THIS WEEK

Be here now

New views of quantum theory that can be tested and have practical applications bring welcome echoes of physics past.

The phrase "Shut up and calculate" is popularly associated with the Cornell University physicist David Mermin, who coined it to describe how many physicists use the mathematics of quantum theory without thinking too hard about its deeper implications.

Mermin himself is clearly not content to shut up and calculate. He has mused on the meanings of quantum theory and of physical theory more generally for many decades, notably in his Reference Frame column for the magazine *Physics Today*. Now, on page 421 of this issue, he considers a new view of quantum theory, called quantum Bayesianism or QBism, and what it implies for a long-standing question. Namely, how can we reconcile our perception that the present moment is special with the relativistic view that space-time is a continuum that reaches from past to future, with nothing to privilege 'the Now'?

Some will see this as a metaphysical question. Certainly, while the gulf remains between the formal machinery of quantum mechanics, with its wavefunctions and probabilities, and our conscious experience of the world, it is hard to see how the question can be framed with the rigour that science usually demands.

Yet QBism offers a way to put us in the picture, even in the absence of a theoretical link between the abstract microworld and the subjective macroworld. It suggests that quantum theory is telling us what an individual can know about a system in the light of what he or she already knows and expects, just as in standard Bayesian probability.

The idea has been given something of a rough ride by physicists, who seem uncomfortable with QBism because they see it as a solipsistic view of the world. Perhaps Mermin's advocacy will secure it a more sympathetic hearing. At any rate, it has the virtue of refusing to ignore quantum theory's long-standing tussle with the role of the observer.

Aside from the merits of the idea, it is striking that Mermin should discuss it at all. Any view of quantum (or indeed classical) physics that borders on the metaphysical has long been out of fashion. Yet the early architects of quantum theory, such as Niels Bohr, Werner Heisenberg and Max Born, had no reservations about examining the philosophical issues it raised, and the problem of 'the Now' troubled Einstein.

Most famously, Bohr and Einstein argued about whether quantum mechanics allowed any room for the idea of realism — of an objective world that exists independently from our efforts to observe and measure it. Bohr insisted that physics was concerned with what we can know, and was silent on the matter of 'how things really are'. He, Born and Heisenberg made claims about quantum theory's challenge to causality and determinism that today look like a bit of an intellectual stretch.

Mermin is not alone in admitting such debates back into science. In a Perspective article on page 443, physicists Artur Ekert and Renato Renner place current work on quantum cryptography in a broader context that encompasses the thorny concept of free will.

Quantum methods of encoding information, they argue, combined with "an arbitrarily small amount of free will are sufficient to conceal whatever we like". To enable foolproof secrecy, "free will is our most valuable asset". This, too, is the kind of claim that a few decades ago would have risked being dismissed, if not ridiculed, as idle coffee-room chat. It reveals researchers' new boldness for engaging with the meanings and corollaries of quantum physics. In doing so, they enrich the discussion as Bohr, Einstein and their colleagues did.

"Researchers have a new boldness for engaging with the meanings and corollaries of quantum physics."

But why now? Interest in the foundations of quantum theory — what it really tells us about the character of the world that we experience in reassuringly classical terms — has flourished since the late 1980s. That has been driven partly by the development of experimental techniques, especially in quantum optics, that can test ideas about phenomena such as entanglement (the codependence of remote quantum states), measurement and wavefunction collapse, and which were previously accessible only to theoretical speculation. As a result, physicists can more clearly see the most fundamental features of quantum theory — in particular the nonlocality and contextuality (contingency on how results are obtained) of quantum systems.

The other driver is an emphasis on quantum theory as a theory of information: of what we can know, transmit and share. This view has already thrown up practical applications such as quantum cryptography and rudimentary quantum computers. But it has also reawakened long-deferred foundational questions in new guises. It shows us that Bohr and Einstein could already see the ramifications for the philosophy and epistemology of science. Lacking the experimental tools to make progress, they doubted that these issues could ever be much more than metaphysical. Now they can be, and it is right that scientists should have the confidence to raise them afresh. ■

Wheat lag

Growth in yields of the cereal must double if the Green Revolution is to be put back on track.

Wheat is widely considered to be the world's most important crop, and Norman Borlaug knew a thing or two about how to grow it. The US agronomist developed varieties that could better resist disease and gave higher yields. In doing so, he saved an estimated one billion people from starvation.

This week marks a century since Borlaug's birth, so what better time to consider why millions still go hungry, and to ponder how the next Green Revolution can be kick-started? At a meeting in Mexico this week, organized by the International Maize and Wheat Improvement Center

(CIMMYT) in collaboration with the Association for Agricultural Research and Experimentation of the State of Sonora, researchers will look again at the prospects for wheat. Although wheat consumption is growing, the investment needed to build on Borlaug's legacy is scarce.

Wheat provides 20% of the dietary energy for the world's population. Yet growth in yields has stagnated at around 0.9% per year over the past decade — by contrast, maize (corn) yields grow by almost double that at approximately 1.6% per year. To meet future demand for food, researchers say that wheat yields must grow by 1.7% each year. That will require investment. The total global spend on wheat breeding and research, around US$500 million per year, is currently one-quarter of that spent to improve maize.

The discrepancy arises because seed companies can make higher profits from maize than from wheat. Maize is a hybrid crop that produces seeds with poor yields, so there is little incentive for farmers to keep and replant them. To get the best results, they must buy new maize seeds each year. Seed producers get no such annual income from wheat because farmers can reap and replant seeds from several successive harvests without losing much yield.

Existing solutions to this are as much economic as agronomic. One strategy is for plant breeders to collect royalties from farmers who save seeds — as is done in countries such as Australia and the United Kingdom. Although this approach is reasonable for farmers who harvest substantial profits, it is less applicable to subsistence farmers in the developing world; for them, science might provide better solutions.

An international research effort to boost wheat yields by 50% by 2035 will be officially launched at this week's meeting. The International Wheat Yield Partnership — a consortium of research institutions including CIMMYT and the United Kingdom's Biotechnology and Biological Sciences Research Council — is a long-term project that aims to raise US$100 million in the next five years.

Researchers plan to use the cash to explore, for example, how to improve the way wheat photosynthesizes. Wheat, frankly, does it badly; it converts just 1% of incoming light to grain. Maize is three times more efficient; sugar cane eight times.

The map of the wheat genome remains a job half-finished. An estimated US$50 million is needed to map the remaining 20 of wheat's 42 chromosomes. Obtaining the complete genome should accelerate the effort to make wheat more tolerant to heat and drought through both conventional breeding and genetic-modification techniques.

Crucial work to monitor the spread of wheat pathogens must continue. Borlaug championed this research through the Durable Rust Resistance in Wheat project led by Cornell University in Ithaca, New York, and the Borlaug Global Rust Initiative, a consortium of more than 20 institutions, which has its secretariat at Cornell. Famously, he said that "rust never sleeps", and the continued spread of the devastating Ug99 stem-rust fungus in Africa and other new strains in Ethiopia (see page 404) demonstrate the truth of his words.

"The total global spend on wheat breeding and research is currently one-quarter of that spent to improve maize."

Today, yield is not the only impact by which agriculture is judged. The Intergovernmental Panel on Climate Change will next week report on farming's direct and indirect consequences for the planet. The focus on the carbon footprint of food production is necessary, but the message cannot be boiled down to the misleading idea that all agricultural techniques bar organic farming are a plague on the environment.

Borlaug believed that increasing yields through the use of new and improved varieties, along with the responsible application of fertilizer and pesticides, could benefit natural ecosystems. Less land would need to be converted into agricultural production to grow food. Recent studies suggest that he was right (J. R. Stevenson *et al. Proc. Natl Acad. Sci. USA* **110,** 8363–8368; 2013).

Misguided opposition to some aspects of big agriculture, such as chemical-fertilizer use in the developing world, could stop many nations from growing the food they need. Objections should be saved for the irresponsible overuse of fertilizer in the West and Asia, which has led to widespread water and air pollution.

One of Borlaug's guiding principles was that food is the moral right of all who are born into this world. He won an important battle against hunger, but the war continues. ■

A parlous state

The decay at ancient Pompeii is symbolic of a deeper malaise in Italy's heritage.

Chuck Palahniuk, the US author of *Fight Club*, noted that "we'll be remembered more for what we destroy than what we create". For anyone who has been following the saga of ancient Pompeii over the past few years, that observation has the bite of reality. Report after report has appeared in the press detailing the collapse of this wall, the closure of that house or, in one extreme case in 2010, the destruction of a whole building.

Pompeii, which attracts some 2.3 million visitors a year, may be adept at grabbing the headlines, but the insidious creep of heritage erosion is far from limited to the Bay of Naples — in fact, it is a problem that encompasses the archaeological wealth of the globe.

Heritage is ensnared in a particularly vicious catch-22. On the one hand, ruins are a key source of inspiration and income, as tourists flock to connect with the past. On the other, that throng so in thrall to history does untold damage merely by turning up to have a look. But even without hordes of tourists inflicting wear and tear on fragile remains, ruins face a determined and dangerous enemy in the shape of the elements.

As soon as an artefact is unearthed, decay sets in. Objects and buildings that have lain preserved for centuries beneath the earth immediately begin to suffer when exposed to the atmosphere. A few kilometres down the road from Pompeii, the site of Herculaneum, which was buried by the same volcanic eruption in AD 79, is faring better than its more famous sibling. True, the exposed site is much smaller, but with a high number of multi-storey buildings and a lot of preserved organic matter, Herculaneum presents unique challenges. The site's enhanced ability to hold decay at bay is largely down to a public–private venture launched in 2001 and the multidisciplinary approach that the ensuing project has taken.

That success is an object lesson not only for Pompeii, which hopes to set up a similar project to help preserve the ruins (see page 411), but also for the malaise that grips Italy's heritage as a whole. Hampered by debt, the country lacks the finances to reverse the decay afflicting its monuments. Instead, it must rely on external funds and philanthropists. The crumbling Colosseum in Rome, for example, is being repaired thanks to some €25 million (US$35 million) from billionaire Diego Della Valle. Yet this is still not enough. Just west of Rome, ancient Ostia is beset by weeds, and was even flooded earlier this year.

Although structural repairs and conservation at any site require money and coordinated expertise, fending off many of the agents of erosion — by controlling vegetation, for example, or employing more guards to keep tourists in check — is relatively easy to achieve. Such steps would be important gains in this war of attrition. Finding ways to maximize the revenues coming into the sites and relaxing the austerity law stopping the employment of public-works personnel would be a start. For if preventable problems are not tackled urgently, then no amount of money from external sources will be able to reverse the tide of decay. ■

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

Fundamental research is the key to eliminating TB

To speed up progress to tackle the disease, basic researchers and those on the front line must talk to each other more, says **Christian Lienhardt**.

Fewer people will contract tuberculosis (TB) this year than last. That is good news, and enough to meet the United Nations Millennium Development Goal that called "to halt and begin to reverse the incidence" of TB. But the pace of progress is too slow. Some 8.6 million people contracted the disease in 2012, and 1.3 million died, including 320,000 people with HIV. As it stands, the world will miss the international target to eliminate TB by 2050. To meet that ambitious goal, we need to modernize the way in which we tackle the disease. This means that fundamental research must play a bigger part in nurturing the development of diagnostics, medicines and vaccines.

Some clinicians in the TB field still do not recognize the power of basic research in developing tools for patients. Indeed, the slow pace of the research often frustrates health workers. In return, basic researchers may not see how their efforts can directly contribute to TB management. The divide between the two approaches sees them often considered at odds with each other.

The past ten years have seen a remarkable rise in the number of research, financing, technical and programme partners involved in building a major pipeline for TB tools. Innovative public–private partnerships and community activism have stimulated faster and more equitable access to such tools. However, without continued studies into the molecular nature of TB, no new interventions will become available to health-care professionals. At a time when health systems still miss about one-third of the new cases each year and as drug-resistant TB strains are signalling a fresh public-health crisis, basic research must be better harnessed and promoted to develop new diagnostic, therapeutic and preventative tools.

To help catalyse this change, the World Health Organization (WHO) has made research one of the three fundamental pillars of its proposed post-2015 global TB strategy. This research pillar seeks to promote the discovery, development and rapid adoption of new tools for TB control. Never before has the WHO so clearly addressed the need for fundamental research in the fight against TB, and it hopes that doing so now will encourage crucial communication across the community. Clinicians and public-health specialists need to help to identify areas in which fundamental research can provide answers, but they also need to learn the rhythms and language of basic research — the constraints, processes and timescales. In exchange, fundamental researchers must learn how to give their studies the most impact on clinical needs and disease burden.

This will require fresh thinking on how to fund and manage cross-disciplinary science. We need to encourage researchers to assemble multidisciplinary teams to work on key public-health research themes with defined outcomes. Longitudinal studies of patients and households can be used, for example, to identify biomarkers that characterize the various clinically relevant stages of TB, together with studies on how to prevent transmission and disease development. The power of this type of study has been illustrated in the Framingham Heart Study, established in 1948 to identify the factors behind cardiovascular disease.

A coordinated approach could help to prevent TB. Understanding the biology of *Mycobacterium tuberculosis* — the function of its intracellular machinery and the biochemical pathways critical for its survival and pathogenesis — is essential to combat the pathogen. More-sophisticated studies of the host immune response could help to identify biological mechanisms in people who become infected but never develop active TB, and might reveal how vaccines could mimic the immune response.

In diagnosis, simple tests that use small quantities of body fluids to detect infection would help to prevent progression to active disease. Rapid identification of drug resistance is needed to initiate appropriate, safe and effective treatment.

In treatment, basic and translational research must focus on improving options for people with drug-sensitive and drug-resistant TB alike. These efforts should go beyond the development of better drug candidates and combinations and study why TB cannot be treated with just a few weeks of chemotherapy — as most bacterial infections can be.

None of these studies, however, will be possible without the close collaboration of patients, physicians and basic, translational and clinical researchers, to jointly guide the rational approach to identification of new diagnosis, treatment and preventive options. Broad, cross-discipline research spans a range of funding agencies and cannot be supported by individual funders. To facilitate the funding of cross-cutting work, projects could be designed on the basis of the disciplines involved, and scientists could seek funds from agencies that support work in these respective disciplines, such that multiple funders could ultimately be involved in larger projects.

Crucially, the development of optimal tools to prevent, detect and treat TB depends heavily on public and private funding; the financing gap for TB research is now estimated at about US$1.4 billion per year. The WHO's objective in promoting this crucial dialogue and change is to bring together the clinical, public-health and research communities to promote and expand TB research and its funding. But the agency also hopes that fundamental researchers from many other scientific fields will consider applying their work to TB. The world needs them. ■

DEVELOPMENT OF OPTIMAL TOOLS TO PREVENT, DETECT AND TREAT TB DEPENDS HEAVILY ON PUBLIC AND PRIVATE FUNDING.

➲ **NATURE.COM**
Discuss this article online at:
go.nature.com/it65i4

Christian Lienhardt *is senior research adviser at the WHO Global TB Programme in Geneva, Switzerland.*
e-mail: lienhardtc@who.int

RESEARCH HIGHLIGHTS

Selections from the scientific literature

ATMOSPHERIC SCIENCE

Aircraft spy on tropical storms

Aircraft flying over tropical oceans could significantly improve the accuracy of storm forecasts by making high-resolution atmospheric measurements.

Jennifer Haase at the Scripps Institution of Oceanography in La Jolla, California, and her team developed an airborne Global Positioning System device, which uses radio signals to sense and precisely locate atmospheric conditions. In 2010, during research flights into the storm cell over the Caribbean Sea that developed into Hurricane Karl, the team found that their system agreed with data from spaceborne systems, and it increased the number of atmospheric profiles taken in the storm region by more than 50%. If deployed on commercial aircraft, the technique could produce huge amounts of data for use in meteorology, the authors note.
Geophys. Res. Lett. **http://doi.org/rzd (2014)**

FEV IMAGES/ALAMY

GENETICS

Loblolly takes genome size prize

A species of pine tree native to the southeastern United States has a genome with 23 billion base pairs, more than 7 times the length of the human genome.

Kristian Stevens at the University of California in Davis and his team sequenced and assembled the genome of the loblolly pine (*Pinus taeda*; pictured), mainly using the tissue from a single pine seed.

Another team, made up of many of the same researchers and led by Jill Wegrzyn at the University of Connecticut in Storrs, characterized around 50,000 of the genes and estimated that 82% of the loblolly genome is made from repetitive elements. This work, the first pine genome assembled so far, provides a foundation to study the biology of conifers, the authors say.
Genetics 196, **875–890; 891–909 (2014)**

ZOOLOGY

How squid control their shine

PALOMA T. GONZALEZ-BELLIDO

Squid control the dynamic iridescence of their skin separately from the way they control their pigmentary colours.

Paloma Gonzalez-Bellido at the University of Cambridge, UK, Trevor Wardill at the Marine Biological Laboratory in Woods Hole, Massachusetts, and their colleagues show that iridescence in Atlantic longfin squid (*Doryteuthis pealeii*) is partly controlled outside the brain through the stellate ganglion. Severing input to this ganglion on one side of the body results in the loss of iridescence on that side (**pictured**) within 10 minutes, with the structures responsible for this iridescence becoming transparent. Both iridescence and the control of pigmentary colours require input from the brain through the pallial nerve, but the iridescence signal is routed through the stellate ganglion. By contrast, the signals for skin-pigment changes are controlled separately through the fin nerve.
J. Exp. Bio. 217, **850–858 (2014)**

MATERIALS

Strong, stiff and shell-like ceramic

A ceramic material created using ice crystals as a template has demonstrated a rare combination of high strength, toughness and stiffness.

Sylvain Deville of the French National Centre for Scientific Research in Cavaillon and his colleagues made the material from alumina platelets, alumina nanoparticles and a silica–calcia glass. When they froze the material, ice crystals grew that caused the platelets to align into layers, trapping the nanoparticles between them. These layers can be pressed and sintered into pieces a few centimetres in diameter and a few millimetres thick. Under the microscope this material resembles nacre, which makes up many seashells, and is as strong and tough as many aluminium alloys used in engineering, but with higher hardness and stiffness, which it maintains up to 600 °C.
Nature Mat. **http://doi.org/r29 (2014)**

NEUROSCIENCE

Protein seed as Alzheimer's test

A test that exploits the abnormal stickiness of misfolded proteins could one day be used to diagnose Alzheimer's disease, reports a team led by Claudio Soto at the University of Texas Medical School in Houston.

The team adapted an amplification assay, originally developed for prion diseases. The new test used certain proteins as seeds that trigger clumping of malformed amyloid-β proteins, creating aggregates that are typical of the disease. When tested on patient samples from three centres, the assay distinguished the cerebrospinal fluid of patients with Alzheimer's from that of people with other neurological diseases, with 90% sensitivity and 92% specificity. The researchers are investigating why the test did not work on samples from a fourth centre.

Cell Rep. **http://doi.org/r3b (2014)**

NANOTECHNOLOGY

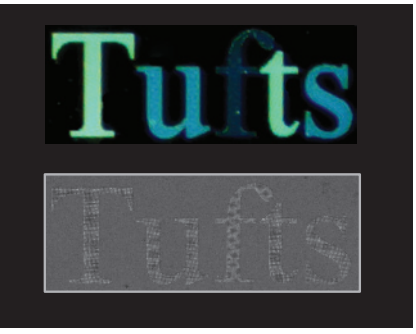
Writing with water and silk

Electron beams can create very fine nano-structured patterns on thin films of silk, with only water required to complete the etching process.

Fiorenzo Omenetto from Tufts University in Massachusetts and his team say that their silk-and-water system avoids the toxic chemicals and complex processing steps needed in other forms of lithography. The authors created photonic lattices (**pictured**), which manipulate light, with resolution as high as 30 nanometres. They also etched patterns out of silk films to which active materials, such as enzymes, proteins or quantum dots had been added.

Nature Nanotechnol. **http://doi.org/r3c (2014)**

SILKLAB/TUFTS UNIV.



ZOOLOGY

Chickens' Pacific voyaging detailed

Suggestions that ancient Polynesians had contact with South America could have been biased by contaminated samples of ancient-chicken DNA, say Alan Cooper and Jeremy Austin at the University of Adelaide, Australia, and their colleagues.

The authors sequenced genetic information from 37 ancient chicken bones and 124 modern samples, all from Polynesia and the islands of southeast Asia. They found a unique and distinctive set of DNA variations in all of the ancient and many of the modern specimens. This 'Polynesian motif' is not found in early South American chickens, suggesting that humans transported chickens from Micronesia across the Pacific Ocean, but only as far as Easter Island (Rapa Nui). Contamination of some ancient samples with modern-chicken DNA probably explains previous suggestions that the migration continued all the way to South America, the authors say.

Proc. Natl Acad. Sci. USA **http://doi.org/rx6 (2014)**

ASTROPHYSICS

Early quasars ate like the rest

The giant black hole in the most distant-known quasar, which formed just 750 million years after the Big Bang, engulfed matter at the same rate as much younger quasars.

Alberto Moretti of the Brera Astronomical Observatory in Milan, Italy, and his colleagues used the European Space Agency's XMM-Newton observatory to study ULAS J1120+0641, a galactic centre that produces huge amounts of radiation powered by a supermassive black hole. They found that its X-ray spectrum, an indication of the rate at which the black hole sucks in matter, was indistinguishable from those of quasars seen later in the life of the Universe. The authors had expected that ULAS J1120+0641 would collect matter at a much higher rate because its mass is 2 billion times that of the Sun. The slow growth rate of the galactic centre raises questions about how it could have reached its huge size so early in the Universe's life.

Astron. Astrophys. **563, A46 (2014)**

COMMUNITY CHOICE

The most viewed papers in science

CLIMATE SCIENCE

Climate change endangers culture

HIGHLY READ on www.IOP.org in March

Many of the world's most important cultural sites could find themselves below sea level owing to climate change.

Using high-resolution topography and predictions of sea-level rise as a result of global warming, Ben Marzeion at the University of Innsbruck, Austria, and Anders Levermann of Potsdam University, Germany, calculated the temperature at which each cultural site on the World Heritage List would be affected by swelling oceans. The authors found that if current global mean temperatures were sustained for two millennia, around 40 sites, 6% of the total, would be damaged if left unprotected. If global temperatures warmed by 3 kelvins the number of sites would rise to 136, or 19% of the total.

Environ. Res. Lett. **9, 034001 (2014)**

IMMUNOLOGY

HIV vaccine success secrets

The production of certain antibodies could explain the partial success of an HIV-vaccine trial and the failure of another.

In 2009, researchers reported that an experimental HIV vaccine had reduced infection risk by 31%, the only HIV vaccine ever shown to be effective. Two teams have now compared the immune responses of people in that trial, known as RV144, with those of participants in a different trial, VAX003, in which a vaccine did not prevent infection. Georgia Tomaras at Duke University in Durham, North Carolina, and her colleagues found that volunteers in the RV144 trial produced a greater response, in terms of IgG3 antibodies, which recognize a portion of HIV's outer shell, than people in the unsuccessful trial.

Another team, led by Galit Alter at Massachusetts General Hospital, Boston, also found that the differences between the two vaccines may be explained by higher levels of IgG3, and by another antibody called IgG1, in the successful trial. These antibodies might have spurred other immune cells to vanquish cells infected with HIV, both teams conclude.

Sci. Transl. Med. **6, 228ra38; 228ra39 (2014)**

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

SEVEN DAYS *The news in brief*

POLICY

Nuclear handover

Japan will relinquish more than 300 kilograms of weapons-grade plutonium and around 200 kilograms of highly enriched uranium to the United States. The handover was announced on 24 March at the Nuclear Security Summit at The Hague in the Netherlands. Japan had kept the materials for use in an experimental Fast Critical Assembly reactor in Tokaimura. But the facility is currently not operating, and Japan's stockpile has raised concerns about potential terrorist threats and nuclear-weapons ambitions.

FACILITIES

Climate-data site

The US government launched a pilot website (http://climate.data.gov) on 19 March to publish and visualize open data on climate change — a move to increase public awareness and promote community planning around the issue. Operated by NASA and the US National Oceanic and Atmospheric Administration, the website will initially focus on coastal flooding and sea-level rise, using data from various federal agencies. It aims to produce a series of web-based tools for climate planning.

Telescope trauma

The budget for NASA's Wide-Field Infrared Survey Telescope (WFIRST) could spiral out of control if the mission expands to include an inherited telescope and an instrument to hunt for extrasolar planets, concludes a report released on 18 March by the US National Research Council. WFIRST is likely to be the agency's next big astrophysics spacecraft after the James Webb Space Telescope launches in 2018. NASA should expect to spend at least US$2.1 billion on it, says the report, even if officials repurpose a 2.4-metre telescope donated by the National Reconnaissance Office, the US agency responsible for intelligence satellites.

DAVE HODSON, CIMMYT

Wheat-disease warning

Wheat farmers in East Africa and the Middle East must be vigilant for a pathogenic fungus that has wiped out more than 10,000 hectares of the crop (pictured) in southern Ethiopia since a 2013 outbreak. The warning comes in a report to the Borlaug Summit on Wheat for Food Security in Mexico on 25–28 March. The strain causing stem-rust disease in Ethiopia — sub-Saharan Africa's largest wheat producer — is not the virulent Ug99 strain that is spreading across the globe, says the report from a coalition of research organizations, including the International Maize and Wheat Improvement Center in El Batán, Mexico. The disease could spread to countries such as Kenya and Uganda, where farmers are planting vulnerable varieties of wheat.

Collection access

US government agencies must produce new rules for maintaining scientific collections of everything from Moon rocks to dinosaur fossils. On 20 March, John Holdren, director of the White House Office of Science and Technology Policy, ordered each agency with a permanent collection to produce a draft management and access policy within six months. The goal is to better preserve and increase access to artefacts including rocks, tissue specimens, seeds and other objects retained for scientific inquiry.

RESEARCH

Top-quark mass

Physicists have produced the most precise measurement yet of the mass of the top quark — the heaviest fundamental particle — by combining data from four experiments at the world's two biggest particle colliders. The new measurement, 173.34 ± 0.76 GeV c^{-2} ($310.28 \pm 1.36 \times 10^{-27}$ kilograms), has a relative uncertainty of 0.44%. This compares with uncertainties of 0.50% and 0.55%, respectively, in measurements made at the (now closed) Tevatron at Fermilab near Batavia, Illinois, and at the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. The result was posted on the arXiv preprint server on 18 March (The ATLAS, CDF, CMS and D0 Collaborations http://arxiv.org/abs/1403.4427; 2014).

Genome editing

The Berkeley and San Francisco campuses of the University of California have announced a US$12-million project focused on the CRISPR/Cas9 gene-editing technology, the powerful technique that allows targeted rewriting of genomes (see *Nature* http://doi.org/rz6; 2014). The Innovative

Genomics Initiative will receive $10 million from the Li Ka Shing Foundation in Hong Kong, with the rest made up by the two universities. Jennifer Doudna, the Berkeley-based executive director of the initiative, says that it will focus on developing the technology for human-health applications and creating a library of research resources so that others can use the technique.

EVENTS

Ebola outbreak

Dozens of people have died in an outbreak of the Ebola virus in southern Guinea. The World Health Organization in Geneva said on 24 March that 86 suspected cases resulting in 59 deaths have been reported. The situation was changing rapidly as *Nature* went to press. The numbers are likely to change as new cases are identified and suspected infections are ruled out through laboratory testing. Possible Ebola cases in neighbouring Liberia and Sierra Leone are also being investigated.

US drought

There is no end in sight for the drought ravaging the southwestern United States, the US National Oceanic and Atmospheric Administration (NOAA) said on 20 March in its Spring Outlook report.

MARCIO JOSE SANCHEZ/AP

California has experienced its warmest and third-driest winter on record, with water reserves depleted (**pictured**) and temperatures from December to February averaging 8.9 °C — almost half a degree warmer than the previous record, set in 1980–81. NOAA also said that rivers in half of the continental United States are at minor or moderate risk of flooding this spring. The southern Great Lakes regions are at greatest risk, the report says.

Polio in Iraq

Iraq last week reported its first case of polio in 14 years. A six-month-old boy, who had not been vaccinated against the disease, developed paralysis in February. The virus was also isolated from his asymptomatic three-year-old sister. The case represents the first regional spread of polio since an outbreak of the crippling disease began last October in Syria, where, as of 20 March, 37 cases have been reported by the World Health Organization.

BUSINESS

Meningitis vaccine

The United Kingdom could become the first country to routinely immunize infants with a broad-spectrum vaccine against meningitis B. The Joint Committee on Vaccination and Immunisation said on 21 March that the vaccine Bexsero — developed by Novartis in Basel, Switzerland — should be used if it can be purchased at low cost. Last year, after the vaccine was approved in Europe, the UK committee had said that it was not worth its price. Novartis plans to seek approval for the vaccine in the United States later this year, after regulators approved its emergency use in 2013 to quell a meningitis B outbreak at universities.

FUNDING

Forest funding

Countries must pay US$30 billion a year for tropical-forest conservation by 2020 to halt deforestation and promote rural development, recommends a report from the United Nations Environment Programme. The report, published on 21 March, comes four months after the United Nations climate summit in Warsaw, where nations agreed on the REDD+ scheme to make carbon payments to developing countries that reduce deforestation and forest degradation (see go.nature.com/pfswg2). Just $6.27 billion has been promised to the scheme so far.

Turing centre

In its budget report on 19 March, the UK government announced a £42-million (US$69-million) 'big data' institute, to be named after British computer scientist Alan Turing. The government will also spend £55 million on a centre to manufacture cell therapies for late-stage clinical trials, £19 million on a centre to help to commercialize the carbon-monolayer material graphene, and £106 million on PhD-training centres. The funding will be spread over five years. But researchers remain disappointed that funding for basic science is frozen at £4.6 billion per year for the next two years, the same level it has been at since 2010. See go.nature.com/fefnls for more.

COMING UP

30 MARCH
As part of its fifth assessment, the Intergovernmental Panel on Climate Change will publish a report on the vulnerability and impacts of climate change on human and natural systems.
go.nature.com/cu4bhk

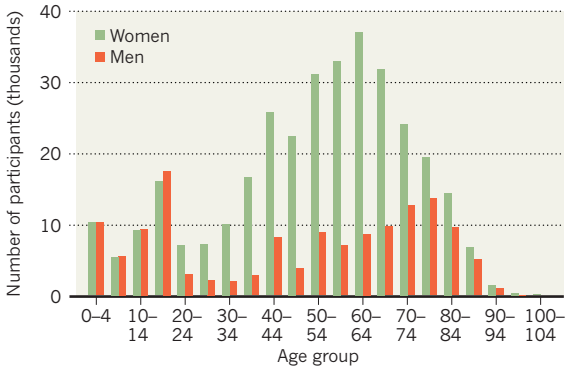
27 MARCH
Southeast Asia will be declared polio free by the World Health Organization. The last wild polio case was reported from India on 13 January 2011.
go.nature.com/5uxoar

TREND WATCH

More than 2.2 million people in the United Kingdom are voluntarily taking part in large-scale population-cohort studies, making for an "unparalleled collection" of data on health and well-being, according to a review of 34 studies by the UK Medical Research Council. The earliest cohort study mentioned in the review, which was published on 21 March, began in 1946 and is still running. However, men aged 20–40 are underrepresented in such studies (see chart).

YOUNG MEN NEEDED IN UK COHORT STUDIES

3.5% of the UK population takes part in medical cohort studies, but few young men are participating.



Data from 26 cohort studies. The Million Women Study (1.36 million women over 50) and UK Biobank (500,000 men and women aged 45–69) are not shown: they eclipse all other studies.

SOURCE: MRC

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

NEWS IN FOCUS

UNIV. MISSISSIPPI COMMUNICATIONS

Marijuana for use in US research is grown at only one federally approved facility, at the University of Mississippi in Oxford.

POLICY

Federal red tape ties up marijuana research

Despite drug's legalization in two US states, biomedical science faces continued restrictions.

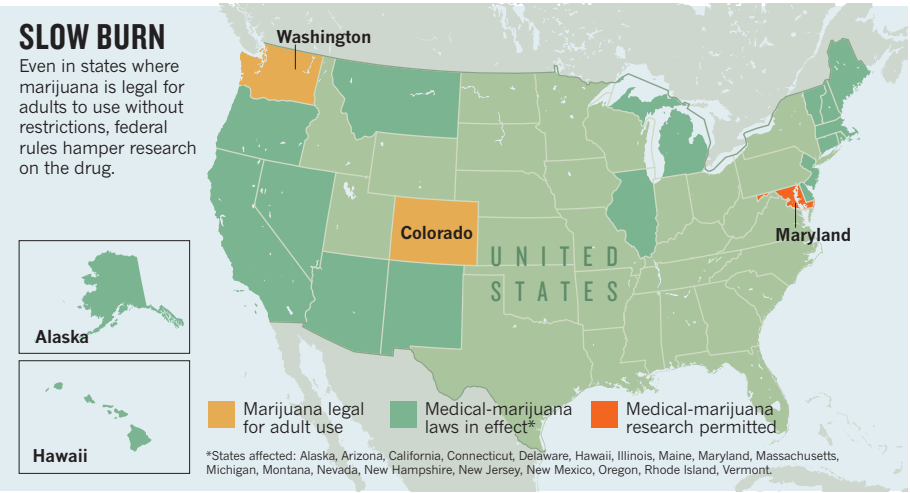
BY HELEN SHEN

Hidden in a locked room on New York University's campus in lower Manhattan, a half-tonne steel safe guards hundreds of vials of a drug extracted from marijuana. To the US government, the drug — Epidiolex — is one of the world's most dangerous substances, ripe for abuse. But to Orrin Devinsky, a neurologist at the university, it is a potential treatment for tremors in children with severe epilepsy.

Although Epidiolex, made by GW Pharmaceuticals in Salisbury, UK, is not psychoactive, it took Devinsky six frustrating months to secure local and national approval for a clinical trial that began in January. "If some teenagers broke in, they'd get nothing out of it," he says. The physical and regulatory barriers that nevertheless surround Epidiolex point to the stigma that continues to stymie research on marijuana and its derivatives, even in US states that have moved to legalize its sale and use, such as Colorado and Washington (see 'Slow burn').

Colorado's legal marijuana sales began in January and are booming. The state's legislators are weighing a proposal to provide US$7 million to study the drug's therapeutic potential. But even if the measure is enacted, Colorado researchers might not be permitted to use locally grown plants for their studies, and would need to clear several levels of government approval before running any clinical trials. The US Drug Enforcement Agency (DEA) still classifies marijuana alongside heroin and LSD as a substance with "no currently accepted medical use and a high potential for abuse".

"It's a terrible, terrible paradox," says Devinsky. "The federal government is ▶



SOURCE: MARIJUANA POLICY PROJECT

severely limiting the research that we can do."

Currently, the US National Institute on Drug Abuse (NIDA) runs the nation's only federally sanctioned marijuana farm, producing a research supply through a contract with the University of Mississippi in Oxford. To obtain the strain, researchers — even those with private or local funding — must gain approval from the Department of Health and Human Services (HHS) or the National Institutes of Health, as well as the Food and Drug Administration (FDA). The DEA and local drug-enforcement officials must also approve research facilities for secure storage and handling of marijuana. The approval process can last for years.

Suzanne Sisley, a clinical psychiatrist at the University of Arizona College of Medicine in Phoenix, knows this well. In 2011, she secured approval from the FDA to conduct a trial of marijuana to treat post-traumatic stress disorder in US military veterans. But it was not until two weeks ago — nearly three years later — that she received clearance from the HHS. Although Sisley already has plenty of volunteers for her 70-person trial, her laboratory will have to pass inspections by drug officials before it can receive limited quantities of the only federally approved marijuana strain.

That variety contains relatively high levels of tetrahydrocannabinol (THC), marijuana's primary active ingredient, and low levels of cannabidiol (CBD), another potentially therapeutic component. Such controlled cultivation helps to maintain chemical consistency in research marijuana, explains Mahmoud ElSohly, who directs the NIDA Marijuana Project at the University of Mississippi.

Still, he anticipates imminent research demand for more plant varieties, especially some containing higher concentrations of CBD. Emerging data suggest that the chemical exhibits anticonvulsant, anti-inflammatory and pain-relieving effects, while lacking — or perhaps even countering — the psychoactive effects of THC[1–3]. A few highly publicized anecdotes involving the use of marijuana products for treating severe epilepsy in children have also stoked intense public interest in high-CBD marijuana. As reports surface of families moving to Colorado to gain access to CBD for their children, Devinsky and other researchers see an urgent need to accelerate scientific investigations of the substance.

"The federal government is severely limiting the research that we can do."

ElSohly has already prepared one variety of marijuana that contains equal amounts of CBD and THC, and is hoping to grow a high-CBD, low-THC variety later this year. "With all the publicity going on right now, I'm sure somebody is going to request that," he says.

But the diversity of recreational marijuana strains flooding the Colorado marketplace — such as 'lemon skunk', 'golden goat' and 'bubba kush' — will not help to fill the research void. In a memo earlier this month, the University of Colorado reminded its faculty members to pursue studies only through federally approved channels, although it acknowledged that the NIDA supply may not be representative of many marijuana strains available to people in the state. "As a university that receives federal financial aid and federal research funding, we have to abide by federal laws," says Ken McConnellogue, a university spokesman.

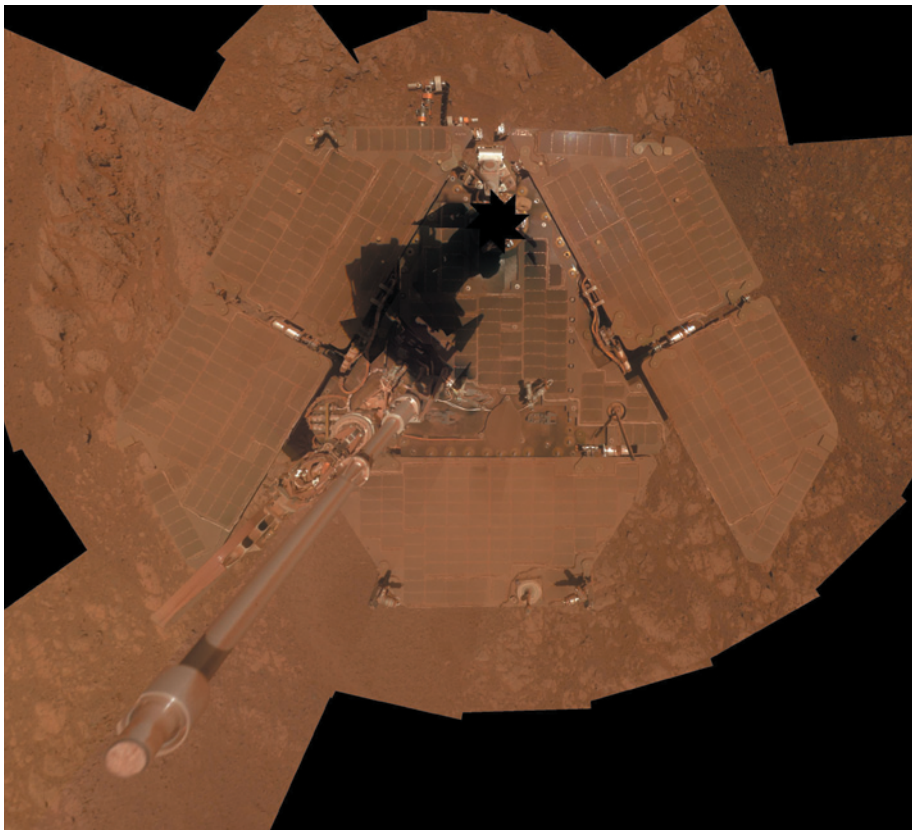
One field that does seem likely to flourish in Colorado and Washington is public-health research. Officials in both states are preparing surveys and studies to track long-term rates of marijuana use as accessibility and social acceptance of the drug increase, and as large-scale data become available for the first time. "We have more people using marijuana, and people are more willing to tell you that they're using it," says Laura Borgelt, a clinical pharmacist at the University of Colorado in Aurora, who is interested in studying the effects of maternal marijuana use during pregnancy on child development.

Other topics of interest include possible changes associated with marijuana in the use of alcohol and other drugs, trends in car accidents and risky behaviour, and school performance. Last year, NIDA awarded more than $1 million in supplemental funding to five groups to study some of these questions in the context of the natural epidemiological experiments under way in Colorado and Washington.

Few data exist to predict the effects of legalizing the drug's use and sale. Studies of the 'non-enforcement' cannabis system in the Netherlands by Robert MacCoun, a social psychologist at the University of California, Berkeley, suggest that marijuana consumption rose in the 1980s, as marijuana shops became widespread[4]. Overall, however, rates of marijuana use in the Netherlands are comparable to those in other European countries.

Cultural differences make it difficult to extrapolate those findings to the United States, warns MacCoun. His group has begun to study patterns of drug use in Washington, and he is looking forward to seeing data from Colorado researchers. "I'm watching quite eagerly to see what we can learn." ■

1. Izzo, A. A., Borrelli, F., Capasso, R., Di Marzo, V. & Mechoulam, R. *Trends Pharmacol. Sci.* **30,** 515–527 (2009).
2. Hill, A. J., Williams, C. M., Whalley, B. J. & Stephens, G. J. *Pharmacol. Ther.* **133,** 79–97 (2012).
3. Niesink, R. J. & van Laar, M. W. *Front. Psychiatry* **4,** 130 (2013).
4. MacCoun, R. J. *Addiction* **106,** 1899–1910 (2011).



NASA/JPL-CALTECH/CORNELL UNIV./ARIZONA STATE UNIV.

The dust-covered Opportunity rover aims to travel 3 more kilometres on Mars to reach clay deposits.

PLANETARY SCIENCE

NASA missions bid for extensions

Scientists must show projects' potential to secure funding.

BY ALEXANDRA WITZE

After spending 10 years and driving almost 39 kilometres on Mars, is there anything left for NASA's Opportunity rover to do?

Mission scientists think so. They say that Opportunity is on the verge of making some of its biggest discoveries yet. Just 3 kilometres south of the rover's current position, satellites have spied a lode of water-bearing clay minerals — a potentially habitable environment from a time when Mars was warmer and wetter than it is today. Matthew Golombek, Opportunity's project scientist at the Jet Propulsion Laboratory (JPL) in Pasadena, California, says that the rover is on the verge of hitting the "jackpot".

But, like six other ongoing NASA missions studying the Moon, Mars and Saturn (see 'Fighting for life'), Opportunity's money is due to run out at the end of the US fiscal year, on 30 September. Managers for each mission are trying to convince the agency to cough up continued funding, and their arguments are due on 11 April. A 'senior review' panel of external planetary scientists will rank the proposals' potential science return, and submit their suggestions to NASA headquarters for a final decision.

Similar reviews take place every couple of years, but the upcoming one is especially fraught. NASA will almost certainly continue to fund its flagship rover, Curiosity, which landed on Mars less than two years ago. That leaves the older missions to battle it out for the rest of a limited pot of money.

Jim Green, who heads NASA's US$1.3-billion planetary-science division, says that the agency is working to find money for all of its operating spacecraft, and wants to dispel the idea that missions are battling to the death. "We'll go through the senior review, order missions and fund them," he told attendees at the Lunar and Planetary Science Conference last week in The Woodlands, Texas.

But hard choices often have to be made, says Nadine Barlow, a Mars researcher at Northern Arizona University in Flagstaff who served on the 2012 senior review panel. "You always have to keep in the back of your mind that this could be the end of the mission," she says.

For this year's senior review, Green asked project scientists to take a slightly different approach from before. Each mission is to outline scenarios for three different price tags: an ordinary operating plan; a larger budget that could cover extra science targets; and a financial floor beneath which little science can realistically be done.

Above all, each mission needs to show what fresh science it has to offer. That can be challenging when a spacecraft has already been exploring for years.

One approach is to argue for studying planetary change over time. The Lunar Reconnaissance Orbiter (LRO) has been snapping high-resolution pictures of the Moon since 2009, providing a long time series for studying surface changes. At the planetary conference, Mark Robinson of Arizona State University in Tempe showed off dramatic pictures of high-speed meteorite impacts on the Moon. On 17 March 2013, for instance, automated cameras on Earth captured the brightest flash ever detected from one of these impacts. Soon after, the LRO photographed the resulting crater.

The spacecraft has found many more examples of fresh impacts. Surprisingly, the craters show debris spread over greater distances than might be expected. That observation could be important for future lunar exploration: astronauts might find themselves sprayed by rocks from far-flung impacts.

The Cassini mission to Saturn also wants to examine temporal changes — in its case, the shifting seasons. Cassini arrived in 2004, just as Saturn's northern hemisphere was beginning to emerge from winter. "We're coming onto northern summer now, and we're excited about the Sun climbing to the highest extent that we've seen," says project scientist Linda Spilker of JPL.

That new Sun angle will give Cassini fresh views of Saturn's rings and of the giant hexagonal jet stream around the planet's north pole. On the planet's largest moon, Titan, which harbours northern polar seas made of methane and other hydrocarbons, the oncoming summer may change wind patterns and ruffle the oceans' surfaces. Researchers studying the moon last week reported that they may have spotted the first waves outside Earth (see *Nature* http://doi.org/r27; 2014); these episodes could occur more frequently as summer approaches. Three close flybys of another of Saturn's moons, Enceladus, are also planned for Cassini, to sample the geyser-like plumes of water vapour that spray out of fractures ▶

FIGHTING FOR LIFE

Seven NASA planetary missions are competing against each other for extensions.

Moon

Lunar Reconnaissance Orbiter
Launched: 2009
Annual operating cost*: $8.1 million

Mars

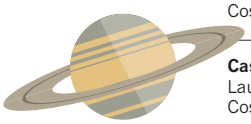
Mars Odyssey
Launched: 2001
Cost: $12.3 million

Opportunity rover
Launched: 2003
Cost: $13.2 million

Mars Express
Launched: 2003, by the European Space Agency.
Cost (to NASA): $3 million

Mars Reconnaissance Orbiter
Launched: 2005
Cost: $29.5 million

Curiosity rover
Launched: 2011
Cost: $59.4 million



Saturn

Cassini
Launched: 1997
Cost: $65.7 million

* Latest year available

▶ near its south pole.

Smaller missions are also angling for a piece of the division's pie. Those running the Mars Odyssey spacecraft, which was launched in 2001 and is the longest-lived Martian spacecraft ever, are moving the craft into a different orbit. That would allow it to look at the red planet from dusk and dawn angles that have not been tried since the 1970s. The Odyssey team is also trying to play up the fact that the orbiter is an essential communications tool: it is in a favourable orbit for relaying information from rovers on the Martian surface to Earth. The newer Mars Reconnaissance Orbiter also acts as a relay, but that mission has suffered several glitches in recent years.

"You always have to keep in mind that this could be the end of the mission."

For now, all the planetary missions under review are running as planned until the end of this fiscal year. Members of the review panel should be named shortly, and are due to meet in May and report their recommendations to Green, for his decision in June.

Stephen Mackwell, head of the Lunar and Planetary Institute in Houston, Texas, thinks that NASA officials will find a way to keep most of the missions going, especially if the project teams can show a way to take the science in a new direction. "Business as usual is not a sustainable approach," he says. ■

LAW

Software patents await legal fate

Supreme Court ruling could affect medical-diagnostics firms.

BY HEIDI LEDFORD

For the first time in more than 30 years, the US Supreme Court is tackling one of the most vexing questions in patent law: can software be patented? Its decision could affect the future of personalized medicine.

On 31 March, the court will hear arguments in *Alice Corporation* v. *CLS Bank International*, a lawsuit over four patents on software intended to reduce risk in financial transactions. The case is the latest example of Supreme Court efforts to rein in the US Patent and Trademark Office, which is notoriously permissive in granting commercial protection to inventions. Two years ago, the court struck down a class of patent often used to protect medical diagnostic tests. Last year, it ended a 30-year tradition of issuing patents on naturally occurring genes.

With the latest case, the court has the potential to restrict software patents, with repercussions not only for the technology industry, but also, yet again, for medical-diagnostics companies, says Jonathan Masur, deputy dean of the University of Chicago Law School in Illinois, who specializes in intellectual property. "It's enormously important," he says.

Software patents are controversial. Critics charge that they tend to be vague and cover obvious inventions, and that they hinder innovation. They are magnets for 'patent trolls' — businesses that collect broad patents mainly to force others to pay licensing fees. Some critics also argue that computer programs are simply the expression of mathematical formulae, and thus can be considered abstract ideas, which, along with laws of nature, are not patentable.

Yet in 1981, the Supreme Court conceded that some software could be patented, and since then the number of software patents has soared.

The *Alice* case began in 2007, when CLS Bank International, based in New York, sued Alice Corporation of Melbourne, Australia, charging that four of Alice's patents were invalid. Alice countersued CLS Bank for infringing the same patents. But CLS Bank could be in a good position, if the Supreme Court's interest in limiting patentability in recent years is any indication (see 'Valid concerns'). Observers are hoping that the court will use the case to develop concrete criteria to determine which software patents are valid. "The problem with the law to date is that it's very subjective," says Robert Sachs, a partner

VALID CONCERNS

The US Supreme Court (**pictured**) is clamping down on the kinds of inventions that can be patented.

JONATHAN ERNST/REUTERS/CORBIS

NOVEMBER 1972 ***Gottschalk* v. *Benson.*** Ruled that computer programs cannot be patented if they merely perform mathematical calculations.

MARCH 1981 ***Diamond* v. *Diehr.*** Ruled that a device controlled by a computer program was patentable — reopening the door to software patents.

APRIL 2007 ***KSR International* v. *Teleflex***. Threw out a patent on an adjustable accelerator-pedal system for being too obvious.

JUNE 2010 ***Bilski* v. *Kappos.*** Struck down a business-methods patent, saying it was based on an abstract idea.

MARCH 2012 ***Mayo Collaborative Services* v. *Prometheus Laboratories.*** Said two medical-diagnostics patents were based on laws of nature, and invalid.

JUNE 2013 ***Association for Molecular Pathology* v. *Myriad Genetics.*** Ruled that natural genes cannot be patented.

2014 ***Alice Corporation* v. *CLS Bank International.*** Arguments heard on software patents.

at the law firm Fenwick and West in San Francisco, California. "Every judge gets to decide: 'Is this the kind of thing that I think a patent claim ought to cover?'"

To bolster its arguments, CLS Bank is citing a 2012 case, *Mayo Collaborative Services* v. *Prometheus Laboratories*, that struck down two medical-diagnostics patents and sent tremors through the diagnostics industry. The patents covered the act of measuring metabolites to determine the proper dosage of certain drugs used to treat autoimmune diseases. The nine judges on the Supreme Court unanimously decided that the patents merely laid claim to a law of nature — the body's breakdown of the drugs — and were therefore invalid.

The *Mayo* ruling represented a fundamental change, says Brian Dorn, a patent attorney at the Barnes & Thornburg law firm in Minneapolis, Minnesota, who counts diagnostics companies among his clients. "It's been hard to get around," he says. "We've been getting patent rejections based on that case." Some even blame the court's decision for a drop in US venture-capital investment in medical-diagnostics companies — from US$395 million in 2011 to $278 million in 2013, according to London-based consultancy firm PricewaterhouseCoopers.

Mayo is also setting a precedent for lower courts. Last October it was used, along with the 2013 *Association for Molecular Pathology* v. *Myriad Genetics* ruling that invalidated patents on genes (see *Nature* **498,** 281–282; 2013), to throw out key patents on a prominent non-invasive prenatal test for Down's syndrome, held by the diagnostics company Sequenom of San Diego, California. The ruling, by a California district court, used a strict interpretation of *Mayo*, says Christopher Holman, a law professor at the University of Missouri–Kansas City Law School. "If that is how the courts are going to use that case, it's a very bad sign for industry," he says.

The *Alice* case could be a chance for the Supreme Court to clarify how *Mayo* should be interpreted, says Masur: rulings on abstract ideas (as is likely to arise from *Alice*) are often applied to arguments about patenting laws of nature (as in *Mayo*). If the Supreme Court invalidates many kinds of software patents, for example, diagnostics firms could be in for more trouble, he says.

The Supreme Court's tendency towards restricting patents is a concern for Gregory Graff, an agricultural economist who studies intellectual property at Colorado State University in Fort Collins. He argues that the US patent system should not be reformed through a series of radical court decisions. Instead, he would prefer to see legislation and patent-office procedures tweaked as fields of technology evolve. "You don't need the nuclear option of blowing a hole in what's deemed patentable," he says. ■

CIRO DE LUCA/REUTERS/CORBIS

Conservation work under the Great Pompeii Project began at the site in 2013.

ARCHAEOLOGY

Sister city inspires Pompeii rescue

Archaeologists hope that funding and conservation boost can halt decay of ancient city.

BY DECLAN BUTLER

Ancient Pompeii, which suffered three wall collapses this month, may look to its smaller sister city Herculaneum for tips to ensure its survival. Both cities, located on Italy's Neapolitan coast, were simultaneously destroyed and preserved by the eruption of Mount Vesuvius in AD 79, but the historic site of Herculaneum is faring much better, thanks to a sustainable conservation project.

Now an international research consortium is planning a similar pilot scheme for Pompeii. And researchers and officials are hoping that a €105-million (US$145-million) European Union project to restore Pompeii will also draw on lessons from Herculaneum.

The rescue plan cannot come soon enough. "A lot of buildings are at risk [of collapse]," one anonymous observer close to the restoration efforts told *Nature*, adding that, as well as the repair effort, "they need a system to maintain the site so that it doesn't happen again".

Excavations of Pompeii, which, with Herculaneum, is a World Heritage Site of the United Nations Educational, Scientific and Cultural Organization, began in the mid-eighteenth century. But conservation took off only in the early twentieth century, by which time much of the city had been reduced to walls. However, decay at the 66-hectare site has accelerated since the 1960s as a result of haphazard funding and the disbanding of in-house maintenance crews.

This month's collapses are the latest in a series of incidents since the 1980s. Water is the site's biggest enemy: unexcavated embankments absorb rainfall and exert pressure on adjoining structures, and poor drainage allows surface water to accumulate. A lack of regular repair work means that rain also seeps into the walls.

Much of Pompeii's masonry consists of rubble stonework, which is highly vulnerable. Typically, this consists of two wall coverings assembled with roughly placed stones of volcanic tuff (*opus incertum*), together with a core of smaller stones set in a lime-mortar grout (*opus caementicium*).

Such constructions show excellent longevity if kept dry. But water infiltration has led to repeated wetting and drying, causing stones to chip and flake, and the mortar to decay and lose cohesion. In addition, many twentieth-century conservation interventions at Pompeii have since proved inappropriate and damaging.

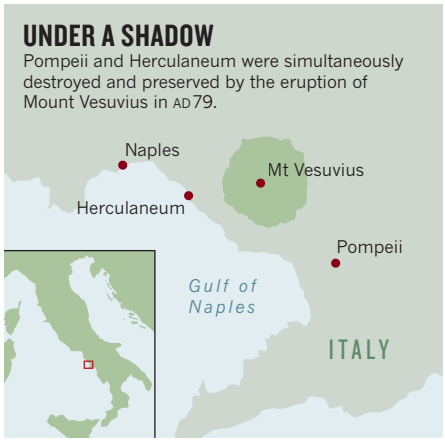
The authorities are responding, however. ▶

In February 2013, the European Union and the Italian government launched an emergency €105-million project to reverse the decades of neglect. Dubbed the Great Pompeii Project, it focuses on draining and securing embankments, restoring masonry and decorated surfaces, and protecting buildings from weathering. But the project, the budget for which must be spent by the end of 2015, has been beleaguered by concerns that delays and bureaucracy may affect the quality of the work. In response, the European Commission said on 6 March that it will carry out an urgent public review of progress.

Johannes Hahn, the European commissioner for regional policy, says that the review's aim is clear. "We have professionals working on the site. The most helpful thing we can do for them is to remove the bureaucratic barriers and let them get down to work."

Massimo Osanna, appointed this month as head of the Special Superintendency for the Archaeological Heritage of Naples and Pompeii, which is in charge of the site's conservation, says that although a longer time frame would be preferable, criticism of the pace of progress is unfair. High-quality archaeological conservation of an open-air site as vast as Pompeii inevitably has long lead times, he argues, and plans for using the money are already well under way.

Having a proper plan for the sustainable management of Pompeii is also a major issue. The outsourcing of routine maintenance to contractors since the 1980s has had severe effects, with repeated delays to urgent repairs, says Andrew Wallace-Hadrill, a social and cultural historian at the University of Cambridge, UK.



Wallace-Hadrill heads what many see as a model for conserving Pompeii: the Herculaneum Conservation Project (HCP). This was launched in 2001, when Herculaneum was in a dire state, as a public–private venture between the philanthropic Packard Humanities Institute of Los Altos, California, the British School at Rome and the superintendency. In just over a decade, and having spent €20 million in total, it has helped to reverse the decay and put the town on a sustainable conservation footing (see 'Under a shadow').

"I'm impressed; Herculaneum was a very sad, dirty and abandoned place; now it's fantastic," says Albrecht Matthaei, an archaeologist at the Roma Tre University. Although under the jurisdiction of the superintendency, the HCP's level of autonomy frees it from state bureaucracy, notes Wallace-Hadrill.

The HCP also brings together archaeologists, engineers and other scientists with conservators. This, together with an emphasis on routine maintenance, has been key, says Wallace-Hadrill. And the project differs from the typical conservation practice of carrying out complete restorations on a house-by-house basis by focusing on urgent site-wide restorations and repairs.

Now, a proposed project, the Pompeii Sustainable Preservation Project (PSPP), is taking inspiration from the HCP. An international consortium of research institutions, including the Fraunhofer research organization and the Technical University of Munich in Germany, hopes to raise US$10 million in philanthropic funding to launch a ten-year pilot project later this year that will study long-term, sustainable conservation and restoration approaches at Pompeii. "HCP is best practice," says Matthaei, a coordinator of the PSPP. The new project seems laudable, but until work begins it is difficult to assess, adds Wallace-Hadrill.

Osanna says that Pompeii's greater size makes comparisons with the HCP difficult. But he is confident that the Great Pompeii Project can fix the city. "People need to give the project time to get up to speed, after which any controversy will hopefully be put to rest," he says. ■

POLITICS

Chile puts plan for science ministry on hold

Government rethinks former president's bid to create first unified science department.

BY MICHELE CATANZARO

The new government of Chile has shelved plans to create the country's first science ministry, disappointing thousands of researchers who had campaigned for it.

Former president Sebastián Piñera submitted plans for the ministry to the Senate on 7 March — days before leaving office — in response to long-standing demands by several Chilean scientific organizations. The plans were part of a last-minute legislative package that included three other, unrelated laws.

But now, politicians supporting new president Michelle Bachelet, who was sworn in on 11 March, have derailed the proposal. Senator Jaime Quintana, spokesman for the centre-left New Majority coalition that backs Bachelet, said that the draft laws "are not legitimate" because to enact them would require money that the current government intends to allocate to its own priorities.

"It is a regrettable situation: it was a hard fight until we got the draft law, and now we feel a big discomfort," says Eduardo Bravo, director of the National Association of Postgraduate Researchers in Santiago, which was among the organizations that had called for the science ministry.

Chile is a small but highly productive player in Latin American science. In 2010, it spent US$1.15 billion on research and development. Although that is equivalent to just 0.44% of its gross domestic product, or 2.8% of Latin America's total investment in science, the country ranks behind only Brazil, Mexico and Argentina in terms of the number of articles published.

However, science governance in Chile is currently spread across the ministries of education, economy and agriculture. "This has resulted in multipolar governance that has not been capable of establishing a clear agenda of strategies and priorities, or strong networks of scientists," says Jesús Sebastián Audina, a researcher in science policy at the Spanish National Research Council in Madrid.

Jorge Babul, president of the Chilean Council of Scientific Societies in Santiago, adds:

IVAN ALVARADO/REUTERS/CORBIS

Current Chilean President Michelle Bachelet has her own plans for science governance.

"Putting science in a ministry not dedicated to it prevents the minister from spending enough time on research policy." He thinks that a single unified ministry would ensure "a big leap" in the ability to transfer knowledge within Chile. Bravo says that a dedicated institution would "have more financial independence".

In April 2013, after years of campaigning, a group of organizations — including several scientific academies and societies and Santiago-based pro-science lobby group the More Science Foundation — presented Piñera with an open letter asking that science policy be unified in a single ministry. The petition had 5,000 signatories.

Piñera responded by creating a working group of academics and politicians, led by engineer and entrepreneur Bruno Philippi, to evaluate Chilean scientific governance and propose changes. In a report released last May, the group urged the government to create the ministry.

Piñera's 7 March proposal to the Senate was strongly based on that report. According to his draft submission, the ministry would have had two sub-secretariats: one for higher education and another for science, technology and innovation. Furthermore, Chile's main science-funding body, the National Commission for Scientific and Technological Research (CONICYT), would regain a role in advising the president on science policy that was taken from it by the authoritarian regime of General Augusto Pinochet in 1973. CONICYT's resumption of this role would ensure that the president is advised by a dedicated science-governance body.

Most of Piñera's proposal is widely supported by the Chilean scientific community, says Sergio Hojman, a physicist at the University of Chile in Santiago and a member of the working group headed by Philippi. But some scientists are sceptical about the feasibility and benefits of the plan. "Building a ministry that works may take decades in Chile," says Mario Durán Toro, an engineer at the Pontifical Catholic University of Chile in Santiago. "It would be enough to re-establish the role that CONICYT had before 1974: that would be much more effective and quick," he says.

As well as potentially taking funds away from the government's own priorities, the proposed ministry conflicts with the coalition's plans: Bachelet's electoral manifesto put science in a sub-secretariat of the economy ministry, not a ministry of its own. Still, Hojman thinks that Bachelet is not necessarily opposed to the science ministry in principle, given that several members of the Philippi commission are supporters of government coalition parties.

"Building a ministry that works may take decades in Chile."

The poor timing of the proposal may also have affected its chances. "Presenting it at the last minute is something that was going to be seen with a bad eye by the new government," says Carolina Muñoz, a chemist and executive director of the More Science Foundation. "But withdrawing it would show a very serious lack of respect towards Chilean scientists."

Finance minister Alberto Arenas is expected to advise on the financial viability of the science ministry in a report on the state of the country's finances on 7 April. ■

CORRECTION

The News story 'Global seismic network takes to the seas' (*Nature* **507,** 151; 2014) wrongly located USGS seismologist Cecily Wolfe at the University of Hawaii in Honolulu instead of at Reston, Virginia.

The Neanderthal in the family

By **Ewen Callaway**

Thirty years after the study of ancient DNA began, it promises to upend our view of the past.

PHOTO ADAPTED FROM: TETRA IMAGES/ALAMY

Before ancient DNA exposed the sexual proclivities of Neanderthals or the ancestry of the first Americans, there was the quagga.

An equine oddity with the head of a zebra and the rump of a donkey, the last quagga (*Equus quagga quagga*) died in 1883. A century later, researchers published[1] around 200 nucleotides sequenced from a 140-year-old piece of quagga muscle. Those scraps of DNA — the first genetic secrets pulled from a long-dead organism — revealed that the quagga was distinct from the mountain zebra (*Equus zebra*).

More significantly, the research showed that from then on, examining fossils would no longer be the only way to probe extinct life. "If the long-term survival of DNA proves to be a general phenomenon," geneticists Russell Higuchi and Allan Wilson of the University of California, Berkeley, and their colleagues noted in their quagga paper[1], "several fields including palaeontology, evolutionary biology, archaeology and forensic science may benefit."

At first, progress was fitful. Concerns over the authenticity of ancient-DNA research fuelled schisms in the field and deep scepticism outside it. But this has faded, thanks to laboratory rigour that borders on paranoia and sequencing techniques that help researchers to identify and exclude contaminating modern DNA.

These advances have fostered an ancient-genomics boom. In the past year, researchers have unveiled the two oldest genomes on record: those of a horse that had been buried in Canadian permafrost for around 700,000 years[2], and of a roughly 400,000-year-old human relative from a Spanish cavern[3]. A Neanderthal sequence every bit as complete and accurate as a contemporary human genome has been released[4], as has the genome of a Siberian child connecting Native Americans to Europeans[5].

Enabling this rush are technological improvements in isolating, sequencing and interpreting the time-ravaged DNA strands in ancient remains such as bones, teeth and hair. Pioneers are obtaining DNA from ever older and more degraded remains, and gleaning insight about long-dead humans and other creatures. And now ancient DNA is set to move from the clean-rooms of specialists to the labs of archaeologists, population geneticists and others. Thirty years after the quagga led the way, *Nature* looks to the field's future.

A MILLION-YEAR-OLD GENOME

Ludovic Orlando, an evolutionary biologist at the University of Copenhagen, had low expectations when he started sequencing DNA from a 560,000-to-780,000-year-old horse leg bone. His colleague, Eske Willerslev, had discovered the bone buried in the permafrost of the Canadian Yukon in 2003. Then he had chucked it into a freezer, waiting for technological improvements that would allow the bone's degraded

DNA to be read. (Freezers in ancient-DNA labs brim with such 'wait and see' samples.)

On a Sunday evening in 2010, Willerslev called Orlando to say that the time had come. Orlando was unconvinced: "I started the project with the firm intention of proving that it was not possible," he says.

Sequencing ancient DNA is a battle against time. After an organism dies, the long strands of its DNA fissure into ever shorter pieces, helped along by DNA-munching enzymes. Low temperatures slow this process, but eventually the strands become so short that they contain little information.

To read the horse's genome, Orlando needed to shepherd useful DNA fragments through the harsh enzymatic treatments used to extract them and ready them for sequencing. Orlando and his team found that the preparation lost vast quantities of fragments. But with a few tweaks to the experimental protocol, such as reducing the extraction temperature, the researchers captured ten times more scraps of DNA than before — and produced a draft of the oldest genome on record[2].

Using a similar approach, Svante Pääbo, a geneticist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and his team turned their attention to 400,000-year-old remains from the Sima de los Huesos cavern in northern Spain, which may have been a burial pit for recent relatives of modern humans called hominins (see 'Hidden heritage'). In the pit, the bones remained at stable, low temperatures, slowing the breakdown of DNA. "If you could have told the hominins where to leave their bones, you may have chosen that site," says Matthias Meyer, a molecular biologist at Pääbo's institute who is leading the efforts.

Last December, the team reported[3] roughly 16,300 letters of a Sima de los Huesos individual's mitochondrial genome — the DNA from power-generating structures in its cells. The sequence revealed an unexpected relationship between the Sima de los Huesos remains and the Denisovans, an archaic group of humans that Pääbo's team had discovered in Russia's Altai Mountains thousands of kilometres away. Meyer and his colleagues hope to improve their methods enough to obtain some or all of the Sima de los Huesos individual's nuclear genome, the DNA from the nuclei of its cells. "It must be possible," says Meyer. "I won't rest until this has been done."

It is now a matter of when, not if, someone will produce a genome from an Arctic animal buried in permafrost for longer than 1 million years, says Meyer. But he and Pääbo want to push the limits of ancient DNA in hominin specimens from warmer locales, such as fossils of *Homo erectus*, the common ancestor of humans and Neanderthals, found in Asia. And Orlando says that researchers may have luck using new extraction techniques on previously vexing remains such as Egyptian mummies or *Homo floresiensis*, a small hominin at least 18,000 years old that was found in a cave on the Indonesian island of Flores. "It opens a great number of places where there are lots of important stories going on, such as the Middle East or the tropics," he says.

GHOSTS IN THE CODE

A few years ago, David Reich discovered a ghost. Reich, a population geneticist at Harvard Medical School in Boston, Massachusetts, and his team were reconstructing the history of Europe using genomes from modern people, when they found a connection between northern Europeans and Native Americans. They proposed that a now-extinct population in northern Eurasia had interbred with both the ancestors of Europeans and a Siberian group that later migrated to the Americas[6]. Reich calls such groups ghost populations, because they are identified by the echoes that they leave in genomes — not by bones or ancient DNA.

Ghost populations are the product of statistical models, and as such should be handled with care when genetic data from fossils are lacking, says Carlos Bustamante, a population geneticist at Stanford University in California. "When are we reifying something that's a statistical artefact, versus when are we understanding something that's a true biological event?"

Sometimes these statistical spectres get a body. Last year, Willerslev's team reported[5] the genome from 24,000-year-old remains dubbed the Mal'ta boy. The results showed that the boy, who had been found in central Siberia, came from a population related to both modern Native Americans and modern Europeans, matching Reich's prediction (see *Nature* http://doi.org/r2b; 2013). "It's a spectacular find," he says.

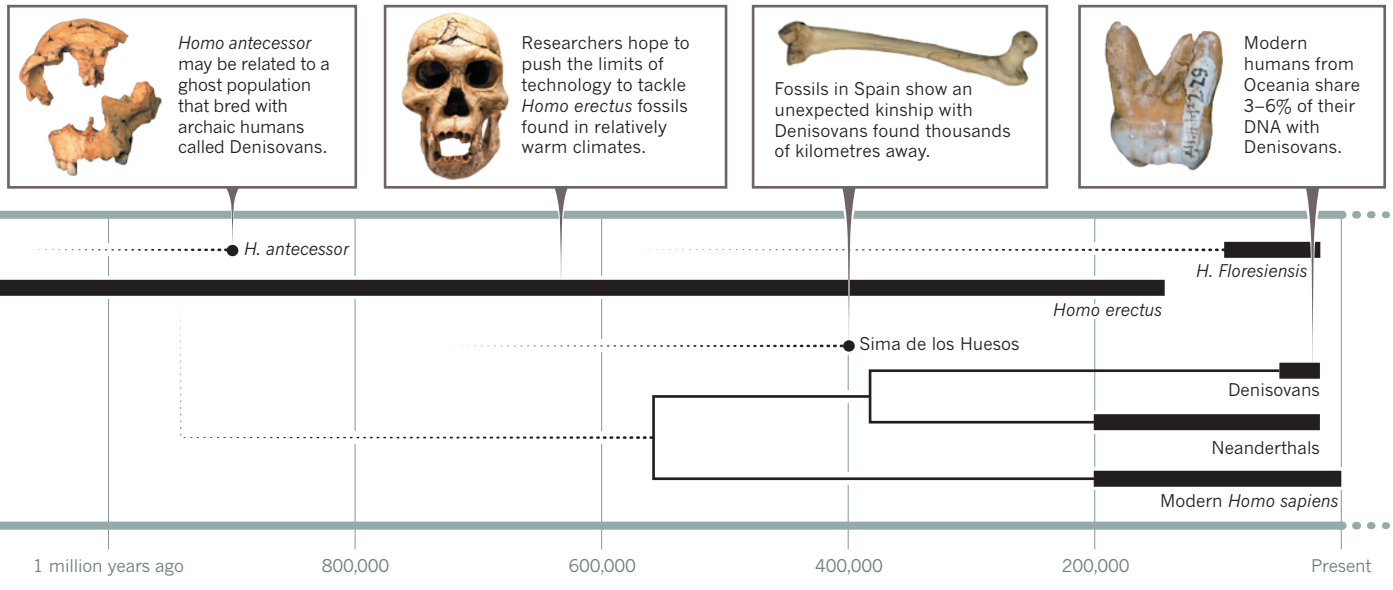
Ghost populations also lurk in ancient DNA. While analysing high-quality genomes of a Neanderthal and a Denisovan, a team led by Reich and Montgomery Slatkin at the University of California, Berkeley, noticed a peculiar pattern: present-day sub-Saharan Africans are more closely related to Neanderthals than they are to Denisovans[4]. But evidence from other ancient genomes suggested that the two archaic groups were equally related to present-day Africans. After weighing the possibilities, the scientists realized that they might have uncovered another ghost population.

The puzzle could be solved, they theorized, if Denisovans had interbred with a species that

PHOTOS (L–R): JAVIER TRUEBA/MSF/SPL; MARKUS SCHIEDER/ALAMY; REF. 3; D. REICH *ET AL. NATURE* **468**, 1053–1060 (2010)

HIDDEN HERITAGE

The study of ancient DNA is revealing connections between archaic humans — and the traces they left behind in modern genomes.



had left Africa perhaps more than 1 million years ago and branched off from the common ancestor of humans, Neanderthals and Denisovans. Subsequent Denisovans would have inherited DNA sequences that present-day Africans lack, explaining why Neanderthals seem to be closer kin to Africans.

Reich's team is analysing genetic signatures in humans with Denisovan DNA to establish when the Denisovans mated with this mystery population — information that could narrow the range of fossils to which it might belong. Genomes studied by Pääbo's lab, principally the Sima de los Huesos remains, may also reveal clues.

Reich is not the only one conjuring ghosts. Chris Stringer, a palaeoanthropologist at the Natural History Museum in London, has proposed that the 900,000-year-old hominin *Homo antecessor*, known from fossils found near Sima de los Huesos, could be part of the ghost population. If it had interbred with an ancestor of the Denisovans and the Sima de los Huesos hominins, it could explain the relationship between the two groups of remains. Testing that hypothesis would require the elusive Sima de los Huesos nuclear DNA. But Reich is optimistic that Pääbo and his team will pull it off. "They've done miracles before in that lab and they may succeed again."

THE NEANDERTHAL WITHIN

"We don't need bones necessarily to find ancient DNA," says Josh Akey, a population geneticist at the University of Washington in Seattle. "We can find the remnants of ancient DNA floating around in contemporary populations."

If early human populations bred with Neanderthals and Denisovans, their descendants should carry short segments of archaic-human DNA. Researchers such as Akey are beginning to catalogue these segments to learn about the biology of archaic humans. Unlike the hunt for ghost populations, which relies on statistical population models, this approach allows researchers to identify specific regions of the genome acquired by interbreeding.

In January, independent teams led by Akey[7] and Reich[8] pieced together a substantial portion — about 20% and 40% respectively — of the Neanderthal genome from bits lurking in the genomes of hundreds of living humans. Their research indicated that some Europeans and Asians had gained genes involved in skin and hair from Neanderthals, possibly helping their ancestors to adapt to cold climates by providing thicker skin, more hair and fewer pores (see *Nature* http://doi.org/rz9; 2014). But giant swathes of the modern genomes were devoid of Neanderthal ancestry, hinting that many Neanderthal genes might have been harmful in modern humans. Akey's team identified[7] one such region around the gene *FOXP2*, which is involved in speech and language. "It's extremely compelling evidence that there were fitness costs to interbreeding," he says.

These discoveries are only the beginning. The Akey and Reich teams found that the genomes of east Asians possess, on average, slightly more Neanderthal DNA than do people of European ancestry. Akey sees this as possible evidence that Neanderthals interbred with ancient humans on at least two separate occasions: once with the ancestors of all Eurasians, and later with a population ancestral only to east Asians. And Akey believes that humans are likely to bear genetic scraps from other extinct species, including some that interbred with the ancestors of humans in sub-Saharan Africa.

> ***With tools that make sequencing ancient DNA cheaper and easier, the field is becoming more egalitarian.***

ANCIENT DNA FOR THE MASSES

For much of the past 30 years, the sensitivity of the polymerase chain reaction (PCR), the method used to amplify ancient DNA, made it prone to contamination. The field's leaders often greeted the work of outsiders with suspicion, earning some of them the title 'the PCR police'. And in recent years, palaeogenomics has been the domain of specialist labs such as Pääbo's, with the expertise and money to obtain and screen hundreds of fossils to find the few that yield enough DNA to sequence an entire genome.

That is set to change. New procedures mean that researchers can now reliably obtain DNA from all but the most degraded samples, and then sequence only the portions of a genome that they are interested in. "I'm still surprised that there are so few labs in the world that do this," says Johannes Krause, a palaeogeneticist at the University of Tübingen, Germany, who led much of the Denisovan work while in Pääbo's lab. "It's not rocket science."

Gradually, new researchers are entering the field. "If I can break in, then anyone can," jokes Bustamante. His research originally focused on ancestry in current human populations. Then, a few years ago, he got a phone call about a mummy.

An international team had sequenced the genome of Ötzi, a 5,300-year-old frozen corpse found in the Tyrolean Alps of Italy in 1991. The researchers wondered if Bustamante could help them to make sense of the ice-man's ancestry. Together, they showed that Ötzi was more closely related to humans who now live in Sardinia and Corsica than those in central Europe, evidence that the population of Europe when he was alive looked very different to how it does today[9].

Bustamante has since plunged into the world of ancient DNA. His team is sequencing samples that chart the arrival of farming in Bulgaria, the transatlantic slave trade in the Americas and dog domestication. The group is developing tools to make sequencing ancient DNA cheaper and easier. "We want to democratize the field," says Bustamante.

Reich, too, sees ancient DNA becoming more egalitarian. His lab's growing interest in areas of human history such as the advent of agriculture or the history of the Indian subcontinent has led it to analyse — often in bulk — remains less rarefied than the scarce Neanderthal samples that first lured him to the field.

Last year, Reich was part of a team that reported[10] an analysis of mitochondrial DNA from 364 European samples between 5,500 and 1,550 years old, to identify major population shifts in Neolithic Europe. Ancient genomics is also set to solve long-standing questions about when and where humans domesticated animals such as dogs, cattle and chickens. A 2013 study[11] of 18 mitochondrial genomes from ancient dogs and wolves, for instance, suggested that European hunter-gatherers domesticated wolves from a population that is now extinct.

Researchers are also returning to the questions that launched the field 30 years ago. Around the time that Orlando's team began sequencing the 700,000-year-old horse, it also turned its attention to a much younger sample — from the quagga.

The effort to sequence the full quagga genome is part of large project to understand the evolutionary relationship between living and extinct horses, zebras and donkeys, and to identify the genetic basis for certain traits. "I was thinking it would be cool to do the oldest, but also the first — where ancient DNA started," says Orlando. "It shows the progress the field has made." ■

Ewen Callaway *writes for* Nature *in London.*

1. Higuchi, R. Bowman, B., Freiberger, M., Ryder, O. A. & Wilson, A. C. *Nature* **312,** 282–284 (1984).
2. Orlando, L. *et al. Nature* **499,** 74–78 (2013).
3. Meyer, M. *et al. Nature* **505,** 403–406 (2014).
4. Prüfer, K. *et al. Nature* **505,** 43–49 (2014).
5. Raghavan, M. *et al. Nature* **505,** 87–91(2014).
6. Patterson, N. *et al. Genetics* **192,** 1065–1093 (2012).
7. Vernot, B. & Akey, J. M. *Science* **343,** 1017–1021 (2014).
8. Sankararaman, S. *et al. Nature* **507,** 354–357 (2014).
9. Keller, A. *et al. Nature Commun.* **3,** 698 (2012).
10. Brandt, G. *et al. Science* **342,** 257–261 (2013).
11. Thalmann, O. *et al. Science* **342,** 871–874 (2013).

A tolerant approach

Despite a long record of failure, a few immunologists continue to pursue precisely targeted therapies for autoimmune diseases.

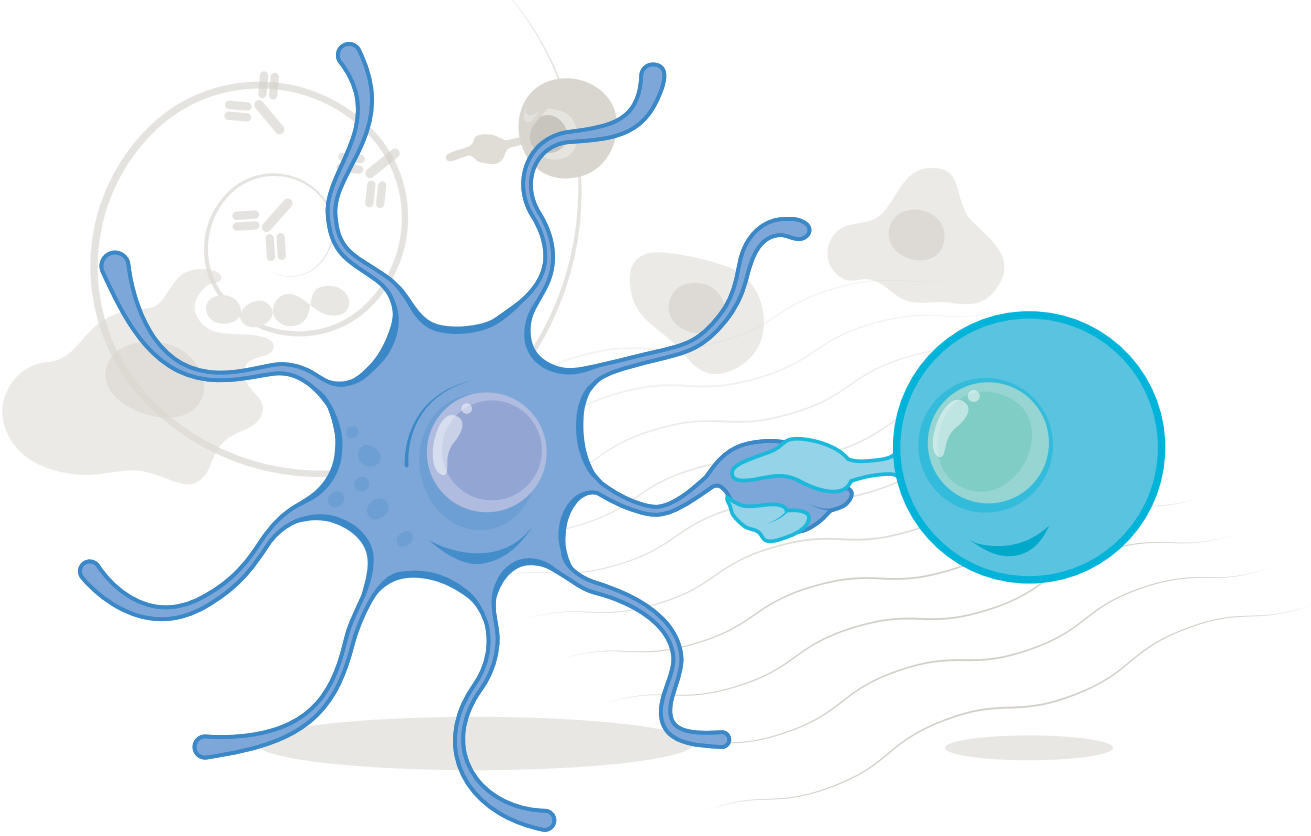


ILLUSTRATION BY NIK SPENCER/*NATURE*

BY KEN GARBER

Ever since Ed Wiley learned that he had type 1 diabetes in 1997, he has fretted over his meals, blood glucose levels and the daily programming of his insulin pump. Wiley, a statistician who lives outside Boulder, Colorado, and works on big data analytics, has learned to live in a state of hypervigilance. Finding the right dose of insulin turned out to be more art than science and, like many with the disease, his control began slipping away with time. By 2008, he says, "my insulin doses just basically didn't work any more". Unable to reliably anticipate what he needed, Wiley was having severe hypoglycaemic episodes and was at risk of diabetic seizures and long-term disability.

On his endocrinologist's advice, he enrolled in a clinical trial of a novel drug called BHT-3021. Although technically a vaccine, BHT-3021 is not designed to stimulate an immune response, but rather to shut it down, stopping the body's errant attack against cells in the pancreas that produce insulin. The goal is to achieve immune tolerance.

Drugs that broadly suppress immunity are the standard treatment for autoimmune disorders such as multiple sclerosis (MS), rheumatoid arthritis and lupus. But these drugs can lead to life-threatening

infections, and do not address the cause of the disease. Tolerance therapies are different. They aim to target only the immune cells that react to a specific antigen, a substance — in Wiley's case, the insulin precursor proinsulin — that might trigger a response. "Why shut down a major arm of the immune system, if we're just trying to restore tolerance to one antigen?" asks immunologist Larry Steinman of Stanford University in California, who developed BHT-3021.

This strategy, known as antigen-specific tolerance, is simple in concept. But, so far, dozens of clinical trials have failed to achieve a categorical success. And there is a fine line between calming the immune system and stimulating it, so these efforts risk making a disease worse — as happened in an MS trial some 15 years ago.

BHT-3021 is one of a new wave of treatments conceived by five veterans of the field that promises to do better. Early-stage trials show encouraging results in people with MS and type 1 diabetes. "A number of these approaches really are going to work," predicts David Wraith, an immunologist at the University of Bristol, UK, and one of the few persisting in pursuing the work. "The science has caught up."

AUTOIMMUNE OVERRIDE

The approaches are varied, but they all rely on the body's natural ability to distinguish its own substances from those of foreign intruders. When bacteria or viruses invade, some are swallowed by specialized antigen-presenting cells, or APCs. These chop up the bacterial or viral antigens and present them to T cells, white blood cells that orchestrate the immune response. The T cells then proliferate and launch a coordinated attack.

APCs also ensure that normal daily maintenance does not turn deadly. As the body's own cells continuously die and are replenished, APCs mop up the debris and present those self-antigens to T cells along with an array of proteins that signal that these cellular remnants pose no danger. In autoimmunity, for unknown reasons, this protective mechanism goes awry. The new therapies are designed to override this dysfunction by deliberately sending the relevant antigen to tissues where the body is likely to see it as a non-threatening part of itself (see 'Teaching tolerance').

Most of the therapies developed so far target MS, which occurs when the immune system attacks the myelin sheath that protects neurons in the brain and spinal cord. Immunologist Stephen Miller of Northwestern University in Chicago, Illinois, designed a therapy[1] that he and neurologist Roland Martin, now at University Hospital Zurich in Switzerland, began testing in patients in 2009. During the treatment, the patients' white blood cells are extracted, chemically linked to seven myelin antigens, then reinfused. The cells make their way to the spleen, where they die and release the antigen, which is picked up by APCs.

Wraith's drug[2], ATX-MS-1467, uses four peptides, or pieces, of a myelin protein commonly attacked in MS. These injected antigens are taken up by immature APCs, which are incapable of stimulating T cells and instead inactivate them or convert them to a T-cell type that maintains tolerance. Krzysztof Selmaj's group at the Medical University of Lodz in Poland designed a similar therapy for the disease using three myelin peptides delivered through a patch that users wear on their skin[3].

Instead of protein fragments, Steinman's diabetes treatment consists of circular pieces of DNA carrying the proinsulin gene, injected into muscle[4]. The proinsulin protein is manufactured in the muscle cells, and is then secreted, taken up by APCs and presented to T cells. This produces "a signal that doesn't invoke danger but invokes tolerance", says Steinman. So far, this treatment and the other new therapies have been tested in fewer than 150 patients, but industry onlookers say they show promise.

DANGEROUS GROUND

The treatments not only have to overcome strong autoimmunity, they must also avoid making it worse. "We need to be extremely cautious," says Gerald Nepom, an immunologist at the Benaroya Research Institute in Seattle, Washington. Any novel manipulation of the immune system involves some risk. In 2006, an antibody-based treatment developed by the German drug company TeGenero was given to six healthy volunteers in a UK trial. Designed to quell the autoimmune response by a mechanism different from antigen-specific tolerance, it instead caused a massive immune response and multiple-organ failure[5]. The participants survived, and investigators have since implemented safer dosing protocols, but researchers know that an immune response can quickly go wrong. "People ask me what keeps me up at night," says Steinman. "Until we have quite a few patients under our belts for some time, it's the worry that we're going to make things worse."

> *"We have to take small steps and see what happens."*

Antigens can easily trigger immunity instead of tolerance, because there is a delicate balance between the two fates. Drug dose, delivery route, tissue destination and unpredictable changes in T-cell identity all matter. In a trial for an MS treatment that began in 1998 (ref. 6), doctors gave patients a single modified myelin peptide, but the trial was halted after three of the first eight recipients suffered worsening symptoms; one was left unable to walk. All three recovered with immunosuppressive treatment, but tests clearly implicated the experimental therapy, and the researchers ultimately worked out that they had extrapolated too high a dose from earlier *in vitro* studies. "We were both frustrated and also shocked," says Martin, a lead investigator on the trial.

Most trials so far have simply failed to work. In 2009, a trial of a myelin peptide antigen involving 612 people with MS showed no benefit over a placebo[7]. One likely reason is that the immune response in most autoimmune diseases can shift from one antigen to another as tissue damage progresses. Miller documented this phenomenon, known as epitope spreading, in animal models 20 years ago[8]. Martin points out that the failed 2009 trial used a single antigen, and adds that the patients involved were at a very advanced stage of the disease, when the immune system is no longer the main cause of neuronal damage.

Wraith says that the field has learned from its mistakes. The new therapies all incorporate multiple antigens to anticipate epitope spreading, for example. And researchers are paying close attention to other relevant factors, such as how the drugs are administered. In the past, says Christophe Benoist, an immunologist at Harvard Medical School in Boston, Massachusetts, it was often more of a gamble. The approach, he says, was, "Let's just put the antigen in and hope something good happens".

Richard Ransohoff, an MS researcher at the Cleveland Clinic in Ohio, has faith in the new therapies, which build on recent advances in understanding antigens and T cells. "These are all very experienced people who understand the complexity of what they're trying to do," he says. That is not to say that the mechanisms of tolerance have been completely worked out. "We're working furiously," says Steinman, but he admits that they are operating in "a vast ocean of ignorance in between icebergs of knowledge". Waiting for that perfect understanding, however, seems foolish. "We have to take small steps and see what happens," he says.

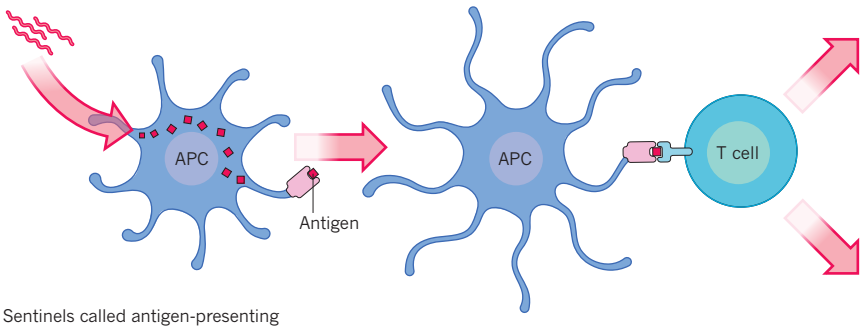
Results from the latest wave of trials are reassuring. Martin's clinical team gave nine patients a single injection of the manipulated immune cells, in escalating doses. The treatment seems to be safe, and the four patients receiving the highest doses showed a reduction in the number of T cells targeting self-antigens[9]. "That was a very positive proof-of-concept study," says Nepom, who was not involved in any of the trials.

The most promising trial, he adds, was a 30-patient test of Selmaj's therapy. Compared with a placebo, the treatment achieved a statistically significant drop in MS disease activity, as measured by magnetic resonance imaging of the brain. Patients assigned to the therapy also had far fewer relapses[10].

In Wraith's ATX-MS-1467 trial, sponsored by Apitope in Diepenbeek, Belgium, 43 patients received the treatment in a series of five escalating

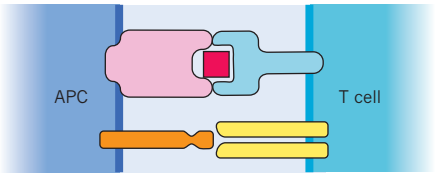
Teaching tolerance

To stop the immune system from attacking healthy tissues, as in autoimmune disease, researchers are exploring ways to induce tolerance, essentially achieving the opposite goal of a conventional vaccine. They focus on how immune cells called T cells are trained to recognize antigens — peptides and other molecules that the immune system can identify as a threat.

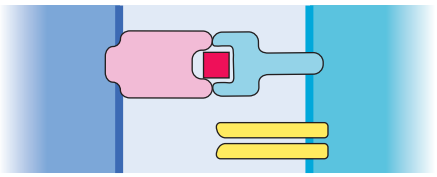


Sentinels called antigen-presenting cells (APCs) engulf and chop up biological molecules, and present them on specialized surface proteins.

A T cell has a receptor that binds to the presented antigen.



In conventional vaccination, the antigen generally appears along with other 'co-stimulatory' molecules on the surface of the APC. This combination puts the T cells on the alert and helps immunize against anything bearing that antigen.



Without co-stimulation — or in the presence of other inhibitory signals — the T cell dies, becomes inactivated or takes on a regulatory role, preventing other immune cells from attacking healthy tissue. This is the goal in antigen-specific tolerance.

doses either under or into the skin. Those in the latter group showed a large reduction in MS activity, says Wraith, although this returned three months after the treatment ended. The data have not yet been published.

Steinman's 80-patient, placebo-controlled, BHT-3021 diabetes trial established that the drug was safe, and one dosage group showed a statistically significant increase in a marker of insulin secretion and a simultaneous decrease in the number of T cells targeting proinsulin[11].

Wiley believes the treatment has helped. During the 12-week study, he says, he regained control of his insulin dosing. And, subjectively at least, the positive effects have persisted.

MOVING FORWARD

But the legacy of failure has left many researchers — and drug companies — deeply sceptical. Even after positive trial results, Martin says, there are "very few investigators that are masochistic enough over a long time to continue in this field, because it's difficult to get funding". He and Selmaj are trying to get support for their new MS trials. And Steinman and the company he co-founded, Tolerion of Portola Valley, California, are seeking between US$20 million and $30 million to take their diabetes vaccine to the next stage. "It's impossible to say when this is going to happen," Steinman says. His first diabetes trial was sponsored by the biotech giant Genentech, based in South San Francisco, California. But after acquiring the company in 2009, Swiss drug firm Roche sought to get out of type 1 diabetes therapies, Steinman says. The company returned the therapy's licence even before the trial ended, and terminated all future commitments.

Others have had more success. Another Swiss drug company, Merck Serono, has already moved Wraith's therapy into phase II to confirm its efficacy. Miller, meanwhile, is planning a phase I trial of a variation on his original approach that uses biodegradable nanoparticles instead of blood cells as antigen carriers[12]. And Chicago-based COUR pharmaceuticals — co-founded by Miller — has signed partnerships with two drug companies for trials in people with diabetes and coeliac disease.

The next round of trials, Nepom says, should incorporate detailed mechanistic studies "in order to learn whether we've chosen the right antigen, and whether we've chosen the right dose and route" in treated patients. This is because, in diseases such as MS and diabetes, researchers can only guess which antigen triggers the immune response at any point in time. "That's an incredibly challenging problem," says Nepom.

He heads the Immune Tolerance Network (ITN), a US-led international consortium that allocates roughly $27 million a year from the National Institutes of Health for clinical trials and related studies. In his view, antigen therapies should be used in combination with drugs from a second major category of tolerance treatment. Unlike antigen-specific therapies, these treatments do not inactivate, convert or destroy the relatively small number of T cells that attack a given antigen. Instead, they are designed to tilt the balance of the body's T-cell repertoire away from the subtypes that promote inflammation and towards those that maintain a state of tolerance, without impairing normal immunity against pathogens. These tolerance treatments have worked only temporarily, if at all, in autoimmunity trials. Nepom thinks that combining the two approaches will produce longer-lasting results. The ITN will soon require that all of the antigen trials it supports be combination trials.

"That's definitely the path to go," agrees Benoist. "Whether that will work or not, who knows. But at least that's a more rational way of doing things."

But pushing combinations does not sit well with some in the field. Miller, for example, doesn't think that the second approach achieves true tolerance, and he fears that combining it with antigen therapies will convolute the results. "If you want to test tolerance, it really has to be done as an individual entity," he says. Wraith thinks that adding the other class of treatment may even interfere with antigen-specific tolerance.

But the combination approach should reduce risk, say Nepom and ITN founder Jeff Bluestone, an immunologist at the University of California, San Francisco. This is because shifting the overall system towards tolerance should help to blunt any unexpected reaction to antigens.

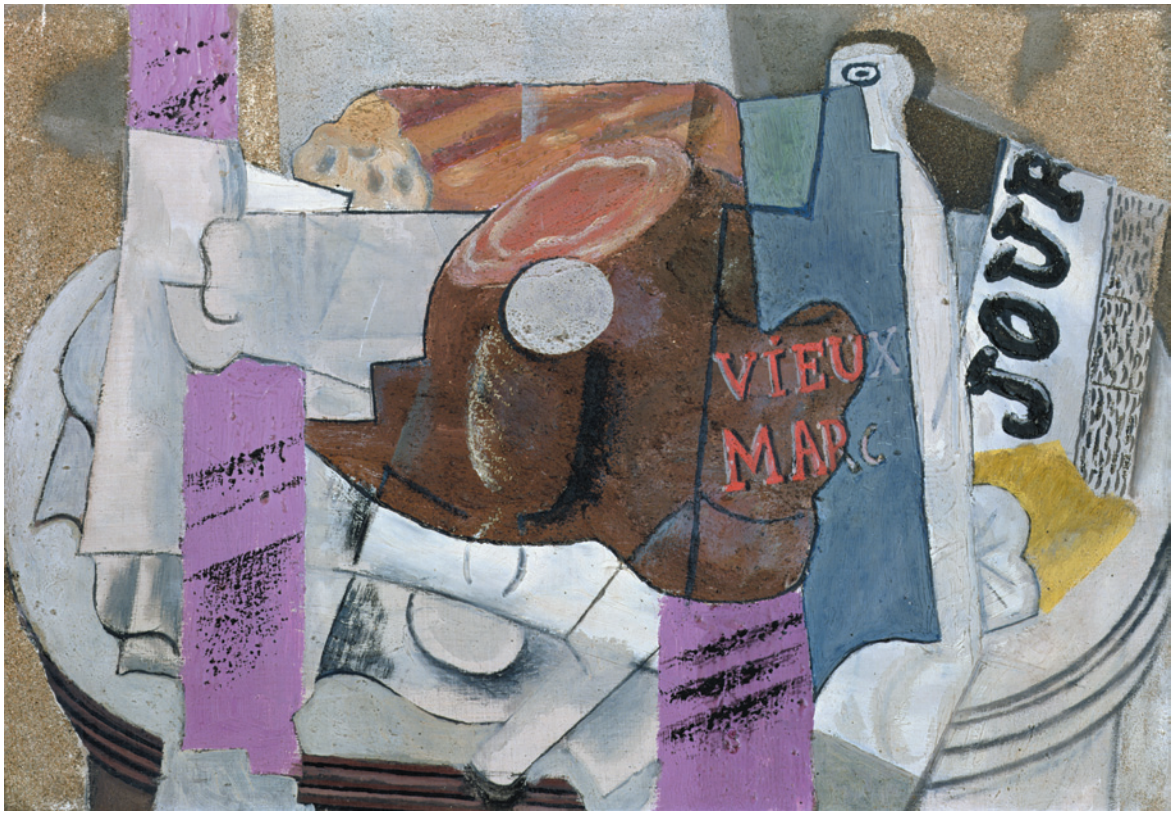
Miller and Wraith say that their treatments have already proved safe in patients, and Wraith debated with Bluestone during a meeting in the Netherlands last October. Wraith calls Bluestone's concerns unfounded. Bluestone's reply: "I hope he's right."

Wiley, for his part, had no problems while taking BHT-3021. The only side effect he noticed was the welcome disappearance of some painful plantar warts. He thinks that BHT-3021 stabilized his body's ability to produce insulin, and would gladly take part in a longer study. "No question," he says. "I would jump at the opportunity." ■

Ken Garber *is a science writer in Ann Arbor, Michigan.*

1. Getts, D. R. *et al. J. Immunol.* **187,** 2405–2417 (2011).
2. Gabryšová, L. *et al. J. Exp. Med.* **206,** 1755–1767 (2009).
3. Juryńczyk, M. *et al. Ann. Neurol.* **68,** 593–601 (2010).
4. Solvason, N. *et al. J. Immunol.* **181,** 8298–8307 (2008).
5. Suntharalingam, G. *et al. N. Engl. J. Med.* **355,** 1018–1028 (2006).
6. Bielekova, B. *et al. Nature Med.* **6,** 1167–1175 (2000).
7. Freedman, M. S. *et al. Neurology* **77,** 1551–1560 (2011).
8. McRae, B. L., Vanderlugt, C. L., Dal Canto, M. C. & Miller, S. D. *J. Exp. Med.* **182,** 75–85 (1995).
9. Lutterotti, A. *et al. Sci. Transl. Med.* **5,** 188ra75 (2013).
10. Walczak, A., Siger, M., Ciach, A., Szczepanik, M. & Selmaj, K. *J. Am. Med. Assoc. Neurol.* **70,** 1105–1109 (2013).
11. Roep, B. O. *et al. Sci. Transl. Med.* **5,** 191ra82 (2013).
12. Getts, D. R. *et al. Nature Biotechnol.* **30,** 1217–1224 (2012).

COMMENT



MUSEE D'ART MODERNE DE LA VILLE DE PARIS, PARIS, FRANCE/GIRAUDON/THE BRIDGEMAN ART LIBRARY/© SUCCESSION PICASSO/DACS, LONDON 2014

Pablo Picasso, *Le Vieux Marc* (oil on canvas), 1912.

QBism puts the scientist back into science

A participatory view of science resolves quantum paradoxes and finds room in classical physics for 'the Now', says **N. David Mermin**.

Physical science describes the objective external world: particles, waves and fields; how they change in time; and how they give rise to the forms of matter, terrestrial and extraterrestrial, microscopic and macroscopic. This world makes itself known to each of us through our own private internal perceptions. Yet physical science has ignored the 'subject' — the scientist — even though their subjective experience constitutes their only link with the external world.

In *Nature and the Greeks*[1], Austrian physicist Erwin Schrödinger traced the removal of the subject from science back more than two millennia. Alongside the spectacular success of physical science, this exclusion of personal experience has given rise to some vexing and persistent puzzles and paradoxes.

Two such unrelated long-standing problems are both resolved by recognizing that the perceiving subject has as important a role to play in understanding the nature of physical science as does the perceived object.

The first problem is the notorious disagreement, confusion and murkiness that for almost a century has plagued the foundations of quantum mechanics, in spite of the theory's extraordinary usefulness and power. The second, less famous, problem has been with us at least as long: there seems to be nothing ▶

in physics that singles out 'the present moment'. Albert Einstein called this the problem of 'the Now'. Both problems are symptoms of the exclusion from physical science of the perceiving subject, and are solved by restoring what the ancient Greeks removed.

QUANTUM MECHANICS

Schrödinger wrote in a little-known 1931 letter[2] to German physicist Arnold Sommerfeld that quantum mechanics "deals only with the object–subject relation". Another founder of quantum mechanics, Danish physicist Niels Bohr, insisted in a 1929 essay[3] that the purpose of science was not to reveal "the real essence of the phenomena" but only to find "relations between the manifold aspects of our experience".

In spite of these early hints, it was only in the twenty-first century that US physicist Christopher Fuchs and British–German physicist Rüdiger Schack[4–6] put forth an understanding of quantum mechanics that restored the balance between subject and object. They call their new point of view 'QBism': Q is for quantum and B is for Bayesian — a view of probability that includes an agent who makes bets and updates odds. QBism attributes the muddle at the foundations of quantum mechanics to our unacknowledged removal of the scientist from the science.

Much of this muddle is associated with the 'wavefunction' that quantum mechanics assigns to a physical system. This irritatingly uninformative term reveals the lack of clarity present in the field from its very beginning in 1925. People argue to this day about whether wavefunctions are real entities, like stones or ripples on a pond, or mathematical abstractions that help us to organize our thinking, like the calculus of probabilities.

Fuchs and Schack adopt the latter view. They take a wavefunction to be associated with a physical system by an agent — me, for example, based on my past experience. I use the wavefunction, following rules laid down by quantum mechanics, to calculate the likelihood of what I might experience next, should I choose to probe further. Depending on what I then perceive, I can update the wavefunction on the basis of that experience, allowing me to better assess my subsequent expectations.

People who believe wavefunctions to be as real as stones have invested much effort in searching for objective physical mechanisms responsible for such changes in the wavefunction: a novel manifestation of gravity, for example, or a new kind of fundamental all-pervasive friction. But according to QBism, the change is only in my personal expectations, which I revise to accommodate my new experience.

Another celebrated part of the muddle produced by the exclusion of the perceiving subject is 'quantum non-locality', the belief of some quantum physicists and many mystics, parapsychologists and journalists that an action in one region of space can instantly alter the real state of affairs in a faraway region. Thousands of papers have been written about this mysterious action at a distance over the past 50 years. A clue that the only change is in the expectations of the perceiving subject[7] is that to learn anything about such alterations one must consult somebody in the region where the action took place.

Most physicists who have paid some attention to QBism have rejected this explicit introduction of subjective personal experience into science, together with its consequences for our understanding of quantum physics. It offends their sense that science is strictly objective.

"The Now is neither an illusion nor a spurious manifestation of temporal chauvinism."

QBists are often charged with solipsism: a belief that the world exists only in the mind of a single agent. This is wrong. Although I cannot enter your mind to experience your own private perceptions, you can affect my perceptions through language. When I converse with you or read your books and articles in *Nature*, I plausibly conclude that you are a perceiving being rather like myself, and infer features of your experience. This is how we can arrive at a common understanding of our external worlds, in spite of the privacy of our individual experiences.

THE NOW

The QBist conversation can be broadened to include issues in which neither quantum mechanics nor probability plays a part, such as the problem of the Now, which arises in purely classical (pre-quantum) physics. I change the term to 'CBism' when describing applications of the QBist view of science in such classical settings. Here C stands for classical and B, for Bohr, whose wisdom went beyond quantum mechanics when he taught that physical science studies our experience.

Philosopher Rudolph Carnap[8] recalls that the problem of the Now worried Einstein seriously. Einstein told him that the experience of the present moment means something special for mankind, essentially different from the past and the future, and that physics cannot describe such a difference. Carnap described Einstein as painfully resigned to the inability of science to grasp this experience.

The issue for Einstein was not the famous revelation of relativity that whether or not two events in two different places happen at the same time can depend on your frame of reference. It was simply that physics seems to offer no way to identify the Now even at a single event in a single place, although a local present moment — Now — is evident to each and every one of us as undeniably real. How can there be no place in physics for something as obvious as that?

My Now — my current state of affairs — is a special event for me while it is happening. I can tell my Now from earlier events, which I only remember, and from later events which I can only anticipate or imagine. The status of an event as my Now is transitory: it becomes a memory as subsequent Nows emerge.

Yet clear, evident and banal as this is to us all, there is no Now in the usual physical description of space and time. Physicists represent all the events experienced by a single person as a line in four-dimensional space-time, called that person's 'world-line'. There is nothing about any point on my world-line that singles it out as my Now.

When I recently mentioned to an eminent theoretical physicist that I was writing an essay explaining how the QBist view of science solves the strictly classical problem of the Now, he said: "Ah, you're going to explain why we all have that illusion." And a distinguished philosopher of science recently derided the attitude that there ought to be a Now on my world-line as "chauvinism of the present moment"[9].

But the Now is neither an illusion nor a spurious manifestation of temporal chauvinism. The problem of the Now is laid to rest by recognizing the mistake behind the conclusion that it is missing from our physical description of the world. That is the very error that led us into the quantum muddle: the exclusion of personal experience from physical science. Einstein's pain at the inability of science to contain a Now was of a piece with his stubborn refusal to accept quantum mechanics as an adequate view of the world.

Physicists reify space-time. They elevate it from a four-dimensional diagram used to record their experience into the kind of "real essence" that Bohr warned us not to seek. My space-time diagram lets me represent events from past experiences, along with deductions or conjectures about events that were not experienced or have yet to happen. By identifying my abstract diagram with an objective reality, I fool myself into regarding that diagram as a four-dimensional arena in which my life is lived. Actual experiences are spread out in time and in space, and actual clocks used to associate times with our experiences are extended physical objects. To represent the rich spatio-temporal structure of human experience as mathematical points in a space-time continuum is a smart strategic simplification, but we ought not to confuse our actual experience with a cartoon.

That there is a place for the present moment in physics becomes obvious when I take my experience of it as the reality it clearly is to me and recognize that space-time is an abstraction that I construct to organize such

experiences. At any moment I can represent my past experience as my world-line, terminating in my Now. As it turns into a memory, I expand my diagram to contain my subsequent Nows. The motion of my Now along my world-line reflects the fact that as my watch advances I acquire more experiences to record.

This provides the place in physics for the Now of any one person. But could the problem of the Now lie in relating the present moments of several different people? When you and I are communicating face-to-face I cannot imagine that a live encounter for me could be only a memory for you, or vice versa. When two people are together at an event, if the event is Now for one of them, then it is Now for both. Although this is only an inference for each person, I take it to be as fundamental a feature of two perceiving subjects as the Now is for a single subject.

Our present moments must overlap at every one of our meetings — whenever we have a conversation, move apart and then come back together and have another conversation. But throughout human history people have only moved at low speeds. The complicating effect of relativistic 'time dilation' — the slowing down of rapidly moving clocks — on the advances of our different individual Nows has been far too small to notice. We can, however, entertain the question of whether our present moments would coincide when we came back together, regardless of how rapidly we moved back and forth and regardless of how long the journey.

It is a basic fact of relativity that my personal time — the progress of my present moment — keeps pace with the reading of my watch. If it did not, I would be aware that the rate of my watch had changed as it moved with me, in violation of Einstein's (and Galileo's) principle of relativity. This is all we need. Consider two twins. When they are together at home, their Nows coincide. Then Alice flies off to a nearby star at 80% of the speed of light, turns around and flies back home to Bob at the same speed. Relativity requires that if Bob's watch has advanced ten years in the meantime, Alice's has advanced only six. But because each of their present moments has advanced in step with the watch each is carrying, the moment of their reunion continues to be Now for them both.

So it is incorrect to claim that physics has nothing to say about local Nows at single events. Physics predicts that our experiences of the Now will continue to have the same familiar features in a future world of interstellar travel at speeds near the speed of light, even for the distinct Nows of many different agents.

Because it solves diverse conundrums in quantum mechanics as well as in the strictly classical problem of the Now, QBist (or CBist) thinking needs to be taken more seriously by physicists. It is time to consider what other foundational puzzles can be resolved by restoring the balance between subject and object in physical science.

As another Viennese investigator even more famous than Schrödinger — Sigmund Freud — put it in 1927 (ref. 10): "The problem of a world constitution that takes no account of the mental apparatus by which we perceive it is an empty abstraction." ■

N. David Mermin *is emeritus professor in the Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York, USA. He started to take QBism seriously while at the Stellenbosch Institute for Advanced Study, South Africa.*
e-mail: david.mermin@cornell.edu

1. Schrödinger, E. *Nature and the Greeks and Science and Humanism* (Cambridge Univ. Press, 1996).
2. Schrödinger, E. *Eine Entdeckung von ganz ausserordentlicher Tragweite* (ed. von Meyenn, K.) 490 (Springer, 2011).
3. Bohr, N. *Atomic Theory and the Description of Nature* 18 (Cambridge Univ. Press, 1934).
4. Caves, C. M., Fuchs, C. A. & Schack, R. *Phys. Rev. A* **65,** 022305 (2002).
5. Fuchs, C. A. Preprint at http://arxiv.org/abs/1003.5209 (2010).
6. Fuchs, C. A. & Schack, R. *Rev. Mod. Phys.* **85,** 1693–1715 (2013).
7. Fuchs, C. A., Mermin, N. D. & Schack, R. Preprint at http://arxiv.org/abs/1311.5253 (2013).
8. Carnap, R. *The Philosophy of Rudolf Carnap* (ed. Schilpp, P. A.) 37–38 (Open Court Publishing, 1963).
9. Price, H. *Science* **341,** 960–961 (2013).
10. Freud, S. *The Future of an Illusion,* in *Mass Psychology and Other Writings* (Penguin, 2004).

Make mouse studies work

More investment to characterize animal models can boost the ability of preclinical work to predict drug effects in humans, says **Steve Perrin**.

Mice take the blame for one of the most uncomfortable truths in translational research. Even after animal studies suggest that a treatment will be safe and effective, more than 80% of potential therapeutics fail when tested in people. Animal models of disease are frequently condemned as poor predictors of whether an experimental drug can become an effective treatment. Often, though, the real reason is that the preclinical experiments were not rigorously designed[1,2].

The series of clinical trials for a potential therapy can cost hundreds of millions of dollars. The human costs are even greater: patients with progressive terminal illnesses may have just one shot at an unproven but promising treatment. Clinical trials typically require patients to commit to year or more of treatment, during which they are precluded from pursuing other experimental options. Launching a clinical trial without the backing of robust animal data keeps patients out of tests for therapies that may have a better chance of success.

One such group of patients is those with amyotrophic lateral sclerosis (ALS), the fatal neurodegenerative condition also known as Lou Gehrig's or motor neuron disease. Over the past decade, about a dozen experimental treatments have made their way into human trials for ALS. All had been shown to ameliorate disease in an established animal model. All but one failed in the clinic, ▶

ILLUSTRATIONS BY CLAIRE WELSH/*NATURE*

and the survival benefits of that one are marginal.

At the ALS Therapy Development Institute (TDI) in Cambridge, Massachusetts, we have tested more than 100 potential drugs in an established mouse model of this disease (mostly unpublished work). Many of these drugs had been reported to slow down disease in that same mouse model; none was found to be beneficial in our experiments (see 'Due diligence, overdue'). Eight of these compounds ultimately failed in clinical trials, which together involved thousands of people. One needs to look no further than potential blockbuster indications such as Alzheimer's and cancer to see that the problem persists across diseases.

After nearly a decade of validation work, the ALS TDI introduced guidelines that should reduce the number of false positives in preclinical studies and so prevent unwarranted clinical trials. The recommendations, which pertain to other diseases too, include: rigorously assessing animals' physical and biochemical traits in terms of human disease; characterizing when disease symptoms and death occur and being alert to unexpected variation; and creating a mathematical model to aid experimental design, including how many mice must be included in a study. It is astonishing how often such straightforward steps are overlooked. It is hard to find a publication, for example, in which a preclinical animal study is backed by statistical models to minimize experimental noise.

The experiments necessary for this type of characterization are expensive, time-consuming and will not, in themselves, lead to new treatments. But without this upfront investment, financial resources for clinical trials are being wasted and lives are being lost.

KNOW YOUR ANIMALS

Investigations at the ALS TDI exemplify how initial physiological descriptions of an animal model rarely encompass all salient features, including how closely the model captures what is observed in patients. Such models are often inadequate for studying how a drug affects various aspects of disease.

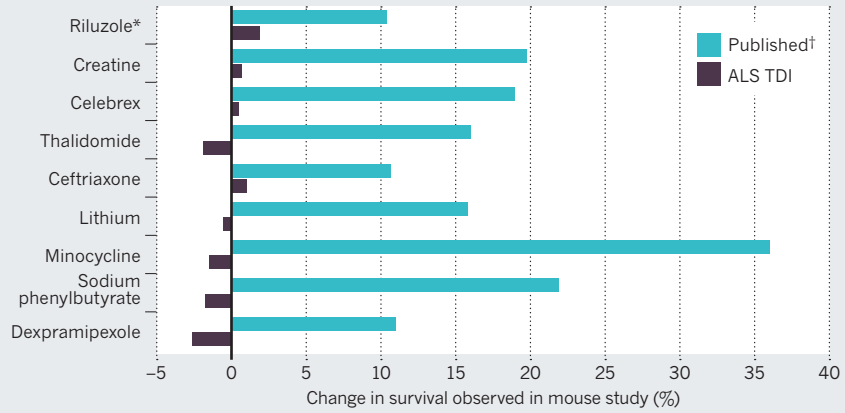
ALS progression is characterized by a deterioration in the neurons that innervate skeletal muscles. Sequencing and genetic studies implicate RNA-binding proteins as crucial for maintaining the health of motor neurons[3]. Mouse models expressing a mutant form of the RNA binding protein TDP43 show hallmark features of ALS: loss of motor neurons, protein aggregation and progressive muscle atrophy[4].

But further study of these mice revealed key differences. In patients (and in established mouse models), paralysis progresses over time. However, we did not observe this progression in TDP43-mutant mice. Measurements of gait and grip strength showed that their muscle deficits were in fact mild, and post-mortem examination found that the animals died not of progressive muscle atrophy, but of acute bowel obstruction caused by deterioration of smooth muscles in the gut[5]. Although the existing TDP43-mutant mice may be useful for studying drugs' effects on certain disease mechanisms, a drug's ability to extend survival would most probably be irrelevant to people.

Scientists who use animal models for translational research must proceed with caution, and be prepared to do further characterizations themselves.

DUE DILIGENCE, OVERDUE

Results of rigorous animal tests by the Amyotrophic Lateral Sclerosis Therapy Development Institute (ALS TDI) are less promising than those published. All these compounds have disappointed in human testing.



*Although riluzole is the only drug currently approved by the US Food and Drug Administration for ALS, our work showed no survival benefit.
†References for published studies can be found in supplementary information at go.nature.com/hf4jf6.

CANCEL THE NOISE

ALS TDI scientists performed a meta-analysis on nearly 5,500 mice that had been used in treatment or control groups over four years[1]. All mice expressed a specific defective version of the SOD1 gene, which is mutated in about 10% of people with inherited ALS. This work, and that of others[6], revealed both unexpected variation in the animals and ways to control for it.

Almost 90% of the mice had an average lifespan of 134 days, give or take 10 days. Careful inspection of animals that lived shorter or longer revealed four factors that produced considerable noise in the data and could have led to spurious conclusions (see 'Four ways to fight noise'). Crucially, understanding such variation requires careful monitoring of hundreds of mice over several generations.

One factor is the failure to exclude animals whose deaths are unrelated to the disease being studied. Other factors are failing to split littermates between control and treatment groups, and not taking gender into account. Male SOD1 mice show symptoms as much as a week before females and die about a week earlier. Given that a week is a 4% variability in survival, such differences could easily be misconstrued as a drug effect.

The fourth factor regards the genes introduced to induce disease. All too often, a disease phenotype is lost as a colony of breeding mice is built up. For many diseases, including ALS, animal models carry multiple copies of the disease-causing gene, and these repeated genes are often not passed on in a stable fashion as cells divide to make gametes. Regular genotyping assays are essential to make sure that mice in subsequent generations do

FOUR WAYS TO FIGHT NOISE

Simple steps to avoid spurious conclusions

- **Exclude irrelevant animals** As often done in clinical trials, subjects that die for reasons unrelated to disease (such as mishandling) should not be counted in results. Reasons for exclusion should be well documented.
- **Balance for gender** Males and females can show differences in symptoms that obscure modest drug effects.
- **Split littermates among experimental groups** Putting siblings into the same treatment group can bias results.
- **Track genes** Genes that induce disease are often not inherited reliably. When copies are lost, symptoms can be less severe and drugs can seem more effective than they are.

not have fewer copies of the transgene, and therefore less severe disease.

At the ALS TDI we have seen this several times. When first described in 2010, all TDP43-mutant mice died within 200 days[7]. When we ordered mice from a breeding colony established from those used in this initial publication, the mice lived for up to 400 days without showing signs of disease. To perform the characterization work on TDP43 described above, we first spent several months backcrossing the strain to create a stable phenotype.

Characterization can flag more subtle potential problems for translation. This is exemplified by a study showing that lithium can boost survival of SOD1 mice by 30 days, an astoundingly long time[8]. A small clinical trial showed that it also extended life in people with ALS[8]. Lithium is already sold to treat schizophrenia, and many people with ALS began taking the drug off label in hope of slowing down their disease progression. Three separate phase III clinical trials were launched in parallel to assess the drug's effects. These enrolled hundreds of patients with a total cost of well over US$100 million. None of the three trials showed any therapeutic benefit[9–11].

Concurrently, other groups attempted to reproduce the preclinical data and could not[12,13]. Although it is difficult to determine why the first study showed such a dramatic effect, its initial results are curious. The median survival time of untreated animals was 20 days shorter than that observed elsewhere, suggesting other anomalies.

For studies that aim to predict treatment benefits, such as extended survival or a delay of symptom progression, a mathematical simulation is in order. This incorporates the variation typically observed in an animal model to calculate how many animals should be assigned to the experimental groups. According to our calculations, highly variable animal models could require hundreds of animals per group; even homogeneous ones require as many as ten.

And before assessing a drug's efficacy, researchers should investigate what dose animals can tolerate, whether the drug reaches the relevant tissue at the required dose and how quickly the drug is metabolized or degraded by the body. We estimate that it takes about $30,000 and 6–9 months to characterize the toxicity of a molecule and assess whether enough reaches the relevant tissue and has a sufficient half life at the target to be potentially effective.

"Public and private agencies should fund characterization studies as a specific project."

If those results are promising, then experiments to test whether a drug can extend an animal's survival are warranted — this will cost about $100,000 per dose and take around 12 months. At least three doses of the molecule should be tested; this will help to establish that any drug responses are real and suggest what a reasonable dosing level might be.

Thus, even assuming the model has been adequately characterized, an investment of $330,000 is necessary just to determine whether a single drug has reasonable potential to treat disease in humans. This seems worthwhile given that it could take thousands of patients, several years and hundreds of millions of dollars to move a drug through the clinical development process.

COMMUNITY EFFORT

As academic labs shift their focus to translational research, the burden to characterize animal models will fall on them. Although the costs are meagre compared with those of clinical trials, the investment required in time and funds is far beyond what any one lab should be expected to do. This burden and the resulting mouse models should be shared. At the very least, researchers should place new animal models in a public repository so that other teams can repeat the characterization, and share the costs of doing it well.

Public and private agencies should fund characterization studies as a specific project. A good example is the Alzheimer's Disease Neuroimaging Initiative, a large, collaborative study to find diagnostic biomarkers of the disease. Competitive bidding and milestone-driven payments could persuade qualified groups to perform the necessary experiments and to make results publicly available. This is unglamorous work that will never directly lead to a breakthrough or therapy, and is hard to mesh with the aims of a typical grant proposal or graduate student training programme. However, without these investments, more patients and funds will be squandered on clinical trials that are uninformative and disappointing. ■

Steve Perrin *is chief scientific officer at the ALS Therapy Development Institute in Cambridge, Massachusetts, USA.*
e-mail: sperrin@als.net

1. Scott, S. *et al. Amyotroph. Lateral Scler.* **9,** 4–15 (2008).
2. Begley, C. G. & Ellis, L. M. *Nature* **483,** 531–533 (2012).
3. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. *Neuron* **79,** 416–438 (2013).
4. Wegorzewska, I., Bell, S., Cairns, N. J., Miller, T. M. & Baloh, R. H. *Proc. Natl Acad. Sci. USA* **106,** 18809–18814 (2009).
5. Hatzipetros, T. *et al. Brain Res.* http://dx.doi.org/10.1016/j.brainres.2013.10.013 (2013).
6. Ludolph, A. C. *et al. Amyotroph. Lateral Scler.* **11,** 38–45 (2010).
7. Stallings, N. R., Puttaparthi, K., Luther, C. M., Burns, D. K. & Elliott, J. L. *Neurobiol. Dis.* **40,** 404–414 (2010).
8. Fornai, F. *et al. Proc. Natl Acad. Sci. USA* **105,** 2052–2057 (2008).
9. UKMND-LiCALS Study Group *Lancet Neurol.* **12,** 339–345 (2013).
10. Verstraete, E. *et al. J. Neurol. Neurosurg. Psychiatr.* **83,** 557–564 (2012).
11. Aggarwal, S. P. *et al. Lancet Neurol.* **9,** 481–488 (2010).
12. Gill, A., Kidd, J., Vieira, F., Thompson, K. & Perrin, S. *PLoS ONE* **4,** e6489 (2009).
13. Pizzasegola, C. *et al. Amyotroph. Lateral Scler.* **10,** 221–228 (2009).

RIA NOVOSTI/SPL

J. D. Bernal (third from left) helped to define the 'science of science' as a discipline.

IN RETROSPECT

The Social Function of Science

Roger Pielke Jr assesses the legacy of J. D. Bernal's science-policy classic on its 75th anniversary.

In 2011 Tom Coburn, the Republican Senator for Oklahoma, issued a report focused on helping the US National Science Foundation to better conduct research that "can transform and improve our lives, advance our understanding of the world, and create meaningful new jobs". It is ironic that this conservative Republican's demands that research be carefully planned and focused on social objectives can be traced directly to the writings of an Irish-born communist crystallographer 75 years ago.

Such is the wholesale acceptance of John Desmond Bernal's views in his 1939 treatise *The Social Function of Science* — covering the organization of research to science and its social role — that they are now part of the fabric of science-policy debates across the political spectrum. For Bernal, usefulness was more than an aspiration: it was the central objective of the scientific enterprise and the desired end of state support of science.

He was among the first to recognize that all public engagement is ultimately political, although his vision of scientists as stalwarts resisting partisan politics might now seem naive: "The scientist … sees the social, economic and political situation as a problem to which a solution must first be found and then applied, not as a battleground of personalities, careers and vested interests."

Bernal was the first to compile estimates of government-wide spending on science, several years before the first gauges of gross domestic product in the early 1940s. On the basis of such estimates, he concluded presciently that the United States was poised to take a long-term leadership role in science. Today, much discussion of science policy (some would say too much) hinges on this kind of number-crunching; 75 years ago, it provided a fundamentally new lens through which to view the scientific enterprise.

Bernal was born in Ireland in 1901. His formative years were marked by the First World War and the 1917 Russian revolution, which, along with the Great Depression in the 1930s, had a lasting negative influence on his view of capitalism. After earning a degree in mathematics and science at the University of Cambridge, UK, in 1922, Bernal did his postgraduate training in X-ray crystallography before joining the Cambridge faculty in 1927. He became part of Britain's left-wing intellectual elite, joining zoologist Solly Zuckerman's dining club Tots and Quots (a reference to Roman playwright Terence's "*Quot homines, tot sententiae*", meaning 'so many men, so many opinions') with biologists Julian Huxley and J. B. S. Haldane, among others. Zuckerman became Britain's first chief scientific adviser in 1964.

The Social Function of Science
J. D. BERNAL
George Routledge and Sons: 1939

Bernal started to write *The Social Function of Science* in 1938 after having "achieved a certain standing in the scientific community", according to his biographer Andrew Brown (*The Sage of Science*; Oxford University Press, 2007). He was far from the first to explore the nexus of science and society. For instance, the theme of the 1936 meeting of the British Science Association was 'Science and Social Welfare', and in 1937 the American Association for the Advancement of Science added "an examination of the profound effects of science upon society" as one of its objectives. Even so, Bernal's book helped to define a new discipline: the science of science.

The book was controversial for two reasons. First, Bernal was presenting a view of science that was directly at odds with the 'pure science' ideal, in which scientists were expected to keep their distance from public affairs. Second was Bernal's vision of science fulfilling its social function by supporting a centrally planned society. He even stated that "science is communism" and argued that the Soviet Union "was one State where the proper function of science was being realized".

In the debates that raged in scientific and other circles, which pitted scientific planning against freedom — essentially, socialist versus capitalist stances — Bernal's arguments were often conflated with his support for the Soviet Union, frequently with his encouragement. A 1946 *Nature* Editorial sought, and largely failed, to find a middle ground: "for though we claim that the impact of science on society is now achieving such importance as to command constant study which is bound to result in conscious planning, we are equally as convinced that it is the man of science who must be allowed to do the planning in consultation with others. In this way his freedom need not be impaired" (*Nature* **158,** 565–567; 1946).

Bernal's great intellectual adversary was Michael Polanyi, a Hungarian chemist who was opposed to Soviet ideals. Polanyi's classic 1962 journal article 'The Republic of Science: Its Political and Economic Theory' (*Minerva* **1,** 54–73; 1962) posits that individual scientists pursuing truth led to the most efficient social outcomes. The parallels with Adam Smith's "invisible hand" guiding capitalist economies could not have been accidental.

The decades-long debate between Bernal and Polanyi played out in many contexts. ▶

For instance, in 1943, when a more-organized approach to government-supported science was first being mooted in the United States, a Democratic Senator for West Virginia, Harvey Kilgore, proposed creating an office to "mobilize the scientific and technical resources of the nation". A correspondence in *The New York Times* warned that this suggestion for government-directed science represented "the latent germs of a form of communism that is most foreign to the ideals which our peoples are fighting to support". After the Second World War, engineer Vannevar Bush, who oversaw the seminal 1945 report *Science — The Endless Frontier* on US science policies, echoed Polanyi's thinking, arguing for public resources with little public accountability.

In the late 1940s, Bernal's career began to fade with his endorsement of the theory of agricultural genetics propounded by Russian agronomist Trofim Lysenko. This was elevated to Soviet policy, yet quickly discredited by scientists in the West. In 1948, Bernal debated against Polanyi on BBC radio, with Polanyi pointing to the suppression of science under Lysenkoism as evidence of the shortfalls of the state planning of science. The BBC's subsequent investigation revealed that nine Soviet geneticists had been put to death for expressing alternative views. Bernal's continuing defence of the theory, and his glowing 1953 obituary of Joseph Stalin in the magazine *Modern Quarterly*, contributed to his diminishing relevance.

Although Bernal lost the intellectual battle over cold-war politics, his ideas on the social function of science have triumphed on nearly every count. The larger and more significant effect of *The Social Function of Science* has been to anticipate and help the ideal of 'pure science' to reach mythical status, ushering in an era of science focused on societal needs, today characterized as 'grand challenges' by scientists and politicians.

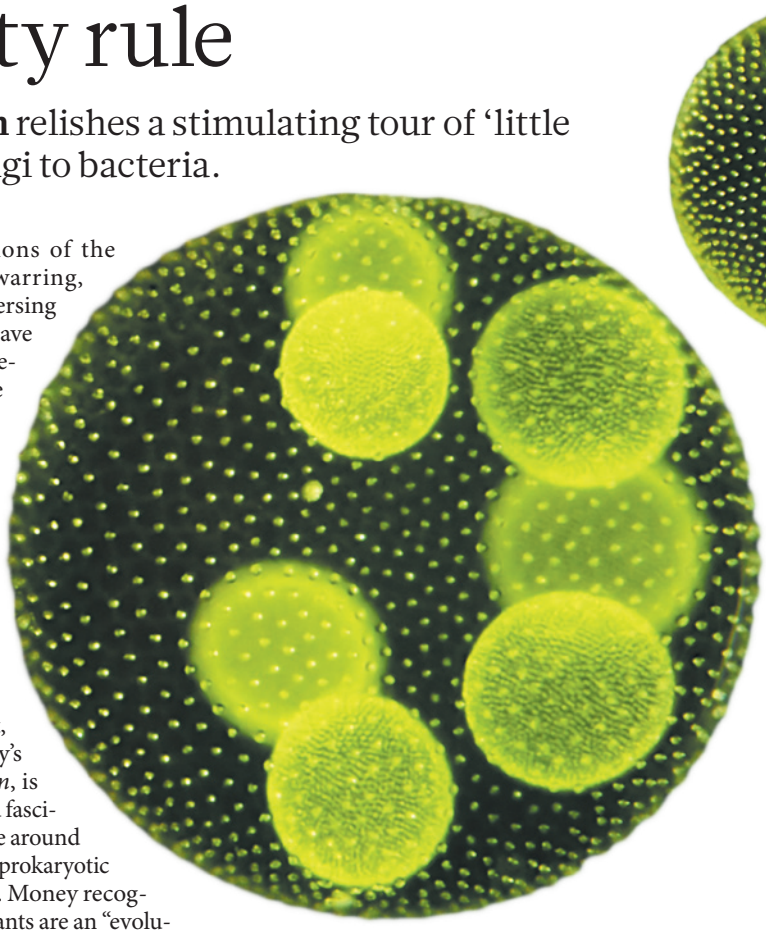
Bernal, looking back at the book 25 years after its publication, emphasized that such needs implied more systematic thinking about science itself: "We need a strategy for research which must be based on a science of science." Today, whether it is a US senator, a British prime minister or a Chinese president expressing a view that science must serve society, each is reflecting Bernal's big idea. Seventy-five years on, that global influence on science-policy thinking is his classic's great legacy. ■

Roger Pielke Jr *is director of the Center for Science and Technology Policy Research and professor of environmental studies at the University of Colorado Boulder.*
e-mail: pielke@colorado.edu

MICROBIOLOGY

Majority rule

Mark O. Martin relishes a stimulating tour of 'little lives', from fungi to bacteria.



Colonies of *Volvox aureus* algae, pregnant with daughter colonies.

LEBENDKULTUREN.DE/SHUTTERSTOCK

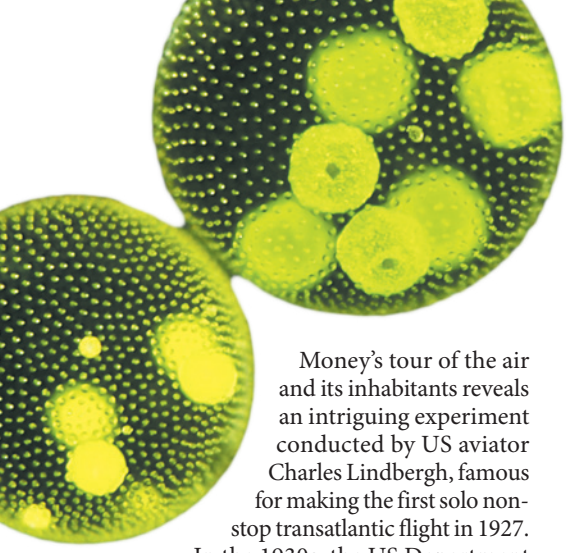
The vast populations of the microscopic — warring, cooperating, dispersing and communicating — have a dizzying array of strategies and forms. These denizens of the universe at the far end of a microscope have much to teach humanity, their roles ranging from the digestion of food to the cycling of carbon and nitrogen in Earth's atmosphere. Several books have championed matters microbial in recent years. The latest, biologist Nicholas Money's *The Amoeba in the Room*, is filled with an impassioned fascination for microscopic life around and within us, in both the prokaryotic and eukaryotic domains. Money recognizes that animals and plants are an "evolutionary afterthought" — as he writes, "the least part of life".

My interests tend toward the bacterial, viral and archaeal. Money covers this territory thoroughly, and goes beyond to sing the praises of fungi, algae and protists. Among others, he evocatively describes the water mould *Haptoglossa*, which uses microbial artillery to attack and consume nematode worms; *Cryptomonas*, a "Russian nesting doll" of an alga, evolved from four different organisms; and *Polychaos dubium*, a huge amoeba with perhaps the largest amount of DNA to be found in a nucleus. Overall, Money delivers a heady mixture of history, philosophy, art and even poetry: the chapters are prefaced with lines from John Milton's *Paradise Lost*.

The Amoeba in the Room: Lives of the Microbes
NICHOLAS P. MONEY
Oxford University Press: 2014.

Money begins with a 'macro' view of his garden pond, then dives into intricate details of the microbial populations in and around it. The seemingly uninhabited water swarms with communities as complex as any seen with the naked eye, and a tree branch is revealed as home to populations of interacting microbes. Money's point is that the diversity of life is clear enough on the macroscopic scale, but the organismal diversity of the microbial world is staggering. I particularly appreciated reading about the eight supergroups of microbial eukaryotes, including some familiar "animalcules" (such as the mitochondria-free pathogen *Giardia*, found among the Excavata).

A chapter on microscopy takes the reader from Assyrian craftsmen to familiar names such as seventeenth-century microbiological pioneers Robert Hooke and Antoni van Leeuwenhoek, as well as characters with whom I was not familiar. Bénédict Prévost, for instance, presented evidence that microbes can cause disease 50 years before Louis Pasteur, whereas Henry Baker popularized the use of microscopes to observe tiny wonders in 1742. We meet Ed Ricketts, US marine biologist and co-author of *Between Pacific Tides* (1939), as well as sculpture-like coccolithophores, numerous *Prochlorococcus* and ubiquitous marine bacteriophages. As in the rest of the book, Money enthusiastically presents evidence of diversity everywhere, no matter the magnification.



Money's tour of the air and its inhabitants reveals an intriguing experiment conducted by US aviator Charles Lindbergh, famous for making the first solo non-stop transatlantic flight in 1927. In the 1930s, the US Department of Agriculture asked Lindbergh and his wife, Anne, to gather air samples in a bid to understand the dispersal of cereal rusts. The two collected abundant rust spores using a plane fitted with a sky-hook, a device incorporating oiled microscope slides. Money also discusses how more recent researchers have speculated that some bacteria and microscopic algae control cloud formation and rainfall.

The co-evolution of human and microbe is amply explored. Those that cause disease or death — think pathogenic bacteria, flu viruses, malaria — have shaped our history. But, as researchers have discovered in the past century, the human body could not function without the estimated thousands of species of bacteria that inhabit our mouths, guts and other recesses. Like the pond and branch that Money probes early on, a human can be seen as a collection of ecological niches, albeit a mobile one.

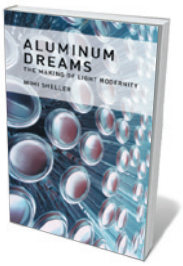
"Seemingly uninhabited water swarms with communities and complex as any seen with the naked eye."

Money goes on to discuss seemingly inhospitable environments (such as deep-sea hydrothermal vents, acidic areas or regions bombarded with high levels of radiation) in which microbes — such as archaeans that can grow at 121 °C — prosper. He sums up with distressing prospects for extinctions and change, but also reassures: the microbial world is one of prodigious diversity, power and resilience.

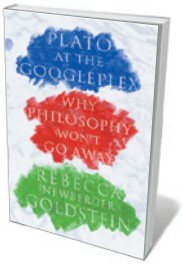
This is a lucid and informative book. There is an impressive afterword of references and notes, and fine line drawings. So much that is lyrical and little-known waits to be discovered here — novelties that will appeal to new undergraduates as well as to incorrigible microbial enthusiasts like myself. ■

Mark O. Martin *is associate professor of biology at the University of Puget Sound in Tacoma, Washington.*
e-mail: momartin@pugetsound.edu

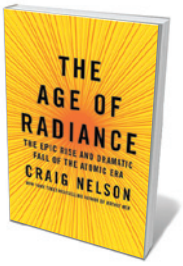
Books in brief



Aluminum Dreams: The Making of Light Modernity
Mimi Sheller MIT Press *(2014)*
It propelled humanity into air and space, transformed communications and fed war machines. Aluminium — in aeroplanes, space capsules, satellites, bombs and baking foil — is welded into the built world. Mimi Sheller's coruscating cultural study reveals how young US chemist Charles Martin Hall and his French counterpart Paul Héroult simultaneously discovered the electrolytic production of aluminium in 1886; how designers were galvanized by its potential for the light and sleek; and how social and environmental problems from bauxite mining and aluminium smelting persist.



Plato at the Googleplex: Why Philosophy Won't Go Away
Rebecca Goldstein Pantheon *(2014)*
Philosopher Rebecca Goldstein probes why Plato — and the philosophical enterprise itself — remains a force in science and culture 2,400 years on. Into a weighty discussion of the Platonic world view Goldstein inserts fictional interludes that see Plato, Chromebook in hand, touring the Googleplex, a neuroscience lab and beyond. Although contrived, this thought experiment usefully casts an eye on our turbocharged century. And it shows what survives of this classical titan: an ability to plumb the deep questions we still grapple with, from the nature of knowledge to morality.



The Age of Radiance: The Epic Rise and Dramatic Fall of the Atomic Era
Craig Nelson Scribner *(2014)*
Tomes on things atomic are rife — from discoveries by seminal physicists to meltdowns and crumbling nuclear arsenals. Craig Nelson braids the strands into a comprehensive chronicle that is also a wry elegy for an ebbing age. Starting with Wilhelm Röntgen's discovery of X-rays, Nelson gives us the panoply of physicists from Lise Meitner to Leo Szilard; offers a masterful reading of Hiroshima and Nagasaki; and takes us through weapons tests, the cold war, Fukushima, and the rest of our long attempt to "live with blessed curses".

The Thing with Feathers: The Surprising Lives of Birds and What They Reveal About Being Human
Noah Strycker Riverhead *(2014)*
Birds intrigue humanity, and in this research round-up Noah Strycker reveals why — in marvels such as the equal-radius flight paths of flocking starlings and the decontamination chamber that is a vulture's stomach. As he notes, such findings can mirror human realities. Clark's nutcrackers, for instance, can recall the location of 5,000 seed caches. Yet, in captivity some wild species lose substantial hippocampal volume (and memory) within weeks — an echo of the rapid losses found in patients with Alzheimer's disease.

The Insect Cookbook: Food for a Sustainable Planet
Arnold van Huis, Henk van Gurp and Marcel Dicke Columbia University Press *(2014)*
The looming squeeze on food security is forcing some scientists to look at the abundance of protein under our feet: insects. Here entomologists Arnold van Huis and Marcel Dicke team up with chef Henk van Gurp for a pragmatic introduction to entomophagy, covering insect farming, nutrition and cuisine. Tarte tatin with chocolate-coated grasshoppers? With 2 billion of us already popping mealworms and more, this is a case of joining the crowd. **Barbara Kiser**

RICHARD GRAULICH/ZUMA PRESS/CORBIS

Before symbols, complex formulae were expressed in words.

MATHEMATICS

Numbers game

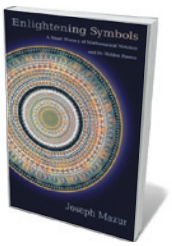
George Szpiro enjoys a history of the slow but irresistible rise of mathematical symbols.

Solving an equation such as $2x - 4 = 10$ is almost ludicrously easy. The symmetry is visible; we just perform identical manipulations on both sides of the equals sign until the symbol for the unknown value, x, remains isolated on the left. But it was not always that simple, as mathematician Joseph Mazur reveals in his history of maths notation, *Enlightening Symbols*.

Until the late Renaissance, there were few symbols to depict clearly and concisely what would later become known as an equation. Numbers and numerical problems were framed in words and sentences. Readers, mostly monks and scholars, had to form mental pictures of the problems at hand to solve, for example, quadratic equations. In the late nineteenth century, even economists still shunned equations and mathematical notation, preferring narrative and anecdotes. In his fascinating narrative, Mazur places the development of mathematical symbolism squarely within cultural evolution.

In the first part of *Enlightening Symbols*, 'Numerals', Mazur traces the depiction of numbers, beginning more than five millennia ago. First, Sumerian farmers calculated the areas of fields using cuneiform numbers; later, Egyptian priests used hieroglyphs for their calendars; and in medieval Europe, astronomers, merchants and officers of courts went about their daily business with abacuses. The chronicle leads from counting on ten fingers, through Egyptian, Hebrew, Greek, Chinese and Roman numerals, to the positional system that we use today, with units in the right-hand column, tens in the next column to the left, hundreds in the next, and so on.

A revolution was fomented in the seventh century AD by Hindu scholars, who invented zero both as a number to denote nothingness and as a placeholder for the positional system. Dissemination of the system in Europe began with translations into Arabic in the late eighth century, and into Latin in the early twelfth century. Mazur discusses many notable early mathematicians, including the Indian Brahmagupta in the seventh century, the Persian Muhammad ibn Musa Al-Khwarizmi in the ninth and the Italian Leonardo Pisano Bigollo — better known as Fibonacci — in the thirteenth.



Enlightening Symbols: A Short History of Mathematical Notation and Its Hidden Powers
JOSEPH MAZUR
Princeton University Press: 2014.

But numerals are only part of the story. There must also be signs to indicate negative and imaginary numbers, symbols to denote special constants, characters to represent unknowns and parameters, and pictograms for operators. Mazur's second section, 'Algebra', covers these. In AD 830, Al-Khwarizmi asked: "What must be the amount of two squares which, when summed up and added to ten times the root of one of them, make up a sum of forty eight dirhems?" Compare such verbosity to our concise $2x^2 + 10x = 48$, and then imagine the even more frightening impression left on readers by the Italian mathematician Scipio del Ferro's wordy prescription of how to solve cubic equations in 1505.

About 50 years later, the Welsh physician Robert Recorde introduced the equals sign, =, "to auoide the tediouse repetition of these woordes: is equalle to … because noe. 2. thynges, can be moare equalle" than a pair of parallels. Recorde was following a trend to symbolize algebra, first with word abbreviations and later with specialized symbols, that had been started by mathematicians in the Maghreb about two centuries earlier.

'The Power of Symbols', the book's third part, is the most challenging. In it, Mazur turns to the theory of mind. Does one need language to have mathematical thought? Drawing on material from luminaries including writer Fyodor Dostoyevsky, philosopher Ludwig Wittgenstein, cognitive neuroscientist Stanislas Dehaene and psychologist Daniel Kahneman, he wrestles with the question.

In literature, for example, meaning comes from associative experiences, and so it is at times with reading mathematics. Mazur argues that symbols, like poems, make connections between experience and the unknown to convey meaning. But when symbols acquire meaning, he warns, one may lose sight of the object represented, and continue with mechanical manipulations without grasping the essence of what is being done — as when one juggles equations that contain π without comprehending the connection to a circle's circumference.

This is a nuanced, intelligently framed chronicle packed with nuggets — such as the fact that Hindus, not Arabs, introduced Arabic numerals. In a word: enlightening. ■

George Szpiro *is a correspondent for the Swiss daily newspaper* Neue Zürcher Zeitung. *He wrote* Pricing the Future, *and is working on a history of mathematical economics and decision-making.*
e-mail: georgeszpiro@gmail.com

Correspondence

Swiss science set to stay international

Curbs on immigration resulting from last month's Swiss referendum (see *Nature* **506,** 265; 2014) have led to Switzerland's exclusion from the competition for European Research Council (ERC) grants. This is a devastating blow, given that the country has the highest share of international researchers on ERC grants. Thanks to the Swiss National Science Foundation (SNSF) stepping in to run a temporary parallel programme, however, Swiss science can remain international and competitive.

The SNSF has supported basic science through single-investigator grants for more than 60 years. The new temporary SNSF schemes will help researchers who are working in Switzerland or negotiating with Swiss institutions and who were planning to participate in the ERC competition. The schemes will have similar deadlines, procedures and success rates to the ERC's, and the evaluation panels will comprise distinguished scientists from Switzerland and abroad.

ERC grants represent a sort of 'Champions League' for researchers in Europe, and so it will not be easy to run a local programme of the same calibre. However, Switzerland will now be able to maintain its strong international track record: 49% of our professorial bodies are international, 33% of our researchers are abroad, and 28% of our students are foreigners.

Martin Vetterli *SNSF, Bern, Switzerland.*
martin.vetterli@snf.ch

Publish results from volunteer computing

In your discussion of waning participation in volunteer computing projects (*Nature* **506,** 16–17; 2014), you omit to mention the most important motivator of all — contribution to scientific progress.

For instance, a review of completed projects on IBM's World Community Grid for biomedical research uncovers a lacklustre track record in scientific output (see go.nature.com/fkheej). Four out of 16 of these projects have yet to produce any publications, even though the earliest project was completed more than 5 years ago. Others have generated only non-peer-reviewed output, such as book chapters or conference proceedings. Of the peer-reviewed publications, several focused solely on the technical aspect of distributed computing, with no analysis of the computed results.

To maintain the viability of volunteer computing projects, researchers should be reminded that public support and confidence are gained through professional and timely communication of results. Otherwise, online titles or badges amount to no more than virtual candy.

Herman Tse *University of Hong Kong, Hong Kong.*
htse@hku.hk

Thwart fatal infant gut parasite

Vaccination against cryptosporidiosis, a devastating disease for children in the developing world caused by a protozoan parasite, has met with limited success (P. Hotez *Health Affairs* **30,** 1080–1087; 2011). A better way to prevent this and other enteric diseases in such settings is to address shortcomings in hygiene.

When funding is in short supply, more investment in basic research is not the only solution (see B. Striepen *Nature* **503,** 189–191; 2013). Increasing efforts to extend interventions that are known to work already are more likely to bring immediate, tangible and cost-effective benefits.

The most significant gains will be achieved through education and the design, deployment and uptake of suitable sanitation and potable-water systems within the affected communities.

Brendon King, Paul Monis *Australian Water Quality Centre, Adelaide, Australia.*
brendon.king@sawater.com.au

Steering plans for neglected diseases

Mary Moran criticizes plans by the World Health Organization (WHO) for tackling neglected diseases in the developing world (*Nature* **506,** 267; 2014). Her arguments are misleading and trivialize global efforts to fix the research and development (R&D) system that is failing to address the health needs of people in these countries.

After 10 years' effort by WHO member states, the WHO Consultative Expert Working Group on R&D Financing and Coordination has proposed a framework to tackle these R&D shortcomings. This aims to promote and coordinate needs-driven innovation, sustainable funding mechanisms and patient access to health technologies.

Moran contends that the current drug pipeline for neglected diseases is "successful". In the past ten years, however, only 4% of all new drugs and vaccines and 1% of all new chemical entities were for neglected diseases — and none was suitable for the 17 neglected tropical diseases (B. Pedrique *et al. Lancet Glob. Health* **1,** e371–e379; 2013). Increased funding and better coordination will help, but what is really needed is a greater global incentive for needs-driven R&D.

To fulfil the working group's criteria, the WHO pilot projects need better coordination, optimization of research and accelerated delivery of health tools. This will be achieved only by strengthening capacity, testing innovative or pooled financing, developing an open-knowledge approach, and unlinking R&D costs from the end price of products.

Bernard Pécoul *Drugs for Neglected Diseases initiative, Geneva, Switzerland.*
bpecoul@dndi.org
Manica Balasegaram *Médecins Sans Frontières, Geneva, Switzerland.*

Research problems in Portugal run deep

As founders of the Portuguese National Association of Researchers in Science and Technology (ANICT), we believe that research problems in the country run deeper than just funding (see *Nature* **507,** 306; 2014).

Evaluation procedures for recruitment, career progression and funding allocation are in need of fundamental reform. They must become more transparent and be based solely on merit.

Selection for nationally funded PhD and postdoc scholarships, for example, is beset by conflicting interests and insufficient transparency (see www.anict.pt). The recent cuts in scholarship numbers have exacerbated these factors.

Currently, merit has little influence on career stability or advancement. Many researchers, even team leaders with students and ongoing grants, are being driven to leave the country because of their non-renewable, short-term contracts.

Instead of bickering and passing the buck, Portugal's politicians need to agree on long-term plans to foster and retain the country's talented young researchers.

Nuno Cerca *University of Minho, Braga, Portugal.*
nunocerca@ceb.uminho.pt
João Lopes *University of Lisbon, Portugal.*
Miguel Jorge *University of Strathclyde, Glasgow, UK.*

NEUROSCIENCE

Updating views of visual updating

Our brains create a stable view of the world even though our eyes dart around. A study of how the brain might compensate for eye movements reveals an unexpected twist in the vision-stabilizing mechanism. SEE LETTER P.504

JOHN A. ASSAD

Vision is the great illusionist. Although scenes before our eyes seem vivid and detailed, the part of the eye dedicated to high-acuity vision, the fovea, can cover only a narrow sliver of visual space — little more than the breadth of a thumbnail held at arm's length. Vision seems so detailed because we constantly move our eyes to scan the high-acuity fovea across a scene. Saccades — quick, jerky movements of the eyes — occur several times per second, filling in the perceptual fogginess of the visual periphery.

Given this perpetual motion of the eyes, how the visual world looks so stable is an enduring mystery. The retina responds whenever the visual scene 'slips' across its surface, and is thus unable to distinguish between the scene flashing in front of the eyes (for example, a speeding train at a railway crossing) or the eyes rushing across the scene. The brain, however, generates eye movements, and thus could perceptually compensate for saccades[1]. On page 504 of this issue, Zirnsak *et al.*[2] offer insight into how this compensation could occur.

In 1992, the neuroscientist Michael Goldberg and his colleagues suggested a potential neuronal mechanism for eye-movement compensation[3]. These researchers studied neurons in the macaque monkey's lateral parietal cortex, a part of the brain that serves as a bridge between vision and eye movements. Parietal neurons, like other visual neurons, have a receptive field (RF), a circumscribed part of the visual field for which visual stimuli activate electrical responses (Fig. 1). These RFs are defined relative to the fixation position (the point in visual space where the fovea is focused when eyes are stationary), and are usually considered spatially fixed relative to that position, reflecting the convergent hard-wired inputs to the neurons that ultimately stem from the retina. Thus, if the fixation position is displaced by a saccade, the RF should 'move' as well, in lock-step with the fixation position.

However, Goldberg and co-workers made the important discovery that parietal RFs are more fluid around the time of saccades: when monkeys made saccades to a point on a computer screen, even before the eyes moved, the neurons could be activated by a spot of light that fell at the future position of the RF relative to the upcoming fixation point after the saccade. The researchers thus suggested that the RF was "updated" or "remapped" to the new position, perhaps to compensate perceptually for the upcoming eye movement. Similar spatial updating was subsequently found in other brain structures that have mixed visual and oculomotor (eye-movement) function, such as the frontal eye fields[4] (FEFs) and the superior colliculus[5], suggesting a common compensatory mechanism.

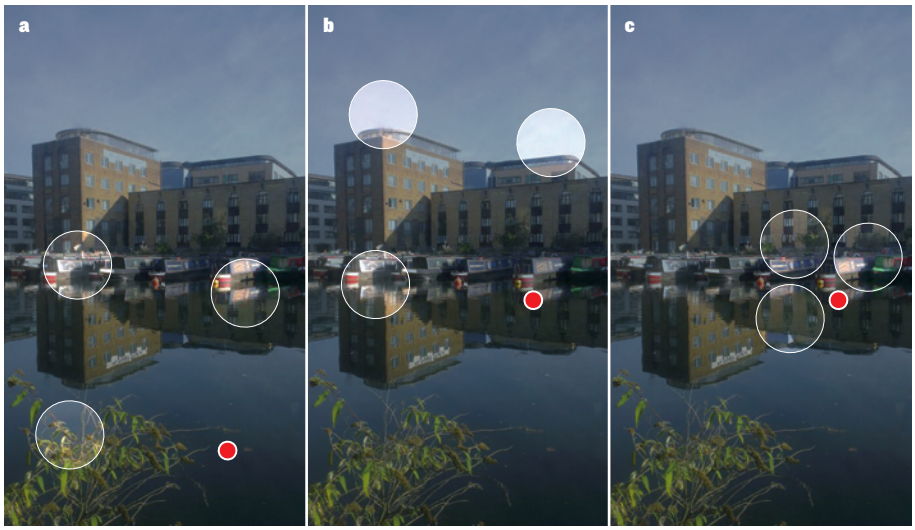


Figure 1 | Movement of receptive fields. **a**, Visual neurons have a receptive field (RF), a region of the visual field for which stimuli activate electrical responses. RFs are defined relative to the position fixated upon by the fovea, the tiny area of the eye that provides high-acuity vision. Here, the positions of three RFs are shown when the fovea is fixated at the red circle. **b**, After the fovea establishes fixation at a new point, the RFs move in lock-step with the new fixation position. Previous studies suggested that RFs shift to their new positions before the eyes move, a process known as predictive remapping. **c**, Zirnsak *et al.*[2] report that the RFs transiently shift to the upcoming fixation point itself, instead of to the future RF locations.

A strongly held notion about the spatial-updating hypothesis is that the brain's entire representation of the visual field rigidly translates in preparation for the abrupt movement of the eyes. However, in the original studies of spatial updating, visual stimuli were focused almost exclusively on the expected upcoming location of the RF. Zirnsak and colleagues have now widened the net in the simplest way imaginable: they presented visual stimuli to macaque monkeys not only at the upcoming (post-saccade) location of the RF of FEF neurons, but at various positions all over a computer screen, to map the entire RF.

Surprisingly, when the full RFs were revealed, the authors found that these fields did not rigidly translate before the eye movement, but instead transiently shifted en masse towards the location of the upcoming saccade target (Fig. 1c; see also Fig. 3a of the paper[2]). Some of the shifts of individual RFs were substantial, up to 18 degrees of visual angle. Moreover, in the few cases in which the original RF was located farther away from the original fixation point than the saccade target (as measured along the direction of the saccade), the shift in the RF to the saccade-target location was actually in the opposite direction to that expected for a rigid translation of the RF relative to the fovea.

The discovery that RFs collapse onto the saccade target is drastically different from the original spatial-updating hypothesis. So why did the authors of the original studies find neuronal responses at the upcoming position of the RF? In fact, the RFs of neurons in the FEF (and related visual–oculomotor structures) are quite large, so even if the RFs

collapsed to the saccade target, the fringes of many of those RFs would have probably overlapped with the upcoming RF location, and would thus respond to stimuli at that location. Indeed, in the original studies, responses at the upcoming RF location were often quite weak, which is consistent with the typically weaker responses at the periphery of RFs.

Zirnsak and colleagues' findings raise many issues. For one, RF shifts to a saccade target resemble the RF shifts that occur towards targets on which our attention is focused, even if the eyes do not move[6,7]. The two phenomena are probably related, inasmuch as shifts of attention to targets of interest precede eye movements. In this perspective, visual stability during saccades could result from the fact that we effectively ignore those parts of the visual scene that are away from the saccade target.

Moreover, spatial–perceptual distortions are known to occur owing to attention shifts and eye movements. These could be interesting phenomena to test further in monkeys, which can signal complex perceptions in experimental settings. More studies will also be needed to understand the cellular mechanisms of rapid RF modulation before saccades. But for experts and non-experts alike, Zirnsak *et al.* provide us with a valuable lesson: interesting things are often revealed when we search for our proverbial lost keys away from the streetlight. ■

John A. Assad *is at the Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto 38068, Italy, and the Department of Neurobiology, Harvard Medical School, Boston, Massachusetts.*
e-mail: jassad@hms.harvard.edu

1. von Helmholtz, H. *A Treatise on Physiological Optics* (transl. Southall, J. P. C.) (Dover, 1963).
2. Zirnsak, M., Steinmetz, N. A., Noudoost, B., Xu, K. Z. & Moore, T. *Nature* **507,** 504–507 (2014).
3. Duhamel, J. R., Colby, C. L. & Goldberg, M. E. *Science* **255,** 90–92 (1992).
4. Umeno, M. M. & Goldberg, M. E. *J. Neurophysiol.* **78,** 1373–1383 (1997).
5. Walker, M. F., Fitzgibbon, E. J. & Goldberg, M. E. *J. Neurophysiol.* **73,** 1988–2003 (1995).
6. Connor, C. E., Preddie, D. C., Gallant, J. L. & Van Essen, D. C. *J. Neurosci.* **17,** 3201–3214 (1997).
7. Womelsdorf, T., Anton-Erxleben, K., Pieper, F. & Treue, S. *Nature Neurosci.* **9,** 1156–1160 (2006).

SOLAR SYSTEM

Stranded in no-man's-land

The discovery of a second resident in a region of the Solar System called the inner Oort cloud prompts fresh thinking about this no-man's-land between the giant planets and the reservoir of comets of long orbital period. SEE LETTER P.471

MEGAN E. SCHWAMB

A decade after its discovery, Sedna[1] still remains one of the strangest objects in the Solar System. This remote icy body is on a highly eccentric orbit that extends to about 1,000 astronomical units (AU; 1 AU is the mean Earth–Sun distance) and has a perihelion (point of closest approach to the Sun) of 76 AU. Its orbit is well beyond the reach of Neptune, which is located at 30 AU, and is a long way from the edge of the Solar System where the Oort cloud, the reservoir of long-orbital-period comets, resides at about 10,000 AU. Although other potential candidates have been found[2–4], Sedna had remained the solitary confirmed member of a proposed inner Oort cloud beyond 70 AU. On page 471 of this issue, Trujillo and Sheppard report[5] the discovery of an object, called 2012 VP_{113}, which joins Sedna as the second confirmed member of the inner Oort cloud. The finding solidifies the existence of a population of icy bodies probably ranging in size from a few to a thousand kilometres.

To all intents and purposes, in the current architecture of the Solar System, Sedna and 2012 VP_{113} should not be there. These objects are in a no-man's-land between the giant planets and the Oort cloud where nothing in the known configuration of the modern-day Solar System could have emplaced them. Effectively frozen in place and untouched as the Solar System evolved to its present state, their orbits preserve the dynamical signature of whatever event scattered these bodies to such distances and detached them from the giant-planet region. With Sedna's discovery in 2003, several formation mechanisms were put forth, each predicting different orbital distributions of inner Oort cloud objects. But with only a single object, it was impossible to unambiguously distinguish between the different hypotheses.

Proposed mechanisms to explain the origin of the inner Oort cloud include the scattering of planetesimals (the building blocks of planet formation) by a distant planet beyond Neptune that may have been ejected from our Solar System, or by the passage of a single star, located at between 500 and 1,000 AU, early on in the Solar System's history[6,7]. The preferred mechanism is the gravitational scattering of leftover planetesimals from close encounters between the nascent Sun and other members of its birth star cluster, which would have contained between 10 and 1,000 stars[6,8–10]. Most stars are born in stellar nurseries that dissolve within a few million years. At the present age of the

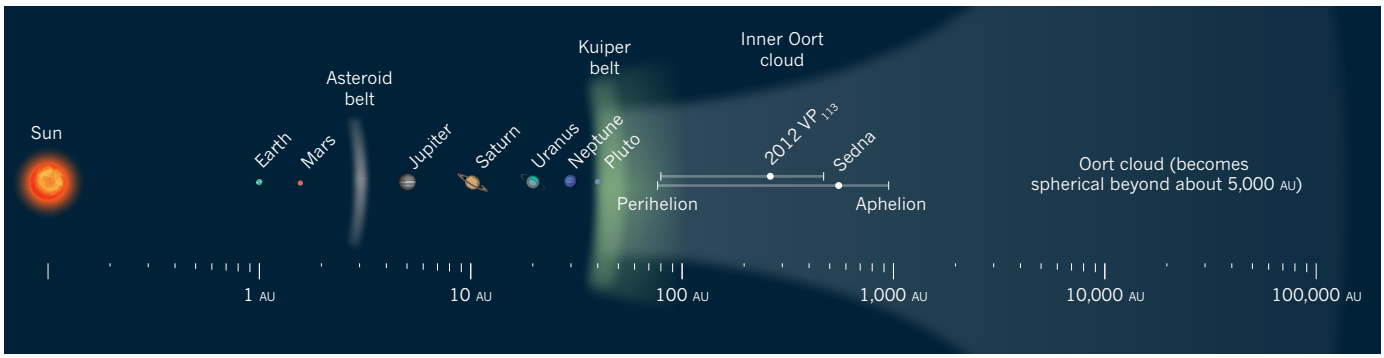


Figure 1 | The inner Oort cloud. Trujillo and Sheppard[5] have detected an object, called 2012 VP_{113}, that joins Sedna as the second confirmed body of the inner Oort cloud, a region believed to lie between the disk-shaped Kuiper belt of icy bodies and the spherical Oort cloud of long-orbital-period comets. Sedna is roughly 1,000 km across, whereas 2012 VP_{113} is estimated to be about 400 km in diameter. Perihelion is the orbital point of closest approach to the Sun; aphelion is the orbital point farthest from the Sun; 1 AU is the mean Earth–Sun distance. Objects in this diagram are not to scale. (Figure adapted from ref. 12.)

Solar System, the Sun and other stars of its birth cluster would be spread across the Galaxy. There is no direct way to identify whether the Sun was ever in a stellar nursery or to find its siblings. But indirectly, the mark from those early stellar encounters, like fingerprints at a crime scene, could provide evidence for the sculpted distribution of orbits in the inner Oort cloud.

Trujillo and Sheppard's detection of 2012 VP_{113}, which was made using the Dark Energy Camera (DECam) at the Cerro Tololo Inter-American Observatory 4-metre telescope in Chile, brings us one step closer to reading the dynamical record of the inner Oort cloud. With its farthest distance from the Sun at 452 AU and coming no closer into the Solar System than 80 AU, 2012 VP_{113} is placed well within the expected inner Oort cloud outside the Kuiper belt, a disk-shaped region of small icy bodies that lies beyond Neptune's orbit (Fig. 1).

It is still true that, with two objects, we cannot unambiguously identify the origin of the inner Oort cloud, but Trujillo and Sheppard find that the orbits of 2012 VP_{113} and Sedna are consistent with models for the Sun's birth cluster and constraints from previous ground-based, wide-field surveys[11]. This suggests that Sedna and 2012 VP_{113} are the tip of the iceberg for this population of distant inner Oort cloud objects. Most objects in the inner Oort cloud would reside at distances farther than Sedna and 2012 VP_{113}. For only an extremely small fraction of their orbits around the Sun would these inner Oort objects be bright enough to be detected in current ground-based surveys.

But there is more to the authors' study than described so far. Trujillo and Sheppard noticed a peculiarity with the orbits of Sedna and 2012 VP_{113}. The two objects have similar values for one of their orbital parameters: the angle between the point of perihelion and where the orbit crosses the plane of the Solar System. Interestingly, the most distant Kuiper-belt objects, with orbital semimajor axes greater than 150 AU and perihelia beyond Neptune, also seem to have values for such angles comparable to those of Sedna and 2012 VP_{113}. Such clustering of orbital angles seems to be unexplainable by the gravitational influence of Neptune alone. This result may be the first hint we have of an identifiable signature of the inner Oort cloud's formation mechanism on the orbits of closer-in Solar System bodies. If true, any formation mechanism proposed for the origin of Sedna and 2012 VP_{113} will need to explain this orbital structure. More detections of objects both in the Kuiper belt and the inner Oort cloud will be needed to confirm this result.

Our knowledge of the inner Oort cloud is, in many ways, in the same state as the study of the Kuiper belt was in the 1990s, when the first Kuiper-belt objects were discovered — 62 years after Pluto's detection. 2012 VP_{113} is the smoking gun for the existence of the inner Oort cloud. We now know that Sedna is not alone. The prospects of future detections using next-generation instruments and facilities, such as the Hyper Suprime-Cam on the 8.2-m Subaru telescope in Hawaii and the 8.4-m Large Synoptic Survey Telescope in Chile, are promising. With greater sky coverage and depth than that obtained by Trujillo and Sheppard, we will find more of these distant bodies lurking in the shadows and will begin to unravel the origin of the inner Oort cloud. ■

Megan E. Schwamb *is at the Institute of Astronomy and Astrophysics, Academia Sinica, Taipei 10617, Taiwan.*
e-mail: mschwamb@asiaa.sinica.edu.tw

1. Brown, M. E., Trujillo, C. & Rabinowitz, D. *Astrophys. J.* **617,** 645–649 (2004).
2. Chen, Y.-T. *et al. Astrophys. J.* **775,** L8 (2013).
3. Millis, R. L. *et al. Astron. J.* **123,** 2083–2109 (2002).
4. Gladman, B. *et al. Icarus* **157,** 269–279 (2002).
5. Trujillo, C. A & Sheppard, S. S. *Nature* **507,** 471–474 (2014).
6. Morbidelli, A. & Levison, H. F. *Astron. J.* **128,** 2564–2576 (2004).
7. Gomes, R. S., Gallardo, T., Fernández, J. A. & Brunini, A. *Celest. Mech. Dyn. Astron.* **91,** 109–129 (2005).
8. Brasser, R., Duncan, M. J., Levison, H. F., Schwamb, M. E. & Brown, M. *Icarus* **217,** 1–19 (2012).
9. Brasser, R., Duncan, M. J. & Levison, H. F. *Icarus* **184,** 59–82 (2006).
10. Kaib, N. A. & Quinn, T. *Icarus* **197,** 221–238 (2008).
11. Schwamb, M. E., Brown, M. E., Rabinowitz, D. L. & Ragozzine, D. *Astrophys. J.* **720,** 1691–1707 (2010).
12. Stern, S. A. *Nature* **424,** 639–642 (2003).

BIOGEOCHEMISTRY

Methane minimalism

A meta-analysis of methane emissions at the ecosystem level reveals a simple exponential dependence on temperature, despite the complex array of factors that control this process. SEE LETTER P.488

TORI M. HOEHLER & MARC J. ALPERIN

Methane is the third-largest contributor to the greenhouse effect, after water vapour and carbon dioxide. The atmospheric concentration of methane rose for much of the twentieth century, was steady between 1999 and 2006, and is now again rising, at a rate of 0.4% annually[1]. The cause of this resumption is not fully understood, but it is probably related to a surge in methane emissions from wetlands — nearly half of global methane emissions comes from wetlands and rice paddies that are expected to be subject to temperature-related and other feedbacks from global climate change. Although a complex set of factors influences methane emission at the ecosystem level, Yvon-Durocher *et al.*[2] report in this issue (page 488) that the average response of methane emissions to temperature across a range of ecosystems is well described by a simple mathematical relationship. Could it be as straightforward as that?

Most natural methane emissions originate with microorganisms called methanogens, the metabolic rates of which, during growth in culture with unlimited substrate, change with temperature according to the Arrhenius equation — a simple exponential dependence of rate constant on the reciprocal of absolute temperature. Such behaviour is unsurprising: although methanogenesis comprises a network of enzyme-catalysed reactions, the Arrhenius dependence reflects the kinetics of a single rate-limiting step. But Yvon-Durocher and colleagues' report of the same temperature dependence at the ecosystem level is notable, because an array of physical, chemical and ecological factors controls the production of methane and its release to the atmosphere[3].

Methanogenesis in soils and sediments is ultimately fuelled by complex organic matter; what proportion of this organic matter is converted to methane, and at what rate, depends as much on ecosystem dynamics as on the enzyme kinetics of methanogens (Fig. 1). For example, organic carbon can be converted to carbon dioxide, rather than methane, when oxidants such as oxygen, nitrate, iron(III) and sulphate are available to fuel methanogens' microbial competitors. And when organic matter does get converted to methane, it must first be broken down by other microbes into the few simple substrates that methanogens can metabolize — an upstream supply that can limit the rate of methanogenesis.

Transport of methane from the site of production to the atmosphere is also subject to effects that might obscure a purely biochemical temperature dependence. Methane is consumed in both aerobic and anaerobic microbial processes, and transport through zones in which such consumption occurs can markedly reduce net emissions. But consumption can largely be bypassed when physical and metabolic factors combine to promote ebullition — the formation and release of methane bubbles, a sight not uncommon in productive freshwater lakes and swamps. The vascular system of plants can also facilitate methane transport, when roots penetrate the methane-producing portions of soils and provide a direct conduit to the atmosphere; but these same conduits also transport oxygen that inhibits methanogenesis

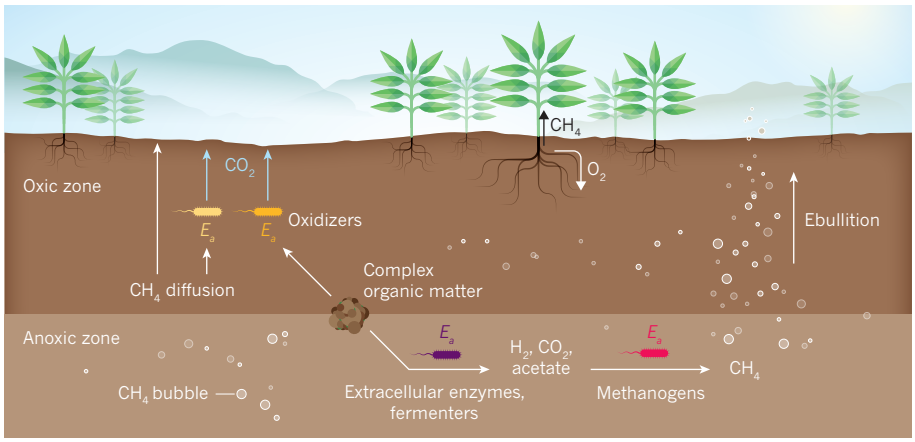


Figure 1 | Emissions network. Methane (CH_4) is generated by microorganisms (methanogens) that metabolize substrates produced through the breakdown of complex organic matter by extracellular enzymes and fermentative microorganisms. The rate of methane emission to the atmosphere is influenced by these microorganisms' individual sensitivities to temperature (indicated by E_a), and by chemical conditions, such as oxygen availability, that divert the flow of carbon to microbial competitors that oxidize organic matter to carbon dioxide. Methane emission also depends on gas transport by diffusion, by bubble ebullition and in the vasculature of plants, and on the fraction of methane that is consumed by methane-oxidizing microorganisms. Despite this complex array of factors, Yvon-Durocher *et al.*[2] report that the response of methane emissions at the ecosystem level can be described by the simple Arrhenius relationship.

and promotes methane consumption.

In their analysis of 127 studies of the ecosystem-level dependence of methane emission on temperature, Yvon-Durocher *et al.* acknowledge this complex array of factors, but conclude that the aggregate temperature response is nonetheless described by the Arrhenius equation, with an apparent activation energy (E_a) of 0.96 electronvolts, similar to the 1.10 eV observed in pure cultures of methanogens. E_a is a measure of temperature sensitivity; for example, 0.96 and 1.10 eV correspond, respectively, to a 3.5- and 4.2-fold increase in rate constant for an increase in temperature from 20 °C to 30 °C.

Statistically speaking, the large number of studies considered allows for a confident statement that the calculated mean E_a (0.96 eV) accurately reflects the mean temperature sensitivity of methane-emitting ecosystems — assuming that the sites that comprise the data set represent a random sample of all such environments. But the impact of factors other than temperature seems evident in the scatter and spread of the individual data sets considered by the authors. For example, about 40% of the studies considered had Arrhenius-plot correlation coefficients (r^2) of less than 0.5, which indicates that less than half of the variance in those emission data is explained by the Arrhenius relationship, and about 10% of the studies measured methane emissions that were higher at lower temperatures (opposite to the effect predicted by the Arrhenius equation).

The reported ecosystem-level E_a is higher than what has been called the "universal temperature dependence" of aerobic metabolism[4] — an E_a of 0.67 ± 0.15 eV that encompasses the metabolism of a wide range of plants, protozoa, invertebrates and vertebrates — and is also higher than the average E_a (0.72 eV) observed for a diverse group of 50 aerobic and anaerobic microorganisms[5]. The higher average E_a reported here for methanogenic ecosystems could reflect either that the biochemistry of methanogens (which have an average E_a of 1.10 eV) directly limits methane emissions in some ecosystems, or that the organisms that supply methanogens with substrates have similarly high temperature dependence. Nevertheless, the clear implication of these findings is that methane production will increase more steeply with temperature than would be captured by climate-change models that assume methane emission is governed by more typical (lower) values of E_a. For example, over the range of global warming projected[6] for this century (1.0–3.7 °C), an E_a of 0.96 eV suggests a 14–63% increase in methane emission compared with 10–40% for an E_a of 0.67 eV.

Yvon-Durocher and colleagues' findings constrain, and perhaps simplify, one piece of a much larger climate-change puzzle. Feedbacks from methane emissions in response to global climate change will ultimately derive from a combination of the direct temperature effects considered here and indirect effects such as thawing of permafrost (and the resultant availability of new organic matter), changes in vegetation, and large-scale inundation or drying of soils and wetlands. Moreover, methane's proportionate contribution to global warming (about 20% over the past century) may actually diminish as carbon dioxide takes an increasingly prominent role in the future[7]. But in this complex problem — unlike methane's flux to the atmosphere — every bit helps. ■

Tori M. Hoehler *is in the Space Science and Astrobiology Division, NASA Ames Research Center, Moffett Field, California 94035, USA.*
Marc J. Alperin *is in the Department of Marine Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3300, USA.*
e-mails: tori.m.hoehler@nasa.gov; alperin@email.unc.edu

1. Kirschke, S. *et al. Nature Geosci.* **6,** 813–823 (2013).
2. Yvon-Durocher, G. *et al. Nature* **507,** 488–491 (2014).
3. Riley, W. J. *et al. Biogeosciences* **8,** 1925–1953 (2011).
4. Gillooly, J. F. *et al. Science* **293,** 2248–2251 (2001).
5. Tijhuis, L. *et al. Biotech. Bioeng.* **42,** 509–519 (1993).
6. Stocker, T. F. *et al.* in *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Technical Summary, 33–115 (Cambridge Univ. Press, 2013).
7. Prather, M. *et al.* (eds) in *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Annex II, 1395–1445 (Cambridge Univ. Press, 2013).

This article was published online on 19 March 2014.

CLIMATE SCIENCE

A high bar for decadal forecasts of El Niño

Climate simulations suggest that multi-decadal periods of high and low variability in the phenomenon known as the El Niño–Southern Oscillation in the tropical Pacific Ocean may be entirely unpredictable.

PEDRO DINEZIO

The episodic warming and cooling of the surface temperature of the tropical Pacific Ocean, known as the El Niño–Southern Oscillation (ENSO), causes year-to-year climate fluctuations, affecting weather, ecosystems and economies around the world. The occurrence of these episodes is not regular. For example, whereas the period covering the years 1970–2000 witnessed the strongest El Niño (warming) events on record, the years since 2000 have experienced fewer and weaker such events. Writing in the *Journal of*

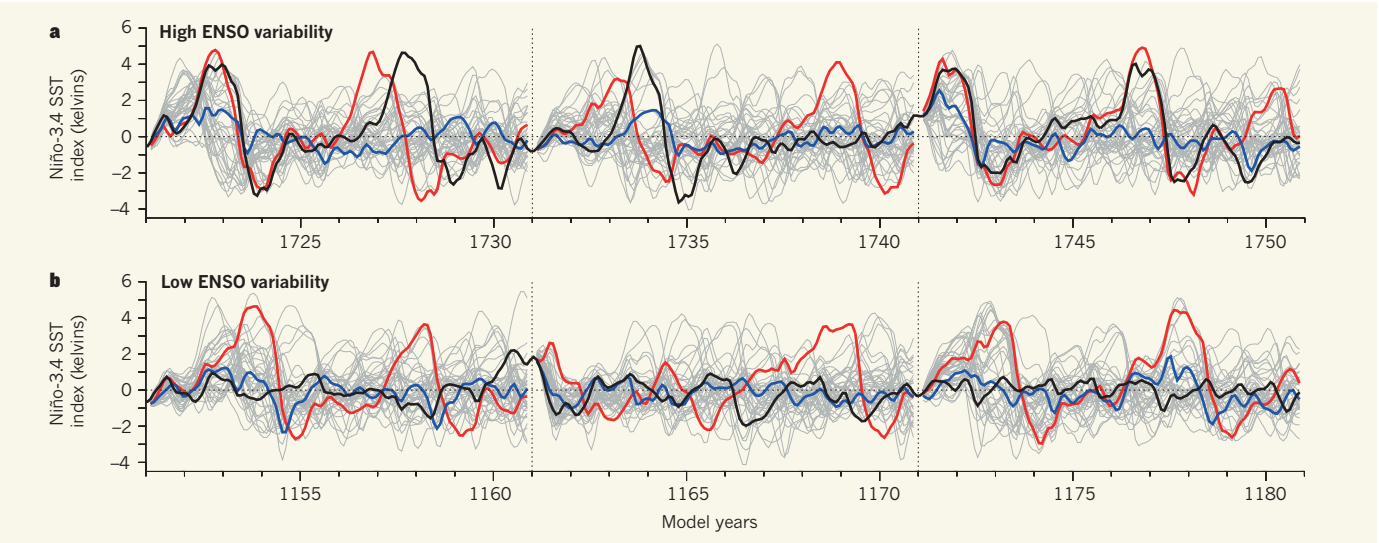


Figure 1 | Decadal forecasts of El Niño–Southern Oscillation activity. Using their GFDL-CM2.1 climate model, Wittenberg *et al.*[1] performed simulations of the variability of the El Niño–Southern Oscillation (ENSO), here quantified using the Niño-3.4 sea surface temperature (SST) index. Two distinct epochs are shown, characterized by high (**a**) or low (**b**) ENSO variability in the model's control run (black curves). For each epoch, 3 sets of 40 simulations are initialized from model years 1721, 1731 and 1741, and from 1151, 1161 and 1171, respectively (faint grey curves). Model years do not coincide with historical years. The simulations differ in only one tiny perturbation applied to the model's initial conditions. For both epochs, at least one simulation out of 40 exhibits high (red curves) or low (blue curves) ENSO variability, indicating that the level of activity is unpredictable. (Graphic courtesy of Andrew Wittenberg.)

Climate, Wittenberg *et al.*[1] make the case that these multi-decadal epochs of enhanced and subdued ENSO activity occur randomly and therefore may be unpredictable.

Changes in ENSO behaviour from decade to decade are commonly seen in historical observations and palaeoclimate proxy records[2–5]. These variations were first put into context by Wittenberg in an earlier study[6], which examined a 2,000-year simulation based on a fairly realistic climate model, known as GFDL-CM2.1. This concluded that decadal- to centennial-scale changes in ENSO behaviour can be internally generated by the model in the absence of any external forcing, such as increases in greenhouse-gas concentration or variations in solar output.

Predicting whether the coming decades will bring an onslaught of strong ENSO events — or none at all — is crucial because of the impact of such events on weather patterns around the world. Individual episodes may be predicted up to two years in advance[7], but on larger timescales our ability to forecast ENSO behaviour accurately may hinge on how ENSO responds to changes in the background climate system. This idea is supported by studies suggesting that the level of activity could be related to natural or man-made changes in the climate of the tropical Pacific[8,9].

In their latest study, Wittenberg *et al.*[1] used the same GFDL-CM2.1 model, this time to forecast epochs of high and low ENSO activity. For each epoch of activity in the model's control run, the authors performed 40 forecasts, each differing by a tiny perturbation of the size of the computer's rounding error to one of the model's numerical grid points. These 'perfect model' forecasts have the best chance of reproducing the extreme ENSO epochs seen in the control run, and permit assessment of the model's intrinsic ability to predict them.

The authors found that, beyond the first two to four years after initialization of the forecasts, the multi-decadal epochs of high and low ENSO activity are completely unpredictable. For each epoch, the model forecasts either active or quiet events with the same probability. That is, the perturbations can alter the forecasts in such a way that the model is capable of forecasting an inactive ENSO decade where it originally simulated a highly active one (Fig. 1a), or an active decade where it simulated a quiet one (Fig. 1b).

This is the 'butterfly effect' of chaos theory[10] applied to ENSO events. Seemingly small perturbations to a system, such as the flapping of a butterfly's wings, may lead to large changes in that system. It is a sobering finding, because it suggests that the changes observed in ENSO behaviour during the twentieth century could very well be random fluctuations unrelated to natural or man-made changes in the climate of the tropical Pacific.

Further research is needed to determine whether the study's conclusions can be extrapolated from the model world to the real world. During the past decade, climate models have progressed substantially in their ability to simulate ENSO events. Many models can now emulate the long-term modulation first seen in the GFDL-CM2.1 simulations, possibly owing to the inclusion of improved wind patterns[11]. But it is not known whether even the best climate models simulate the correct mix of the myriad processes that influence ENSO. One cause of uncertainty might be that the decadal fluctuations in the background climate, which are thought to be the source of ENSO predictability[12], are too weak in the models' simulations[13]. Conversely, models simulate activity that is much stronger than observed[6,14], so this too-strong ENSO might be oblivious to the too-weak changes in background climate, resulting in decreased predictability. The realism of the simulations must be improved if model-based conclusions are to be applied to the real world.

Existing observational records are not yet long enough for us to investigate whether, and how, ENSO responds to long-term climate fluctuations that could be sources of predictability. Progress on this front depends on maintaining and expanding our observational capability in the ocean, which relies on arrays of autonomous profiling floats and tropical moorings. In the meantime, results such as those of Wittenberg *et al.* are reminders of the challenges associated with forecasting ENSO changes. Future attempts to attribute the causes of individual events and their decadal variations now face a much higher bar. ■

Pedro DiNezio *is in the Department of Oceanography, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA.*
e-mail: pdn@hawaii.edu

1. Wittenberg, A. T., Rosati, A., Delworth, T. L., Vecchi, G. A. & Zeng, F. *J. Clim.* http://dx.doi.org/10.1175/JCLI-D-13-00577.1 (2014).
2. McGregor, S., Timmermann, A., England, M. H., Timm, O. E. & Wittenberg, A. T. *Clim. Past* **9,** 2269–2284 (2013).
3. Li, J. *et al. Nature Clim. Change* **1,** 114–118 (2011).
4. Emile-Geay, J., Cobb, K. M., Mann, M. E. &

Wittenberg, A. T. *J. Clim.* **26,** 2329–2352 (2013).
5. Cobb, K. M. *et al. Science* **339,** 67–70 (2013).
6. Wittenberg, A. T. *Geophys. Res. Lett.* **36,** L12702 (2009).
7. Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E. & Huang, D. *Nature* **428,** 733–736 (2004).
8. Wang, B. & An, S. *Clim. Dyn.* **18,** 475–486 (2002).
9. DiNezio, P. N. *et al. J. Clim.* **25,** 7399–7420 (2012).
10. Lorenz, E. N. presented at *139th Annu. Meet. AAAS*, Boston, Mass. (1972); http://eaps4.mit.edu/research/Lorenz/Butterfly_1972.pdf.
11. Neale, R. B., Richter, J. H. & Jochum, M. *J. Clim.* **21,** 5904–5924 (2008).
12. Kirtman, B. P. & Schopf, P. S. *J. Clim.* **11,** 2804–2822 (1998).
13. Ault, T. R., Deser, C., Newman, M. & Emile-Geay, J. *Geophys. Res. Lett.* **40,** 3450–3456 (2013).
14. Deser, C. *et al. J. Clim.* **25,** 2622–2651 (2012).

ALZHEIMER'S DISEASE

A protective factor for the ageing brain

What guards the aged brain against neurodegeneration? A study finds that the REST protein has a central role in protecting ageing neurons from death and in maintaining cognitive capacity in the elderly. **SEE ARTICLE P.448**

LI-HUEI TSAI & RAM MADABHUSHI

Alzheimer's disease is the leading cause of dementia. In the United States, an estimated 13% of people aged over 65, and nearly one-third of those aged over 85, are affected by this disorder[1]. As the population ages, the number of people with the disease is expected to rise precipitously, with no effective therapeutic strategy in sight. A central question is why certain individuals who display the anatomical and molecular features of Alzheimer's disease still preserve their cognitive abilities. On page 448 of this issue, Lu *et al.*[2] provide some answers.

The authors demonstrate that a protein called repressor element 1 silencing transcription factor (REST), which is normally expressed at low levels in the neurons of young human brains, is profoundly elevated in aged brains. By contrast, they observed that REST levels are markedly reduced in the nucleus of neurons in patients with mild cognitive impairment (a condition that often precedes dementia) and Alzheimer's disease. Lu and colleagues also analysed post-mortem samples of prefrontal cortex taken from people who had previously undergone neuropsychiatric assessments[3], and found that nuclear REST levels correlated positively with cognitive function. Taken together, these findings underscore the importance of maintaining REST protein levels in the aged brain for preventing cognitive decline and Alzheimer's disease.

Previous gene-expression studies[4–6] have extensively characterized the changes that occur in the ageing brains of healthy individuals and those with Alzheimer's. Generally, in both groups, neuronal genes implicated in synaptic transmission (signalling at neural junctions), in calcium signalling and in the functioning of inhibitory neurons show decreased expression, whereas genes associated with stress responses, DNA-damage responses, immune responses and apoptotic cell death show increased expression. But there are also crucial differences. For instance, toxic stimuli that are characteristic of Alzheimer's-related neurodegeneration induce the blockage of gene expression associated with neural plasticity (changes in neural pathways and junctions) in mouse models of Alzheimer's and in the brains of people with the disease[7].

In their study, Lu *et al.* used a dazzling array of experimental approaches to demonstrate that the loss of neuroprotective REST functions contributes to neuronal vulnerability in the brains of those with Alzheimer's. The authors observed that, in the brains of healthy aged individuals, nuclear REST both targets and suppresses several pro-apoptotic genes, as well as certain genes that encode enzymes involved in the pathology of Alzheimer's. But in diseased brains this suppression is lost, resulting in the induction of genes that are likely to underlie aspects of neuronal loss and neurodegeneration (Fig. 1).

The authors also report that REST-deficient mice show no neuronal loss at 1 month old, but they do by 8 months. Similarly, the authors found that *Caenorhabditis elegans* worms deficient in *spr-1*, *spr-3* and *spr-4* (genes that evolved from the same ancestral gene as *REST*) are more sensitive to oxidative stress and have shorter lifespans than wild-type worms. Therefore, REST-mediated transcriptional repression confers resistance to neurotoxic stress in several scenarios.

When the researchers treated cultured human neurons with hydrogen peroxide (an oxidizing agent), and then administered some of the resulting culture medium to untreated cells, they observed a clear increase in the expression of REST messenger RNA by those cells. This suggests that a soluble REST-inducing factor is produced by neurons upon oxidative stress. Furthermore, the authors demonstrated that extracts of aged, but not young, human prefrontal cortex robustly induce REST expression in neural cells *in vitro*.

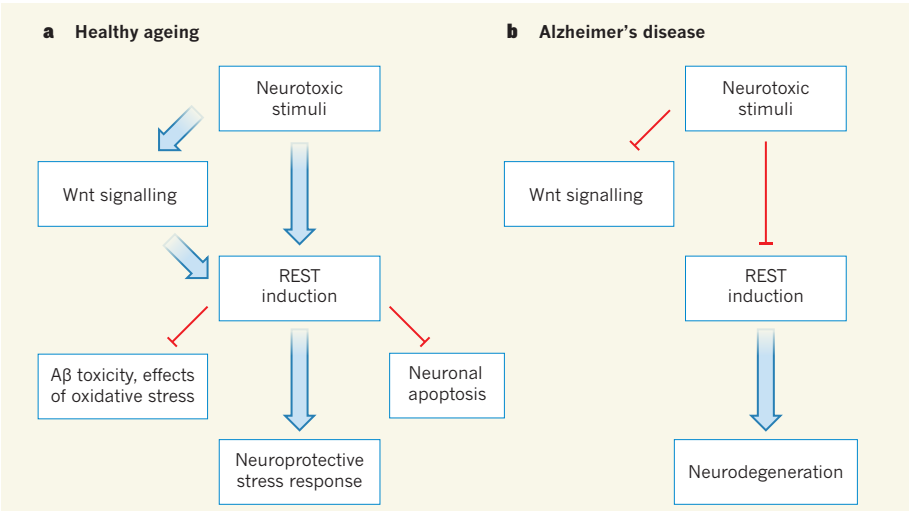


Figure 1 | REST activation distinguishes healthy aged brains from the brains of those with Alzheimer's disease. **a**, Lu *et al.*[2] report that various neurotoxic stimuli in the ageing brains of healthy people cause increased expression of the REST protein, an effect that is at least partially dependent on the Wnt-signalling pathway. REST induction activates neuroprotective stress-response pathways, while suppressing neuronal apoptotic cell death and the toxic effects of the accumulation of amyloid-β (Aβ, the main protein component of the plaques found in the brains of patients with Alzheimer's disease) and of elevated oxidative stress. **b**, In the brains of those with Alzheimer's disease, both Wnt signalling and REST induction are suppressed, leading to neurodegeneration.

Conversely, they found that extracts of cortex from those affected by Alzheimer's had reduced REST-inducing activity compared with age-matched controls.

REST is a target of the Wnt–β-catenin signalling pathway[8]. Lu *et al.* showed that REST can be induced by Wnt-3a and Wnt-7a (two proteins from the Wnt family), and that a Wnt-receptor antagonist partially inhibits REST induction. These results suggest that stress can increase Wnt signalling, which, in turn, induces REST expression. It remains to be determined whether Wnt is elevated in only some parts of the brain to influence REST — the authors report evidence of Wnt signalling in the hippocampus and the frontal cortex, but not in the cerebellum, for example. The question of whether the observed changes in Wnt signalling originate from neurons or from other cells in the brain also deserves further investigation.

Lu and colleagues went on to show that loss of nuclear REST is associated with a substantial increase in the expression of genes implicated in apoptosis and Alzheimer's pathology in both the least severe (Alzheimer's disease 1, AD1) and most severe (AD2) forms of the disease. Unexpectedly, however, they observed that another set of genes targeted by REST — those involved in synaptic transmission and other functions of neural junctions — initially show increased expression in AD1, but then reduced levels in AD2. It is plausible that the increased expression of these genes in AD1 is a compensatory mechanism to maintain neuronal homeostasis. Further analysis of gene expression in specific subtypes of neural cells should help to clarify precisely how the observed gene-expression patterns arise.

The authors' observations also raise the question of why gene-expression profiles differ in AD1 and AD2. Further characterization of the environment of the affected genes' chromatin (the DNA–protein complex in which genetic material is organized) and of the binding of other transcription factors in the genes' regulatory regions in the brains of the young, the aged, and those with AD1 or AD2, should provide greater insight into this matter.

Excitingly, this study provides the first detailed investigation of molecular markers in the brain that differentiate between populations of the young, the aged and those with Alzheimer's. Furthermore, by showing that ageing in the brain might be associated with the activation of a specific stress-response program, it implies that sustained maintenance of this program confers protection from neurodegeneration. Indeed, all the healthy centenarians studied in the research showed uniformly high levels of REST.

Could therapeutically stimulating nuclear REST activity in the brain prevent Alzheimer's and other degenerative diseases? On the basis of Lu and co-workers' results, one strategy would be to activate Wnt signalling in aged individuals. However, such activation is also implicated in the development of various cancers[9], and so this approach would probably require careful targeting of Wnt activation in the brain. Alternative strategies include finding either Wnt-independent REST activators or small molecules that prevent the export of REST from the nucleus. A deeper understanding of the molecular mechanisms that govern REST activation in the ageing brain will be crucial for such efforts to be successful. ■

Li-Huei Tsai *and* **Ram Madabhushi** *are at the Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.*
e-mail: lhtsai@mit.edu

1. Alzheimer's Association *Alzheimer's Dement.* **8,** 131–168 (2012).
2. Lu, T. *et. al. Nature* **507,** 448–454 (2014).
3. Bennett, D. A. *et al. Neurology* **66,** 1837–1844 (2006).
4. Lu, T. *et al. Nature* **429,** 883–891 (2004).
5. Loerch, P. M. *et al. PLoS ONE* **3,** e3329 (2008).
6. Cooper-Knock, J. *et al. Nature Rev. Neurol.* **8,** 518–530 (2012).
7. Graff, J. *et al. Nature* **483,** 222–226 (2012).
8. Willert, J., Epping, M., Pollack, J. R., Brown, P. O. & Nusse, R. *BMC Dev. Biol.* **2,** 8 (2002).
9. Anastas, J. N. & Moon, R. T. *Nature Rev. Cancer* **13,** 11–26 (2013).

This article was published online on 19 March 2014.

LOW-TEMPERATURE PHYSICS

Chaos in the cold

A marriage between theory and experiment has shown that ultracold erbium atoms trapped with laser light and subjected to a magnetic field undergo collisions that are characterized by quantum chaos. SEE LETTER P.475

PAUL S. JULIENNE

The use of magnetic fields to manipulate the interactions of ultracold atoms at temperatures below one microkelvin has allowed the experimental and theoretical study of a plethora of exotic phenomena in quantum physics[1]. The report by Frisch *et al.*[2] on page 475 of this issue adds an entirely new twist to this work. Until now, such research has mainly used simple atoms that have relatively simple interactions. Frisch and colleagues worked instead with a complex atomic species, erbium, and combined theory and experiment to demonstrate the signature of quantum chaos in the collisions between two erbium atoms. The results offer the prospect of exploring new avenues of ultracold physics and chemistry using complex atomic and molecular species.

In their study, Frisch *et al.* used a tightly focused laser to trap samples of around 100,000 erbium (Er) atoms in their lowest-energy quantum state at a temperature of about 400 nanokelvins. The samples consisted either of bosonic isotopes (^{168}Er or ^{166}Er), which have integer spin, or of a fermionic one (^{167}Er), which has half-integer spin. They tuned an applied magnetic field to a fixed value between 0 and 7 millitesla and counted the number of atoms remaining after holding them in the trap for 400 milliseconds. The authors found that the number of atoms left in the trap depended strongly on the particular value of the magnetic field selected. Such atom losses are well known in cold-atom physics, and are used to locate features known as Feshbach resonances, which occur when the total (nearly zero) energy of two colliding atoms matches the magnetically tunable energy of a bound diatomic molecular state (Er_2). Tabulating the magnetic-field values at which losses exist pinpoints the 'positions' of the resonances, from which the dynamics of the underlying collisions can be probed.

Frisch *et al.* found that the number of resonances in their system is surprisingly large, much larger than obtained with alkali-metal species such as lithium, rubidium or

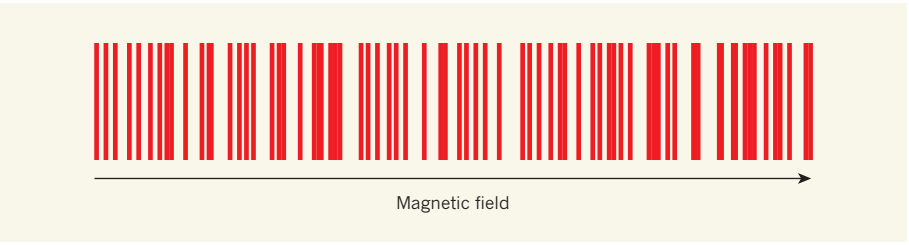


Figure 1 | A barcode of resonances. The lines mark the different values of an applied magnetic field at which erbium atoms from a trapped sample studied by Frisch *et al.*[2] show increased loss from the trap. The distribution of spacing between adjacent lines, which locate features known as Feshbach resonances, indicates that collisions between two atoms have quantum chaotic behaviour.

caesium (Cs). They provide theoretical models that connect this behaviour with chaotic quantum dynamics of the collisions. Because alkali-metal atoms such as Cs have a simple atomic structure (they have zero electronic orbital angular momentum), the resonances of colliding Cs atoms are spaced relatively far apart as a function of increasing magnetic field, and are characterized by simple quantum numbers. By contrast, an Er atom has a complex atomic structure that gives it a total electronic orbital angular momentum of five. Therefore, colliding Er atoms have many more resonances than their Cs analogues, and they are 'mixed' — that is, they are no longer described by simple quantum numbers.

To explain the nature of their measured resonance spectrum (Fig. 1), as well as that calculated by their model, Frisch *et al.* borrowed a tool known as random matrix theory from earlier work in nuclear and other fields of physics. By assuming that the various interactions of their system are described by the parameters of this theory, they could reproduce the average spacing between successive resonances. They found that the distribution of spacing for ^{168}Er or ^{166}Er bosons was much closer to one that corresponds to chaotic dynamics, known as a Wigner–Dyson distribution, than to the Poissonian distribution that characterizes regular, non-chaotic dynamics. The ^{167}Er fermions, which also have nuclear spin, showed an even denser set of resonances than the bosonic isotopes, but the authors have not yet analysed this set in detail because of its greater complexity. A study that was reported[3] this year measured similarly dense sets of resonances in systems of cold bosonic and fermionic isotopes of dysprosium (Dy).

Why are these results significant? Let us consider what is special about cold collisions. Because cold atoms move slowly, on average more than 10,000 times more slowly than atoms at room temperature, the Heisenberg uncertainty principle ensures that the uncertainty in their position is large. The colliding atoms no longer behave like particles, but take on wave-like character, with a wavelength that is inversely proportional to their velocity. These wavelengths can become large, one micrometre or more, much larger than the sub-nanometre length of a chemical bond. In this case, collisions become simple and are characterized by a parameter known as the scattering length.

Part of the power and beauty of cold-atom physics is that the scattering length can be made to take on any value by tuning a magnetic field close to a Feshbach resonance[1]. Its value controls the two-body, few-body and many-body physics of ultracold quantum matter. Thus, controlling the field makes the system dance to our tune. Previously, the Feshbach resonances used for such control have been isolated resonances that vary with the field and the atomic kinetic energy in a simple and well-understood way. Now, cold-atom researchers have to figure out how to understand the scattering length associated with a dense set of chaotic resonances such as that observed by Frisch and colleagues. The variation with field and kinetic energy will be more complex than previously encountered. The same is true of the atom-loss processes that are associated with such resonances and that determine the lifetime of a cold atomic system.

Another reason for working with Er is that its magnetic dipole moment is seven times larger than that of alkali-metal atoms. The dipole moment for Dy is even larger. Such large dipole moments, which result in a long-range interaction between pairs of dipoles, could enable the realization of some of the rich variety of phenomena predicted for an ensemble of cold dipoles[4,5]. It will now be necessary to understand the interplay between the dense set of resonances and the long-range dipolar interactions.

Finally, cold molecules — which have rotational, vibrational and other internal degrees of freedom — are expected to have much denser resonance spectra than Er or Dy atoms. Work is also being done to make systems of cold polar molecules, which can have much stronger dipolar interactions than atoms and exhibit a larger range of dipolar phenomena. Therefore, Frisch and colleagues' study is but a prelude to the interesting work that is to come on cold molecular systems, the resonances of which may have quite long lifetimes[6]. We can anticipate that the considerable body of work carried out on the internal relaxation and chemical reactions of excited molecular complexes[7,8] will come into play when chaos turns up at the heart of cold molecular collisions. ■

Paul S. Julienne *is at the Joint Quantum Institute, University of Maryland, College Park, Maryland 20742, USA.*
e-mail: psj@umd.edu

1. Chin, C., Grimm, R., Julienne, P. & Tiesinga, E. *Rev. Mod. Phys.* **82,** 1225–1286 (2010).
2. Frisch, A. *Nature* **507,** 475–479 (2014).
3. Baumann, K., Burdick, N. Q., Lu, M. & Lev, B. L. *Phys. Rev. A* **89,** 020701(R) (2014).
4. Baranov, M. *Phys. Rep.* **464,** 71–111 (2008).
5. Baranov, M. A., Dalmonte, M., Pupillo, G. & Zoller, P. *Chem. Rev.* **112,** 5012–5061 (2012).
6. Mayle, M., Ruzic, B. P. & Bohn, J. L. *Phys. Rev. A* **85,** 062712 (2012).
7. Jacobson, M. P. & Field, R. W. *J. Phys. Chem. A* **104,** 3073–3086 (2000).
8. Bowman, J. M. & Suits, A. G. *Phys. Today* **64** (11), 33–37 (2011).

This article was published online on 12 March 2014.

OSTEOARTHRITIS

The zinc link

Increased influx of zinc into chondrocytes — the cells that make up cartilage — has been found to activate matrix-degrading enzymes that cause the destruction of cartilage in osteoarthritis.

VIRGINIA BYERS KRAUS

The classic wound-healing response consists of a cascade of events that can be arbitrarily divided into four phases: blood clotting and bleeding cessation; inflammation; cellular proliferation; and wound remodelling with scar-tissue formation[1]. Osteoarthritis, the most prevalent and recalcitrant form of arthritis, has been proposed to be the result of a chronic wound-healing response in joints[2]. Articular cartilage, which is the tissue that lines joint surfaces, has no blood vessels, so wound healing at these sites begins with the inflammatory phase. This phase is characterized by an extensive auto-debridement response, in which dead or damaged tissue is removed. Writing in *Cell*, Kim *et al.*[3] provide compelling data that help to unravel the mechanistic details of this 'clean-up' phase, how it can progress to osteoarthritis, and the role of zinc in this process.

Second only to iron in trace-element abundance in the human body, and well known as a component of wound-healing salves, zinc has long been implicated in wound healing through promoting auto-debridement, anti-infective activity and the re-formation of skin[4]. Now, however, Kim *et al.* show that zinc has a deleterious effect on cartilage through its ability to upregulate enzymes responsible for cartilage degradation in a process that bears a striking resemblance to auto-debridement during wound healing.

The authors found that expression of the zinc-transporting protein ZIP8 increased *in vitro* in chondrocytes — the cells that make up cartilage — in response to IL-1β, an inflammatory cell-signalling molecule that is a key mediator of osteoarthritis. Overexpression of ZIP8 in chondrocytes led to increased levels of intracellular zinc and of zinc-dependent metalloprotease matrix-degrading enzymes that have been implicated in osteoarthritis, including MMP3, MMP9, MMP12, MMP13 and ADAMTS5. These metalloproteases are key mediators of the auto-debridement phase of wound healing: they regulate inflammation;

degrade the tissue matrix surrounding cells, thereby facilitating cell migration to the area; and remodel the new extracellular matrix[5].

Kim *et al.* also found higher ZIP8 and intracellular zinc levels in the cartilage of humans with osteoarthritis, and in that of mice that model this condition, than in normal cartilage. Increasing or decreasing zinc-transport activity through cartilage-specific up- or downregulation of ZIP8 in mice respectively increased or decreased the severity of the disease. One of the factors mediating the effect of zinc seems to be the metal-sensing transcription factor MTF1 — the authors observed that zinc enhanced the nuclear localization and transcriptional activity of this protein, and that its overexpression led to increased expression of matrix metalloproteases and to osteoarthritis. These findings are consistent with the recent demonstration[6] that ZIP8 levels influence the severity of osteoarthritis. Taken together with the knowledge that excessive metalloprotease activity contributes to the development of chronic skin wounds[5], these data support the theory that osteoarthritis is an internal wound-healing process that engenders chronic joint disease.

Because osteoarthritis is a complex disease that involves many interacting tissues, elucidating the temporal course of its development is challenging. Kim *et al.* show that modulation of ZIP8 is sufficient to give rise to typical osteoarthritis-related features in bone — a finding that supports the 'cartilage first' scenario in the long-standing debate regarding whether osteoarthritis develops first in cartilage or in bone[7]. Overexpression of ZIP8 also enhanced cartilage destruction in the absence of apparent inflammation of the synovial lining of the joint, further bolstering the idea that initial cartilage destruction triggers a cascade of disease-aggravating events in neighbouring joint tissues. This is consistent with the emerging model of the disease, in which metalloproteases, in addition to chewing away cartilage, generate fragments of cartilage molecules that are perceived by the immune system as danger signals[8], eliciting an inflammatory response that contributes to disease progression (Fig. 1).

As demonstrated in mice, there is a correlation between wound healing, cartilage regeneration and protection from osteoarthritis[9]. For example, mice of the 'superhealer' strain undergo faster cutaneous-wound healing, experience a less intense and shorter inflammatory response to joint injury, and develop less severe osteoarthritis than non-superhealer mice[10]. Therefore, a better understanding of the phases and mechanisms underlying the wound-healing response is likely to reveal potential avenues for the treatment of osteoarthritis.

How might the human wound-healing response be manipulated to resemble the superhealer phenomenon? Kim *et al.* speculate that local depletion of zinc, inhibition of ZIP8 function and zinc influx and inhibition

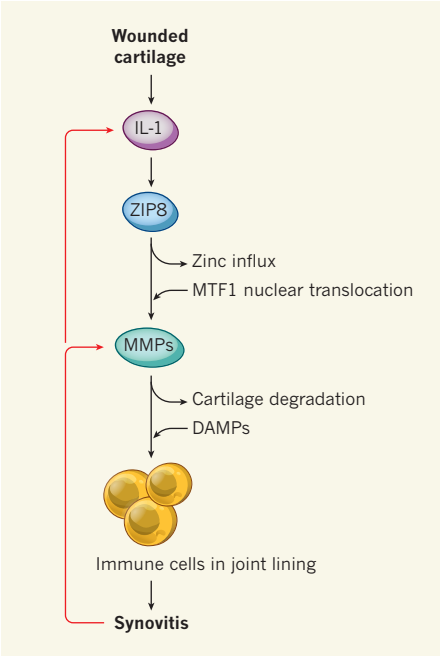


Figure 1 | The downward spiral of osteoarthritis. Osteoarthritis is thought to occur when the wound-healing response to a joint injury fails to occur satisfactorily. This response begins with the activation of cell-signalling molecules, such as members of the IL-1 family, in chondrocytes. Kim *et al.*[3] show that IL-1β induces expression of the protein ZIP8, triggering a series of events that lead to disease development. First, ZIP8 mediates an influx of zinc into the cells, translocating the metal-sensing transcription factor MTF1 to the nucleus and thereby enhancing the activity of zinc-dependent metalloprotease enzymes (MMPs) that induce cartilage breakdown. Fragments of cartilage resulting from this enzymatic activity, referred to as disease-associated molecular patterns (DAMPs), are then recognized as danger signals by immune cells[8], inducing inflammation of the joint lining, known as synovitis, that feeds back to further aggravate the response.

of MTF1 activity in cartilage might be effective therapeutic approaches. Indeed, inhibiting the target of zinc, the matrix metalloproteases, has long been considered a logical treatment strategy. Although early attempts at metalloprotease inhibition suffered from a lack of selectivity and were associated with musculoskeletal side effects[11], more-selective metalloprotease inhibitors are showing promising therapeutic potential[11], as are strategies for inhibiting the zinc–ZIP8–MTF1 axis[6].

In all such studies, it will be important to consider the broader effects of zinc on the body. Might zinc supplementation in humans worsen osteoarthritis? And is it feasible to lower zinc levels in a joint without affecting systemic zinc availability? For successful therapeutic targeting, we will also need to fill gaps in our understanding of the molecular basis of osteoarthritis. What are the initial events that upregulate ZIP proteins in cartilage? And how is IL-1β, the inducer of ZIP8, itself induced?

Mechanical injury of cartilage is a potent inducer of metalloproteases, but not of IL-1β[12], whereas skin wounding induces the local release of IL-1α, which increases intracellular IL-1β in chondrocytes[13]. Because IL-1α has been proposed as the key intermediary in the early stages of the disease, with IL-1β the dominant mediator in more advanced stages[14], it would be interesting to determine whether IL-1α can also activate the zinc–ZIP8–MTF1 axis. Finally, given that ZIP14, a close homologue of ZIP8 and a zinc and iron transporter, is implicated[15] in haemochromatosis (an iron-accumulation condition that is also a cause of osteoarthritis), additional study of metal transporters in general may yield further insight into the pathways to osteoarthritis.

A healed wound has been described[1] as "one that has been returned to a normal anatomical structure, function and appearance of the tissue within a reasonable period of time". But we do not know how often joints completely return to normal after injury, or if they ever do[16]. In humans, osteoarthritis is diagnosed only once it has progressed to revealing overt signs of disease, such as cartilage loss, and such cases presumably reflect joints that have undergone an unresolved wound-healing response. Perhaps our growing understanding of the molecular underpinnings of this disease will allow signs of osteoarthritis to be detected earlier in the wound-healing battle, so that this response can be turned off sooner or channelled into a regenerative response, thereby preventing disease development. ■

Virginia Byers Kraus *is at the Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, North Carolina 27701-2047, USA.*
e-mail: vbk@duke.edu

1. Velnar, T., Bailey, T. & Smrkolj, V. *J. Int. Med. Res.* **37,** 1528–1542 (2009).
2. Scanzello, C. R., Plaas, A. & Crow, M. K. *Curr. Opin. Rheumatol.* **20,** 565–572 (2008).
3. Kim, J.-H. *et al. Cell* **156,** 730–743 (2014).
4. Lansdown, A. B. G., Mirastschijski, U., Stubbs, N., Scanlon, E. & Ågren, M. S. *Wound Repair Regen.* **15,** 2–16 (2007).
5. Xue, M., Le, N. T. V. & Jackson, C. J. *Expert Opin. Ther. Targets* **10,** 143–155 (2006).
6. Song, J. *et al. J. Biomed. Sci.* **20,** 31 (2013).
7. Findlay, D. M. in *Principles of Osteoarthritis: Its Definition, Character, Derivation and Modality-Related Recognition* (ed. Rothschild, B. M.) Ch. 7, 139–154 (InTech, 2012).
8. Liu-Bryan, R. *Curr. Rheumatol. Rep.* **15,** 323 (2013).
9. Rai, M. F. *et al. Arthritis Rheum.* **64,** 2300–2310 (2012).
10. Lewis, J. S. Jr *et al. Arthritis Rheum.* **65,** 660–670 (2013).
11. Gege, C. *et al. J. Med. Chem.* **55,** 709–716 (2012).
12. Lee, J. H., Fitzgerald, J. B., DiMicco, M. A. & Grodzinsky, A. J. *Arthritis Rheum.* **52,** 2386–2395 (2005).
13. de Isla, N. G., Yang, J. W., Huselstein, C., Muller, S. & Stoltz, J. F. *Biorheology* **43,** 595–601 (2006).
14. van den Berg, W. B., Joosten, L. A. B. & van de Loo, F. A. J. *Clin. Exp. Rheumatol.* **17** (Suppl. 18), 0105–0114 (1999).
15. Gao, J., Zhao, N., Knutson, M. D. & Enns, C. A. *J. Biol. Chem.* **283,** 21462–21468 (2008).
16. Kraus, V. B. in *Rheumatology* (eds Hochberg, M. C., Silman, A. J., Smolen, J. S., Weinblatt, M. E. & Weisman, M. H.) Ch. 179 (Mosby, in the press).

PERSPECTIVE

doi:10.1038/nature13132

The ultimate physical limits of privacy

Artur Ekert[1,2] & Renato Renner[3]

Among those who make a living from the science of secrecy, worry and paranoia are just signs of professionalism. Can we protect our secrets against those who wield superior technological powers? Can we trust those who provide us with tools for protection? Can we even trust ourselves, our own freedom of choice? Recent developments in quantum cryptography show that some of these questions can be addressed and discussed in precise and operational terms, suggesting that privacy is indeed possible under surprisingly weak assumptions.

Edgar Allan Poe, an American writer and an amateur cryptographer, once wrote "… it may be roundly asserted that human ingenuity cannot concoct a cipher which human ingenuity cannot resolve …"[1]. Is it true? Are we doomed to be deprived of our privacy, no matter how hard we try to retain it? If the history of secret communication is of any guidance here, the answer is a resounding 'yes'. There is hardly a shortage of examples illustrating how the most brilliant efforts of code-makers were matched by the ingenuity of code-breakers[2]. Even today, the best that modern cryptography can offer are security reductions, telling us, for example, that breaking RSA, one of the most widely used public key cryptographic systems, is at least as hard as factoring large integers[3]. But is factoring really hard? Not with quantum technology. Indeed, RSA, and many other public key cryptosystems, will become insecure once a quantum computer is built[4]. Admittedly, that day is probably decades away, but can anyone prove, or give any reliable assurance, that it is? Confidence in the slowness of technological progress is all that the security of our best ciphers now rests on.

This said, the requirements for perfectly secure communication are well understood. When technical buzzwords are stripped away, all we need to construct a perfect cipher is shared private randomness, more precisely, a sequence of random bits known as a 'cryptographic key'. Any two parties who share the key, we call them Alice and Bob (not their real names, of course), can then use it to communicate secretly, using a simple encryption method known as the one-time pad[5]. The key is turned into a meaningful message by one party telling the other, in public, which bits of the key should be flipped. An eavesdropper, Eve, who has monitored the public communication and knows the general method of encryption but not the key will not be able to infer anything useful about the message. It is vital though that the key bits be truly random, never reused, and securely delivered to Alice and Bob, who may be miles apart. This is not easy, but it can be done, and one can only be amazed how well quantum physics lends itself to the task of key distribution.

Quantum key distribution, proposed independently by Bennett and Brassard[6] and by Ekert[7], derives its security either from the Heisenberg uncertainty principle (certain pairs of physical properties are complementary in the sense that knowing one property necessarily precludes knowledge about the other) or the monogamy of quantum entanglement (certain quantum correlations cannot be arbitrarily shared). At first, the idea of using quantum phenomena to improve secrecy was nothing more than an academic curiosity, but over time, with the progress of quantum technologies, it was embraced by experimental physicists and eventually turned into a viable commercial proposition. But even though quantum cryptography can offer the best security available at present, it is not immune to attacks exploiting botched implementations (see, for example, refs 8–11 for practical illustrations). The flaws in the design may be unintentional, the result of ignorance or negligence on the part of some honest individuals who design quantum cryptosystems; but they can also be malicious, secretly implanted by powerful adversaries. Should we not then dissect our cryptographic devices, analyse them and make sure that they do exactly what they are supposed to do? Given that some of the flaws may be unknown to us, what exactly should we be looking for? It has long been believed that here we reach the limits of privacy, and that at this point whoever is more technologically advanced, be it the NSA, GCHQ or some other agency, has the upper hand. Surprisingly, this is not the case.

Recent research shows that privacy is possible under stunningly weak assumptions. All we need are monogamous correlations and a little bit of 'free will', here defined as the ability to make choices that are independent of everything pre-existing and are hence unpredictable[12,13]. Given this, we can entertain seemingly implausible scenarios. For example, devices of unknown or dubious provenance, even those that are manufactured by our enemies, can be safely used to generate and distribute secure keys. There are caveats, of course: the devices must be placed in well-isolated locations to prevent any leaks of the registered data, and the data must be analysed by a trusted entity. Barring this, once the devices pass a certain statistical test they can be purchased without any knowledge of their internal working. This is a truly remarkable feat, also referred to as 'device-independent' cryptography[14–20]. Needless to say, proving security under such weak assumptions, with all the mathematical subtleties, is considerably more challenging than in the case of trusted devices, but the rapid progress in the past few years has been very encouraging, making device-independent cryptography one of the most active areas of quantum information science.

In fact, some of the device-independent schemes do not even rely on the validity of quantum theory[21–24], and they therefore guarantee security against adversaries who may have access to superior, 'post-quantum', technologies. The adversaries may even be given control over the choices made by Alice and Bob during the key distribution protocol[25]. As long as this control is not complete, Alice and Bob can do something about it. It turns out that 'free will' or, more specifically, the ability to make unpredictable, and, therefore, random, choices can be amplified[26]. Randomness amplification has recently triggered a flurry of research activity, culminating in a striking result: anything that is not completely deterministic can be made completely random[27,28]. This means, as we explain below, that as long as some of our choices are random and beyond control of the powers that be, we can keep our secrets secret.

The power of free choice

If there is one encryption method that comes close to a perfect cipher, it is the one-time pad. As we have already explained, its security critically

[1]Mathematical Institute, University of Oxford, Oxford OX2 9GG, UK. [2]Centre for Quantum Technologies, National University of Singapore, 117543 Singapore. [3]Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland.

relies on the randomness and secrecy of the cryptographic key. There is a snag, however, known as the 'key distribution problem'. Each key bit can be used only once, to encrypt one single message bit. To maintain their private communication, Alice and Bob must find a way to generate and distribute fresh key bits continuously. But how?

Let us put all the practicalities aside, just for a moment, and dream about something that would solve the key distribution problem. For example, imagine that Alice and Bob were given two magically linked coins, which always come out the same side up—either two heads or two tails—with equal probabilities. Alice and Bob can then toss such coins at their respective locations, writing '0' for heads and '1' for tails. The resulting binary strings will be random and identical, but will they be secret? Not necessarily. Technologically superior Eve could have manufactured an additional coin, magically linked to the coins held by Alice and Bob. The three coins always tally and Eve knows all the bits in the string.

Clearly, to achieve secrecy we must let Alice and Bob do something that is beyond Eve's control. For example, Alice and Bob may be given a choice between two different coins; Alice can toss either coin A_1 or coin A_2 and Bob, either B_1 or B_2. For each toss they must choose one of the two; tossing both A_1 and A_2 or both B_1 and B_2 is forbidden. Suppose, again, that the coins are magically linked; Alice and Bob's coins always come out the same, except when they toss A_1 and B_2, which always come out opposite. The magic can be succinctly summarized by the following four conditions[29,30] (Fig. 1):

$$A_1 = B_1,\ B_1 = A_2,\ A_2 = B_2,\ B_2 \neq A_1 \qquad (1)$$

These conditions are clearly contradictory; it is impossible to assign values to A_1, A_2, B_1 and B_2 so that all the four conditions are satisfied. But remember, Alice and Bob can toss only one coin each, and thus they can test only one of the four conditions in equation (1) at a time. Unperformed tosses do not have outcomes, and, hence, there is no contradiction here.

What if, say, Alice could break the rule and toss both of her coins, A_1 and A_2, in one go? It turns out that she would deprive Bob of his free choice. Suppose that Alice tossed first (correlations are not affected by the chronological order of the tosses) and that her outcomes are such that $A_1 = A_2$. Then Bob has no choice but to toss B_1, because this is the only choice compatible with the conditions in equation (1). Similarly, if $A_1 \neq A_2$, the only choice left to Bob is to toss B_2. This simple argument implies that the magic coins cannot be cloned. Having a clone, Z, of, say, A_1 (such that $Z = A_1$), and being able to toss it together with A_2 would lead to the same contradictions as tossing both A_1 and A_2. The existence of Z deprives Bob of his free choice. The conclusion is that if Alice and Bob have free choice then the magic correlations must be monogamous, that is, nothing else can be correlated to their coins. This turns the tables on Eve. Neither she nor anyone else can manufacture a coin that will always tally with any of the coins held by Alice or Bob. All ingredients for secure key distribution are now in place.

Key distribution

To establish a cryptographic key, Alice and Bob toss their magic coins. For each toss, Alice and Bob choose randomly, and independently of each other, which particular coin will be tossed: Alice is choosing between A_1 and A_2, and Bob, between B_1 and B_2. After the toss, they announce publicly the coins they selected, but not the outcomes they registered. The outcomes are secret, because the coins cannot be cloned, and identical, because the coins are magically linked (except when A_1 and B_2 are tossed, in which case either Bob or Alice must flip his or her bit). The net result is that Alice and Bob share one secret bit. To establish a longer key, they simply repeat this procedure as many times as required.

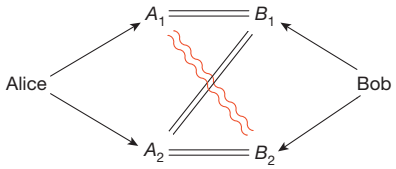


Figure 1 | Magic correlations. Alice and Bob choose and toss one coin each. Their choices are free, random and independent of each other, and the coins always come out the same way up, except when they toss A_1 and B_2, which always come out the opposite way up (represented by the red wiggly lines). Such correlations cannot be shared with a third party; for example, nobody can manufacture a coin that will always tally with any of the coins held by Alice or Bob.

We note that Alice and Bob do not need to make any assumptions about the provenance of the coins; as long as the coins comply with the conditions in equation (1), they are as good as it gets and could have been manufactured by anyone, adversaries included. But this compliance has to be checked. Alice and Bob can do it, for example, by revealing the outcomes of some randomly chosen tosses and checking if they agree with equation (1). Such publicly disclosed tosses are then discarded and the key is composed from the remaining tosses, outcomes of which have never been revealed in public. If Alice and Bob notice a deviation from the magic correlations, they abort the key distribution and try again with another set of coins.

Here we have tacitly assumed that Alice and Bob can communicate in public, but in such a way that nobody can alter their messages; for example, they might use a radio broadcast or an advert in a newspaper, or some other way that prevents impersonations. This communication is passively monitored by Eve and is the only information she gathers during the key distribution, because the coins are tossed in well-isolated locations that prevent any leaks of the registered outcomes. Given this, the secrecy of the key is based solely on the monogamy of the magic correlations and on one innocuous but essential assumption: both Alice and Bob can freely choose which coins to toss.

It seems that we have already achieved our goal. There is only one little problem with our, otherwise impeccable, solution of the key distribution problem, which is that the magic correlations do not exist. That is, we do not know of any physical process that can generate them. But all is not lost, because there are physically admissible correlations that are 'magical' enough for our purposes. Welcome to the quantum world!

The quantum of solace

Quantum theory is believed to govern all objects, large and small, but its consequences are most conspicuous in microscopic systems such as individual atoms or photons. Take, for example, polarized photons. Millions of identically polarized photons form the familiar polarized light, but at the quantum level polarization is an intrinsic property of each photon, corresponding to its spin. Although the polarization of a single photon can be measured along any direction, the outcome of the measurement has only two values, indicating whether the polarization is parallel or orthogonal to the measurement direction. For our purposes, we will label these outcomes 0 and 1.

A number of quantum optical techniques can be employed to generate pairs of polarization-entangled photons. Such photons respond to measurements, carried out on each of them separately, in a very coordinated manner. Suppose that Alice and Bob measure the polarizations of their respective photons along different directions, α and β. It turns out that, although the values 0 and 1 are equally likely to appear, Alice and Bob's outcomes tally with the probability

$$\cos^2(\alpha - \beta) \qquad (2)$$

This is just about everything you need to know about quantum physics for now.

Let us now replace the coin tosses by appropriately chosen polarization measurements: instead of tossing coin A_1, Alice simply measures her photon along $\alpha_1 = 0$; and instead of tossing A_2, she measures the photon along $\alpha_2 = 2\pi/8$. Similarly, Bob replaces his coin tosses B_1 and B_2 by measurements along directions $\beta_1 = \pi/8$ and $\beta_2 = 3\pi/8$, respectively. The resulting joint probabilities of all possible outcomes, obtained using equation (2) and the specified polarization angles, are shown in Table 1.

From a more general perspective, for any value of ε, which can be considered the probability of deviation from the magic correlations, the

Table 1 | Approximating magic correlations

	$A_1 = 0$	$A_1 = 1$	$A_2 = 0$	$A_2 = 1$
$B_1 = 0$	$\frac{1-\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{1-\varepsilon}{2}$	$\frac{\varepsilon}{2}$
$B_1 = 1$	$\frac{\varepsilon}{2}$	$\frac{1-\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{1-\varepsilon}{2}$
$B_2 = 0$	$\frac{\varepsilon}{2}$	$\frac{1-\varepsilon}{2}$	$\frac{1-\varepsilon}{2}$	$\frac{\varepsilon}{2}$
$B_2 = 1$	$\frac{1-\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{1-\varepsilon}{2}$

Joint probabilities of binary outcomes given the choices of A_i and B_j ($i, j = 1, 2$). The parameter ε takes the value 0 for the magic correlations (see equation (1)). The lowest physically admissible value, $\varepsilon = \sin^2(\pi/8) \approx 0.146$, can be obtained by measuring polarizations of appropriately entangled photons at some specific angles, for example 0, $\pi/8$, $2\pi/8$ and $3\pi/8$, corresponding to A_1, B_1, A_2 and B_2, respectively.

table describes 'non-signalling' correlations: Alice, by choosing between A_1 and A_2, cannot communicate any information to Bob, and vice versa Bob choosing between B_1 and B_2 cannot send any information to Alice. Neither of them can see through the statistics of the outcomes what the other one is doing. Correlations with $\varepsilon \geq 1/4$ are called 'classical', because they admit pre-assigned values of A_1, A_2, B_1 and B_2. This is no longer the case when $\varepsilon < 1/4$, because any pre-assignment is bound to violate at least one of the four conditions in equation (1). Surprisingly, as we have just seen, there are physically admissible correlations for which ε can reach $\sin^2(\pi/8) \approx 0.146$, which is the lowest value that can be achieved with quantum correlations[31]. Even though perfect magic correlations, with $\varepsilon = 0$, do not exist, there is still some magic left in quantum correlations, and it can be exploited.

Less reality, more security

The impossibility of assigning numerical values to certain physical quantities, for example the different polarizations of a photon, has been baffling physicists for almost a century[32]. After all, most of us grew up holding it self-evident that there is an objective reality in which physical objects have properties that can be quantified and whose values exist regardless of whether we measure them or not. Shocking as it may be, our world is not of this kind. Statistical inequalities, such as $\varepsilon \geq 1/4$, derived on the assumption that the values of unmeasured physical quantities do exist and commonly referred to as Bell's inequalities[33], have been violated in a number of painstaking experiments[34–44]. We shall not dwell on the philosophical implications of this experimental fact (volumes have been written on the subject), but simply point out that it should be embraced by all those who worry about secrecy because what does not exist cannot be eavesdropped, and so it is much easier to keep secrets in a non-classical world.

Indeed, given the correlations parameterized by ε, it can be shown that the probability of Eve guessing correctly any particular outcome cannot exceed $(1 + 4\varepsilon)/2$ (Box 1). Eve may know something about the outcomes (which is not good) but Alice and Bob, after running a statistical test and estimating ε, know how much she may know (which is good). If ε is low enough, this allows them to distil an almost perfect key from the outcomes, using a technique known as 'privacy amplification'.[45,46] The basic idea behind privacy amplification is quite simple. Imagine that you have two bits and that you know your adversary knows at most one of them, but that you do not know which one. Add the two bits together (modulo 2); the resulting bit will be secret. Needless to say, given more bits, there are more sophisticated ways of achieving secrecy, to mention only two-universal hash functions[47] or Trevisan's extractor[48].

In summary, whenever Alice and Bob are given any devices that generate correlated outcomes, they can run the key distribution protocol supplemented by a statistical 'honesty test' to estimate ε. If this value is small enough, say $\varepsilon = 0.15$, the end result, after privacy amplification, is a perfect cryptographic key. We obtain trusted privacy from untrusted devices, but what constitutes a device? We need to sort out one more thing before we can celebrate the arrival of the ultimate cipher. Should Alice and Bob trust the ultimate measuring and controlling devices; that is, should they trust themselves?

BOX 1

Eavesdropping quantified

Suppose that Eve wants to manufacture a device that outputs binary values, Z, designed to tally with, say, A_1. Regardless of her technological prowess, Eve has limited chances to succeed. For any two outcomes, A_i and B_j, the probabilities that they are equal to Z, that is, $\Pr(Z = A_i)$ and $\Pr(Z = B_j)$, cannot differ by more than $\Pr(A_i \neq B_j)$. This implies a sequence of inequalities:

$$\Pr(Z = A_1) - \Pr(Z = B_1) \leq \Pr(A_1 \neq B_1)$$
$$\Pr(Z = B_1) - \Pr(Z = A_2) \leq \Pr(B_1 \neq A_2)$$
$$\Pr(Z = A_2) - \Pr(Z = B_2) \leq \Pr(A_2 \neq B_2)$$
$$\Pr(Z = B_2) - \Pr(Z \neq A_1) \leq \Pr(B_2 = A_1)$$

Adding these inequalities together and taking into account that $\Pr(Z \neq A_1) = 1 - \Pr(Z = A_1)$ gives

$$\Pr(Z = A_1) \leq \frac{1}{2}(1 + I_2)$$

where the quantity $I_2 = \Pr(A_1 \neq B_1) + \Pr(B_1 \neq A_2) + \Pr(A_2 \neq B_2) + \Pr(B_2 = A_1)$ is the sum of the probabilities that any of the conditions in equation (1) is violated. The derivation presented here works for any A_i and B_j, and, it is worth stressing, does not involve quantum theory.

Although the values A_1, A_2, B_1 and B_2 do not coexist, all the probabilities used here involve only pairs of values, A_i and B_j, which can be measured simultaneously. They can be determined from the statistics of the experimental data. For the polarization measurements described in the text, we would obtain $I_2 = 4\varepsilon$ where $\varepsilon = \sin^2(\pi/8) \approx 0.146$. The bound thus asserts that $\Pr(Z = A_1) \leq 0.793$; that is, Eve's value, Z, will deviate from A_1 in more than 20% of the cases.

The notion of magic correlations can be extended to cases where Alice and Bob choose between $n \geq 2$ different measurements[67,68], with the conditions in equation (1) replaced by

$$A_1 = B_1,\ B_1 = A_2,\ \ldots,\ A_n = B_n,\ B_n \neq A_1 \qquad (3)$$

To approximate such correlations, Alice and Bob may use entangled photons and measure polarizations A_i and B_j, specified by angles α_i and β_j. These angles are chosen to be even and odd multiples of $\pi/4n$, respectively, so that the adjacent values of α_i and β_j are $\pi/4n$ radians apart. Then, according to equation (2), each of the conditions in equation (3) is satisfied, except with an error probability of $\varepsilon_n = \sin^2(\pi/4n) < 1/n^2$. It can then be shown, by the same arguments as for the $n = 2$ case, that any attempt by Eve to compute a prediction, Z, for the outcome of, say, A_1, can succeed with probability at most $(1 + I_n)/2$, where $I_n = \Pr(A_1 \neq B_1) + \Pr(B_1 \neq A_2) + \cdots + \Pr(A_n \neq B_n) + \Pr(B_n = A_1)$. For any classical correlations, $I_n \geq 1$. In contrast, quantum theory admits correlations such that $I_n = 2n\varepsilon_n < 2/n$. Consequently, in the limit of large n, the probability of Eve guessing the value of A_1 correctly becomes 1/2; that is, A_1 is uniformly random and independent of any information held by Eve. This observation is not only relevant for key distribution[21], but has been crucial for randomness amplification[26].

Should we trust ourselves?

We can hardly get more paranoid than that. Can we make free choices or are we held to the ransom of a greater force? In other words, what if we are manipulated?

We have already stressed the power of free choice. Decisions such as which coin to toss and which polarization to measure must be made freely (randomly) and independently. If referring to the experimenter's 'free will' sounds too esoteric, then think about the random number generators that in practical implementations make such choices. Where is

their randomness coming from? What if these random number generators are of dubious provenance, possibly manufactured by the same person who offered the key distribution kit? It is evident that without randomness there is no privacy: if everything is pre-determined, and all possible choices we make (with the help of tweaked random number generators or otherwise) are predictable or pre-programmed by our adversaries, then there is nothing that we can build our privacy on. Or is there?

There is if the manipulation is not complete and there is a little bit of freedom left. If someone we trust tells us that such and such a fraction of the choices made by our random number generators cannot be determined by the adversary, then privacy is still possible because local randomness can be amplified[26]. Randomness amplification can itself be done with device-independent protocols, and it works even if the fraction of initial randomness is arbitrarily small or the devices are noisy[27,28].

It all looks bizarre and too good to be true. Perfect privacy, secure against powerful adversaries who provide us with cryptographic tools and who may even manipulate us? Is such a thing possible? Yes, it is, but 'the devil is in the detail' and we need to look into some practicalities.

Practicalities

Quantum key distribution, in which security is tested by the degree of violation of Bell's inequalities, was proposed some time ago[7] and was followed shortly by a proof-of-principle experiment at what used to be called the Defence Research Agency (now Qinetiq) in Malvern, UK[49]. However, the device-independent character of this protocol has not been recognized until recently[15]. Moreover, proving the security of such a scheme in the presence of noise has not been easy. It has taken over a decade to agree on a useful definition of secrecy, even for trusted devices, and to conclude a long sequence of steadily improved security results[50–53] that eventually took into account all the quantum resources that Eve can muster[54]. Dealing with untrusted devices is even more tricky and keeps many of our colleagues busy[55–57].

Although all security proofs infer secrecy from the monogamy of the correlations, a major challenge is to make these arguments quantitative and robust to noise and imperfections, and applicable to keys of finite size[58,59]. There are other issues as well. For example, here we have taken for granted that Alice and Bob can estimate the parameter ε from a sufficiently large sample of their registered data. In the quantum domain, a statement of that kind requires a quantum version of what is known in classical statistics as de Finetti's theorem[54,60]. It guarantees that, for instance, pairs of photons can be treated as individual objects with individual properties and without any hidden correlations to other pairs. These, and many other results, addressed a number of subtleties and, finally, twenty years after its inception, the original entanglement-based key distribution protocol[7] has been shown to offer security even if the devices are not fully trusted and are exposed to noise[15–20]. This is assuming that quantum theory is all that there is, and that Eve is bound by the laws of quantum physics. However, if Alice and Bob are paranoid enough to give Eve some 'post-quantum' powers (technologies more powerful than quantum technologies which may rely on as-yet undiscovered physical phenomena that are not described by quantum physics), they can still resort to less efficient protocols that do not rely on quantum theory[21–24]. We should stress, however, that device-independent protocols and their security proofs have not yet reached the level of sophistication that is now common for the device-dependent scenario. In particular, more work is needed to improve the efficiency of the key distribution protocols or to identify conditions under which untrusted devices may be reused in multiple rounds of such protocols.

Given that violation of Bell's inequality is an experimental fact, what is it that prevents us from running the experiments that violated Bell's inequality again, but this time under the label of the device-independent key distribution? Convincing as they are, these experiments still leave some loopholes. For example, it is in principle possible that the photons detected in the experiments did not represent a fair sample of all photons emitted by the source (the 'detection loophole') or that the various parts and components of the experiment were causally connected (the 'locality loophole'). Some of these concerns were addressed in more recent experiments[43,44], but, a single experiment that closes all the loopholes at once, demonstrating the ultimate violation of Bell's inequality, is still lacking.

This is not so disturbing for physicists, because nature would have to be very malicious if it were to cheat us selectively—on locality in some experiments and in exploring detection loopholes in some other. In contrast, there is nothing to prevent an eavesdropper being malicious. In this adversarial setting, a proper experimental demonstration of device-independent cryptography requires a proper violation of Bell's inequalities. This is particularly true for the detection loophole. Imagine, for example, that Eve pre-programmed the devices assuming in advance a sequence of settings that Alice and Bob may choose for their measurements. Whenever her guess is correct, the devices will respond with pre-programmed results, and when it is not, one of the devices will simulate failure to respond. If Alice and Bob naively discard all the instances in which at least one of the devices failed to deliver a result, then they can be easily fooled by Eve. Thus, we do need the loophole-free violation of Bell's inequalities.

Closing the detection loop-hole is very challenging, because almost any optical component adds losses and imperfections to the key distribution set-up, but it is within the reach of today's technology, especially with the rapid progress in photodetection techniques. If distance is not an issue, then we can achieve near-perfect detection efficiency using entangled ions rather than photons[40], and this has been used to generate the first device-independent certified randomness[61,62]. Short of full device independence, we can also entertain intermediate scenarios, where some parts of the devices are trusted and some are not. Indeed, proposals that address issues such as untrusted detectors[63,64] offer significant improvements over the existing quantum key distribution schemes[65,66] and move secure communication in interesting new directions.

Experimental device-independent cryptography is far from easy, but technological progress so far has encouraged optimism. The days we stop worrying about untrustworthy or incompetent providers of cryptographic services may be not that far away.

Conclusion

Over the past decade or so, quantum cryptography has come of age, but the field is still an amazingly fertile source of inspiration for fundamental research. The search for the ultimate physical limits of privacy is still very much a work in progress, but we know that privacy is possible under surprisingly weak assumptions. Monogamous correlations, of whatever origin, and an arbitrarily small amount of free will are sufficient to conceal whatever we like. Free will is our most valuable asset. Come to think about it, without free will, there is no point in concealing anything anyway.

Received 12 August 2013; accepted 7 February 2014.

1. Poe, E. A. A few words on secret writing. *Graham's Mag.* **19,** 33–38 (1841).
2. Kahn, D. *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet* (Scribner, 1996).
3. Rivest, R., Shamir, A. & Adleman, L. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM* **21,** 120–126 (1978).
4. Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26,** 1484–1509 (1997).
5. Vernam, G. S. Cipher printing telegraph systems for secret wire and radio telegraphic communications. *J. Am. Inst. Electr. Eng.* **45,** 109–115 (1926).
6. Bennett, C. H. & Brassard, G. in *Proc. IEEE Int. Conf. Computer Syst. Signal Process.* 175–179 (IEEE, 1984).
   This work reported key distribution based on encoding information in complementary bases.
7. Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67,** 661–663 (1991).
   This work reported key distribution based on quantum entanglement.
8. Fung, C.-H. F., Qi, B., Tamaki, K. & Lo, H.-K. Phase-remapping attack in practical quantum-key-distribution systems. *Phys. Rev. A* **75,** 032314 (2007).
9. Lydersen, L. *et al.* Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature Photon.* **4,** 686–689 (2010).
10. Weier, H. *et al.* Quantum eavesdropping without interception: an attack exploiting the dead time of single-photon detectors. *New J. Phys.* **13,** 073024 (2011).
11. Gerhardt, I. *et al.* Full-field implementation of a perfect eavesdropper on a quantum cryptography system. *Nature Commun.* **2,** 349 (2011).
12. Bell, J. S. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy* (Cambridge Univ. Press, 2004).
13. Colbeck, R. & Renner, R. No extension of quantum theory can have improved predictive power. *Nature Commun.* **2,** 411 (2011).

14. Mayers, D. & Yao, A. in *FOCS '98: Proc. 39th Annu. Symp. Foundations Computer Sci.* 503–509 (IEEE, 1998).
This work included a proposal for self-testing cryptographic devices.
15. Acín, A. *et al.* Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **98,** 230501 (2007).
This work included a proposal for device-independent quantum key distribution.
16. McKague, M. *Quantum Information Processing With Adversarial Devices*. PhD thesis, Univ. Waterloo (2010).
17. Hänggi, E. & Renner, R. Device-independent quantum key distribution with commuting measurements. Preprint at http://arxiv.org/abs/1009.1833 (2010).
18. Masanes, L., Pironio, S. & Acin, A. Secure device-independent quantum key distribution with causally independent measurement devices. *Nature Commun.* **2,** 238 (2011).
19. Vazirani, U. & Vidick, T. in *ITCS '14: Proc. 2014 Conf. Innovations Theor. Computer Sci.* (ed, Naor, M.) 35–36 (ACM, 2012).
20. Reichardt, B. W., Unger, F. & Vazirani, U. Classical command of quantum systems. *Nature* **496,** 456–460 (2013).
21. Barrett, J., Hardy, L. & Kent, A. No signaling and quantum key distribution. *Phys. Rev. Lett.* **95,** 010503 (2005).
This work demonstrated that the security of entanglement-based key distribution can be guaranteed without relying on the correctness of quantum theory.
22. Acín, A., Gisin, N. & Masanes, L. From Bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.* **97,** 120405 (2006).
23. Masanes, L. Universally composable privacy amplification from causality constraints. *Phys. Rev. Lett.* **102,** 140501 (2009).
24. Hänggi, E., Renner, R. & Wolf, S. in *Advances in Cryptology – EUROCRYPT 2010* Vol. 6110 (ed. Gilbert, H.) 216–234 (Springer, 2010).
25. Koh, D. E. *et al.* Effects of reduced measurement independence on Bell-based randomness expansion. *Phys. Rev. Lett.* **109,** 160404 (2012).
26. Colbeck, R. & Renner, R. Free randomness can be amplified. *Nature Phys.* **8,** 450–454 (2012).
This work proved that randomness amplification is possible.
27. Gallego, R. *et al.* Full randomness from arbitrarily deterministic events. *Nature Commun.* **4,** 2654 (2013).
28. Brandao, F. *et al.* Robust device-independent randomness amplification with few devices. Preprint at http://arxiv.org/abs/1310.4544 (2013).
29. Khalfi, L. A. & Tsirelson, B. S. in *Symp. Foundations Mod. Phys.* (eds Lahti, P. & Mittelstaedt, P.) 441–460 (World Scientific, 1985).
30. Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys.* **24,** 379–385 (1994).
31. Cirel'son, B. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **4,** 93–100 (1980).
32. Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47,** 777–780 (1935).
33. Bell, J. S. On the Einstein-Podolsky-Rosen paradox. *Physics* **1,** 195–200 (1964).
34. Freedman, S. J. & Clauser, J. F. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.* **28,** 938–941 (1972).
35. Aspect, A., Grangier, P. & Roger, G. Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.* **47,** 460–463 (1981).
36. Aspect, A., Grangier, P. & Roger, G. Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: a new violation of Bell's inequalities. *Phys. Rev. Lett.* **49,** 91–94 (1982).
37. Aspect, A., Dalibard, J. & Roger, G. Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49,** 1804–1807 (1982).
38. Weihs, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **81,** 5039–5043 (1998).
39. Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. Violation of Bell inequalities by photons more than 10 km apart. *Phys. Rev. Lett.* **81,** 3563–3566 (1998).
40. Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409,** 791–794 (2001).
41. Pomarico, E., Bancal, J.-D., Sanguinetti, B., Rochdi, A. & Gisin, N. Various quantum nonlocality tests with a commercial two-photon entanglement source. *Phys. Rev. A* **83,** 052104 (2011).
42. Stuart, T. E., Slater, J. A., Colbeck, R., Renner, R. & Tittel, W. Experimental bound on the maximum predictive power of physical theories. *Phys. Rev. Lett.* **109,** 020402 (2012).
43. Giustina, M. *et al.* Bell violation using entangled photons without the fair-sampling assumption. *Nature* **497,** 227–230 (2013).
44. Christensen, B. G. *et al.* Detection-loophole-free test of quantum nonlocality, and applications. *Phys. Rev. Lett.* **111,** 130406 (2013).
45. Bennett, C. H., Brassard, G., Crepeau, C. & Maurer, U. M. Generalized privacy amplification. *IEEE Trans. Inf. Theory* **41,** 1915–1923 (1995).
46. Renner, R. & König, R. in *Second Theory Cryptogr. Conf.* (ed. Kilian, J.) 407–425 (Lect. Notes Computer Sci. 3378, Springer, 2005).
47. Tomamichel, M., Schaffner, C., Smith, A. & Renner, R. Leftover hashing against quantum side information. *IEEE Trans. Inf. Theory* **57,** 5524–5535 (2011).
48. De, A., Portmann, C., Vidick, T. & Renner, R. Trevisan's extractor in the presence of quantum side information. *SIAM J. Comput.* **41,** 915–940 (2012).
49. Ekert, A. K., Rarity, J. G., Tapster, P. R. & Palma, G. M. Practical quantum cryptography based on two-photon interferometry. *Phys. Rev. Lett.* **69,** 1293–1295 (1992).
50. Biham, E. & Mor, T. Security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **78,** 2256–2259 (1997).
51. Lütkenhaus, N. Security against individual attacks for realistic quantum key distribution. *Phys. Rev. A* **61,** 052304 (2000).
52. Shor, P. & Preskill, J. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.* **85,** 441–444 (2000).
53. Mayers, D. Unconditional security in quantum cryptography. *J. ACM* **48,** 351–406 (2001).
54. Renner, R., *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich (2005).
This work included a comprehensive security analysis of quantum key distribution with trusted devices and composable security.
55. Arnon-Friedman, R. & Ta-Shma, A. Limits of privacy amplification against non-signalling memory attacks. *Phys. Rev. A* **86,** 062333 (2012).
56. Hänggi, E., Renner, R. & Wolf, S. The impossibility of non-signaling privacy amplification. *Theor. Comput. Sci.* **486,** 27–42 (2013).
57. Barrett, J., Colbeck, R. & Kent, A. Memory attacks on device-independent quantum cryptography. *Phys. Rev. Lett.* **110,** 010503 (2013).
58. Scarani, V. & Renner, R. Quantum cryptography with finite resources: unconditional security bound for discrete-variable protocols with one-way postprocessing. *Phys. Rev. Lett.* **100,** 200501 (2008).
59. Scarani, V. *et al.* The security of practical quantum key distribution. *Rev. Mod. Phys.* **81,** 1301–1350 (2009).
60. Renner, R. Symmetry of large physical systems implies independence of subsystems. *Nature Phys.* **3,** 645–649 (2007).
61. Colbeck, R. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, Univ. Cambridge (2006).
62. Pironio, S. *et al.* Random numbers certified by Bell's theorem. *Nature* **464,** 1021–1024 (2010).
63. Lim, C. C. W., Portmann, C., Tomamichel, M., Renner, R. & Gisin, N. Device-independent quantum key distribution with local Bell test. *Phys. Rev. X* **3,** 031006 (2013).
64. Lo, H.-K., Curty, M. & Qi, B. Measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **108,** 130503 (2012).
65. Rubenok, A., Slater, J. A., Chan, P., Lucio-Martinez, I. & Tittel, W. Real-world two-photon interference and proof-of-principle quantum key distribution immune to detector attacks. *Phys. Rev. Lett.* **111,** 130501 (2013).
66. Liu, Y. *et al.* Experimental measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **111,** 130502 (2013).
67. Pearle, P. Hidden-variable example based upon data rejection. *Phys. Rev. D* **2,** 1418–1425 (1970).
68. Braunstein, S. L. & Caves, C. M. Wringing out better Bell inequalities. *Ann. Phys.* **202,** 22–56 (1990).

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 should be addressed to A.E. (artur.ekert@quibit.org) or R.R. (renner@phys.ethz.ch).

ARTICLE

doi:10.1038/nature13163

REST and stress resistance in ageing and Alzheimer's disease

Tao Lu[1], Liviu Aron[1], Joseph Zullo[1], Ying Pan[1], Haeyoung Kim[1], Yiwen Chen[2], Tun-Hsiang Yang[1], Hyun-Min Kim[1], Derek Drake[1], X. Shirley Liu[2], David A. Bennett[3], Monica P. Colaiácovo[1] & Bruce A. Yankner[1]

Human neurons are functional over an entire lifetime, yet the mechanisms that preserve function and protect against neurodegeneration during ageing are unknown. Here we show that induction of the repressor element 1-silencing transcription factor (REST; also known as neuron-restrictive silencer factor, NRSF) is a universal feature of normal ageing in human cortical and hippocampal neurons. REST is lost, however, in mild cognitive impairment and Alzheimer's disease. Chromatin immunoprecipitation with deep sequencing and expression analysis show that REST represses genes that promote cell death and Alzheimer's disease pathology, and induces the expression of stress response genes. Moreover, REST potently protects neurons from oxidative stress and amyloid β-protein toxicity, and conditional deletion of REST in the mouse brain leads to age-related neurodegeneration. A functional orthologue of REST, *Caenorhabditis elegans* SPR-4, also protects against oxidative stress and amyloid β-protein toxicity. During normal ageing, REST is induced in part by cell non-autonomous Wnt signalling. However, in Alzheimer's disease, frontotemporal dementia and dementia with Lewy bodies, REST is lost from the nucleus and appears in autophagosomes together with pathological misfolded proteins. Finally, REST levels during ageing are closely correlated with cognitive preservation and longevity. Thus, the activation state of REST may distinguish neuroprotection from neurodegeneration in the ageing brain.

The preservation of cognitive function during ageing has emerged as one of the major medical challenges of the 21st century. A fundamental question is why some individuals age with their cognitive function relatively intact whereas others decline and develop Alzheimer's disease (AD). Early studies suggested that neuronal loss was an integral feature of the ageing brain. With the advent of stereological neuronal quantification, however, it became clear that neuronal cell number is largely preserved in the neocortex and hippocampus of the ageing human brain, declining only in the setting of neurodegenerative disease[1–3]. Robust stress-response mechanisms must have evolved, therefore, to preserve neurons and cognitive function across an entire lifespan[4,5].

REST is a repressor of neuronal genes during embryonic development that is downregulated once terminal neuronal differentiation has occurred[6–8]. Here we show that REST is induced in the ageing human brain and regulates a network of genes that mediate cell death, stress resistance and AD pathology. This gene network becomes dysregulated at early stages of AD when REST is lost from the nucleus. Conditional REST knockout mice and *C. elegans* models suggest that REST protects neurons from age-related toxic insults. In ageing humans, elevated REST levels are associated with preservation of cognitive function and increased longevity, even in the presence of AD pathology. Hence, REST regulates a neuroprotective stress response that may be central to cognitive preservation during ageing.

REST is induced in the ageing brain and declines in AD

Transcriptional profiling has demonstrated significant changes in the expression of neuronal genes in the prefrontal cortex of ageing humans[9,10]. Analysis of this data set using the Ingenuity Systems IPA platform indicates that the transcription factor most strongly predicted to be activated in the ageing brain is REST ($P = 9 \times 10^{-10}$). Moreover, the 21-base-pair canonical RE1 recognition motif for REST is highly enriched in the age-downregulated gene set ($P = 3 \times 10^{-7}$) (Fig. 1a).

To explore the role of REST in the ageing brain, we measured REST levels in extracts of prefrontal cortex (PFC) from young adult (20–35 years) and aged (73–106 years) individuals without AD. REST expression was significantly increased in the ageing human PFC at both the messenger RNA and protein levels, as determined by quantitative PCR with reverse transcription (qRT–PCR) and western blotting (Fig. 1b, c). Full-length REST was markedly increased; the truncated splice variant REST4 was a minor component, comprising 0.1–0.5% of REST mRNA. Immunofluorescence microscopy using three different antibodies against the amino- or carboxy-terminal domains of REST showed a striking induction of REST in the nucleus of ageing neurons in the PFC and hippocampus (Fig. 1d, e and Extended Data Fig. 1). Much lower levels of REST were detected in microglial cells and astrocytes (data not shown). REST antibody specificity was indicated by ablation of immunoreactivity after antibody preabsorption with a REST blocking peptide, absence of immunoreactivity with matched nonspecific IgG, and loss of immunoreactivity after shRNA-mediated REST knockdown in neural SH-SY5Y cells (Extended Data Fig. 1b, c).

We then investigated whether induction of REST in ageing neurons leads to increased REST–RE1 site binding. To assess REST targeting specifically in neurons, we isolated neuronal nuclei from the PFC by fluorescence-activated cell sorting (FACS) of NeuN-positive nuclei[5] (Methods). Chromatin immunoprecipitation (ChIP)-PCR analysis showed a marked induction of REST binding to canonical RE1 motifs in REST target genes in the aged PFC (Fig. 1f). These results indicate that REST expression and function is increased in ageing neurons.

We next examined REST in ageing individuals with mild cognitive impairment (MCI) or AD. REST was almost absent from the nucleus of cortical and hippocampal neurons in AD (Fig. 1d and Extended Data Fig. 1a, d). Punctate cytoplasmic REST immunoreactivity was detected in some AD neurons, but was not detectable in other neurons (Fig. 1d). Loss of nuclear REST was confirmed by quantitative immunocytochemical

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Biostatistics and Computational Biology, Dana-Faber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02115, USA. [3]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois 60612, USA.

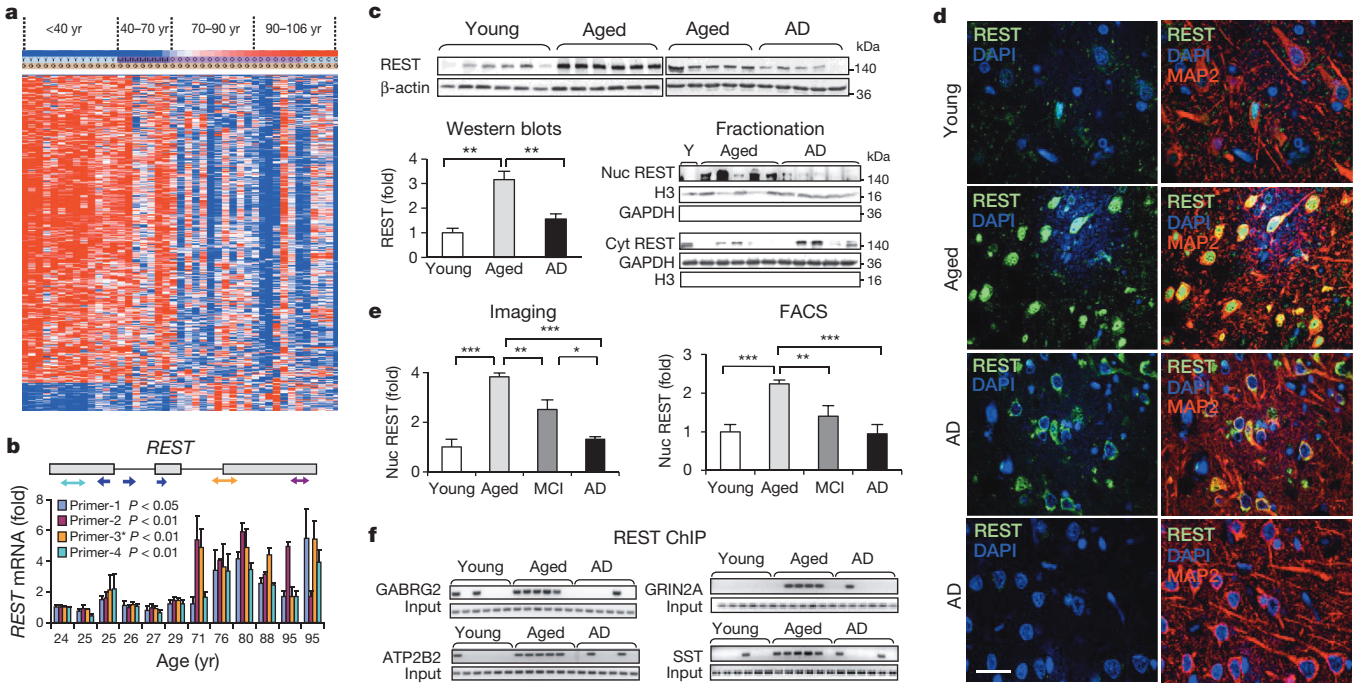


Figure 1 | Induction of REST in the ageing human prefrontal cortex. **a**, Hierarchical cluster analysis of predicted REST targets, based on the presence of the canonical RE1 motif[29], shows relatively high expression in young adults (red) and lower expression (blue) in the ageing population. Each lane represents an individual prefrontal cortical brain sample. **b**, qRT–PCR shows age-dependent induction of REST mRNA in the PFC. Values represent the mean ± s.d., $n = 3$. Shown are P values indicating significance for the mean of aged (71–95 years) versus young (24–29 years) cases for each primer. **c**, Age-dependent increase in total REST protein level in the PFC of controls but not AD patients. REST levels were also measured in isolated nuclear (Nuc REST; lower right panel, upper blot) and cytoplasmic (Cyt REST; lower right panel, lower blot) fractions. Each lane represents an individual case. Young, $n = 12$; Aged, $n = 15$; AD, $n = 10$. **d**, Confocal immunofluorescence labelling for REST (green), the neuronal marker MAP2 (red) and DNA (DAPI, blue) in the PFC of young adult, aged and AD cases. Scale bar, 25 μm. **e**, Quantitative analysis of nuclear REST levels by *in situ* imaging (left panel: Young, $n = 11$; Aged, $n = 77$; AD, $n = 72$; MCI = 11) or FACS analysis of isolated PFC neuronal nuclei (right panel: Young, $n = 11$; Aged, $n = 22$; AD, $n = 11$; MCI $n = 12$). For **c** and **e**, values are expressed as fold change relative to the young adult group, and represent the mean ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by Student's unpaired t-test. **f**, ChIP analysis shows induction of REST–RE1 site binding in normal ageing PFC neurons but not in AD.

analysis of 77 cases of sporadic AD compared with 72 age-matched cognitively intact controls