform the analytical work. "To solve a straightforward structure, you really don't have to understand the theory and the maths, and that's a bit of a pity," says Luger. "I'm a little worried that we're running out of people who know how to handle problems or complex situations."

Bosak notes that positions related to crystallography are frequently available at ESRF, and that they are hard to fill. "It's very difficult to find a good crystallographer these days," he says. Beamline scientists must have a thorough understanding of crystallography theory and instrumentation, skills that many modern training programmes do not emphasize. This means that a crystallographer with the right skill set can find that he or she is in demand.

There is also a growing list of contract companies that specialize in crystallography. Firms such as Proteros Biostructures in Planegg, Germany; Shanghai Medicilon in China; and Emerald Bio in Bedford, Massachusetts, provide full-service crystallography to clients, many of which are pharmaceutical companies. The firms employ scientists at bachelor's, master's and PhD levels to carry out all steps of crystallography, from protein design to structural analysis. But pharmaceutical companies such as Merck, based in Whitehouse Station, New Jersey, and Novartis, based in Basel, Switzerland, still have their own crystallography programmes centred on structure-based rational drug design, which also employ scientists at all levels. These companies are potentially a better fit for those who wish to focus on a specific protein or biological process rather than a plethora of them.

D'Arcy advises students with an interest in X-ray crystallography to take the time to learn its theoretical underpinnings and all the techniques involved. "Don't let people do things for you," she says. "There are a lot of senior people who know how to do things, and there's always a time crunch to get data — you get crystals, and you just want to see the structure. Taking the time to sit down and teach yourself the theory and computer programs is going to pay in the long run — because you really learn when things go wrong." ■

Laura Cassiday *is a freelance writer based in Hudson, Colorado.*

TURNING POINT
Nicholas Wright

As a student, Nicholas Wright pursued interests in biology and public policy, securing four degrees and a fellowship in the department of government at the London School of Economics (LSE). He now uses his neuroscience training and insights into human decision-making to inform nuclear-security policy as a fellow at the Carnegie Endowment for International Peace in Washington DC.

CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE

Did you always have dual interests?
Yes. I went straight to medical school at University College London (UCL), but I also did a year at Imperial College London studying health policy and management, which proved a turning point. While there, I did research in Chile on how best to incorporate scientific findings into clinical medicine. I learned that, to be effective, public policy must always take cultural and organizational factors into account; and I learned how best to ask questions so that they are relevant to public policy.

How did you combine your interests?
At the end of my medical degree, I went to a series of lectures by economist Richard Layard from the LSE, who talked about what neuroscience might be able to tell us about economic and social decision-making. I read up on neuroscience and decided to do a master's degree. My research into functional magnetic resonance imaging (fMRI) dispelled the hypothesis that only one area of the brain specializes in reading. The technique surpassed my expectations and proved itself to be a new source of information that could be relevant to public policy.

How did you delve into decision-making?
It wasn't by chance. After my postgraduate medical exams, I did a PhD project to study how risk perception influences decision-making, hoping to apply the concepts to issues of public policy. I worked with the Wellcome Trust Centre for Neuroimaging at UCL and stayed on as a fellow doing fMRI after I finished my PhD.

How did you position yourself for a policy job?
During a year-long fellowship at the LSE, I built up my contacts, planned events with policy-makers and created a narrative about my experience. Several policy-oriented job opportunities in Washington DC came up, but a position at the Carnegie Endowment for International Peace was most exciting.

What appealed to you about that post?
There was a lot of great work done in the 1970s on applying decision-making and cognitive psychology to nuclear strategy, but much less had been done recently. The ideas coming out of neuroeconomics hadn't yet been applied to international relations, so there was enormous potential for doing interesting work that could have a positive impact on the world.

Has your work had real-world impact?
In January, a colleague and I published an article called 'The neuroscience guide to negotiations with Iran' in *The Atlantic*. We combined insights from neuroscience, behaviour and history to better understand Iranian motives in the ongoing nuclear talks. For example, conciliatory gestures are more effective when they're unexpected. Neuroimaging experiments detail how the brain computes the difference between what is expected and what actually happens, and the more surprising the reward or punishment, the more impact it has on decision-making. Last year, Iranian President Hassan Rouhani unexpectedly used social media to engage on political issues, raising hopes for a diplomatic breakthrough. We argued that neuroscience provides a new, important source of evidence relevant to nuclear talks with Iran. Our article was read by US and UK defence policy-makers, and I have been asked to continue providing briefs to the US Department of Defense.

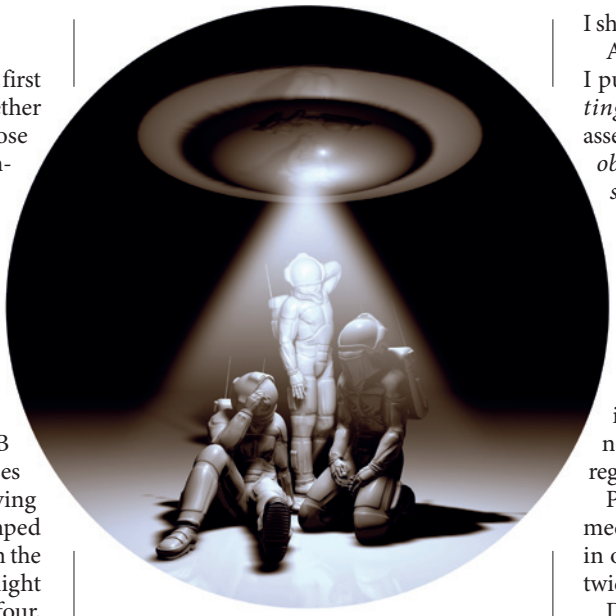
Do policy-makers value a science background?
In the world of public policy, there are so many competing priorities that there is a limit to how much science can be used. Winston Churchill once said that scientists "should be on tap, but not on top". Although science is not the only consideration, I am on tap to provide it. ■

INTERVIEW BY VIRGINIA GEWIN

ME AND MY FLYING SAUCER

A cosmic ride.

BY IAN WATSON



JACEY

Fellow people, this is probably your first ever sight of a genuine and altogether shimmering flying saucer! All those thousands of books about close encounters, abductions by aliens, things seen in the sky and in cornfields, are complete cow pies. Am I cleared for landing? I have something on board you'll be glad to see. Namely your four marooned Martian astronauts, sound in wind and limb. Juno, Jim, Chuck and Barbara won't suffocate on Mars on account of their ascent engine failing.

I couldn't possibly let J, J, C and B stifle in their tin can, even if this exposes my UFO to becoming an identified flying object. I could have clandestinely dumped the four of them anywhere hospitable in the world, near some highway, but that might have led to personal problems for the four, such as conspiracy theories, or getting disappeared. So I chose a public approach — though at the same time still fairly anonymous, like a caped crusader plus mask. My passengers haven't set eyes on me. For all they know, I might be an AAI, an alien AI.

Oh, and I shan't be hanging around for long, an inch above the landing strip, so don't bother breaking out Château Rohypnol or dozed-up 7 Up on my account. As soon as the heroic quartet are safely on the ground, I'll be buzzing off promptly. And don't bother chasing me with those jets. I can do Mach 100 without batting an eyelid. Plus: cloak of invisibility. Just keep well back from this saucer, as you've no idea what energies it deploys. That includes your fire engines and ambulances with SWAT teams on board.

Your Mars quartet left their bags of rocks and dust back on the red planet; I didn't want mess in my hold. Oops, sorry, people, if I seem to be patronizing the brave quartet who took four months to get to Mars, whereas I brought them back in four hours, but it does take even little me several days to get to Neptune; I have limitations. As do fairly featureless blue Neptune and extremely chilly Triton: –235 °C.

Why spend several days going to Triton? Why, to patrol the Far Frontier! Admittedly, there's an even farther frontier out in the Kuiper belt, not to mention the Oort cloud beyond, but … I confide nothing about refuelling my power system. Get going now, bold astronauts, out of the hold with you!

And of course the orbit of Neptune implies a very large *sphere* of mostly empty space, but I'm not heading way out to gape at stars or to view the Sun as a very bright spot.

If you're already trying to do voice print recognition, just in case you strike lucky: waste of time. This is a synthvoice, although totally naturalistic.

Ah, Juno, I see you're outside. Skedaddle, lady! Don't make such a meal of it. You aren't *that* heavy on Earth.

How come I have an authentic flying saucer? I've been wondering whether to say, as this seems such a wet dream, pardon, such a fantasy, for a young geek like me.

It all began when I took a bit of tech of my own devising to a park at dusk. *Don't* jump to the conclusion that this might be a 4-qubit iPhone. I tapped in a very long number, nothing to do with *pi*, and I got a surprisingly swift response in the form of a glowing little globe rising lazily like a luminous golf ball and drifting towards me. I guess I happened to be in the right place, unless the rest of the world and the ocean depths are littered with globes imitating golf balls.

This does rather suggest, don't you think, that the earliest arrivals on our world from the Outside postdate our fourteenth century, although I suppose earlier arrivals might have masqueraded as fruits or nuts …

Now you, Chuck, stop your loitering. All the way home I was gradually increasing the grav in the hold to re-accustom you, oops, maybe I shouldn't have said that.

As the mini-globe hovered before me, I put my hand upon it tentatively. *Tingle tingle tingle*. The gizmo read me and assessed: *highly intelligent, atheist, ingenious, obsessional, dedicated, responsible, loves solitude, bold but not rash*, et cetera … I cottoned on instantly.

Evidently here was a very compact von Neumann first-contact gizmo. If you don't know what a vNfc gizmo is, you may as well stop listening. I guess the vNfc gizmo could have communicated with me efficiently in Korean, if I happened to be Korean. Although not Koranian — atheism seemed highly regarded in the aptitude test.

Proposal: would I accept to be the intermediary between *Homo sap sap* (so brainy, in our opinion, that we named ourselves twice) and the Outside intelligences?

Did Darwin go to the Galapagos Islands? Enough said.

Ah, Jim, I was wondering when. Lug your legs after Juno and Chuck. Bye-bye.

Basically my duties are to keep watch and transmit *by the Outsider way* — bet you'd like to know what that is — summaries of significant up-to-date Earth news. Likelihood of nuclear war or other global catastrophes, breakthroughs in nanotech or a star-drive. And no, I don't use tachyons to transmit. But I like the idea that the cosmos is non-causal, deep down; could that be a clue?

And then there was one … Come along now, Barbara. One small step for a lady mission commander. You know you can do it.

You *aren't* going to leave? Do you imagine you'll stall me here till someone fires a big net of green kryptonite over my UFO?

The ingratitude!

I swear I'll shut the hold, take off in 30 seconds, counting, and forcibly eject you — well, it can't be on Triton 'cause of Nep's radiation belt, but, damn it, back on Mars. Enough food, water and air, and I'll resupply every couple of months, but it'll be marooned on Mars for you — and no, I don't need a lady friend, who might take control of my UFO, even if you *are* an astronaut. I have higher, cosmic priorities.

Hatch closed, here we go. Can't say I didn't warn you … Bye, folks. *Whooooooosh!* ■

Ian Watson's *most recent novel is* The Waters of Destiny *(with Andy West). The first volume is free to download from www.watersofdestiny.com.*

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm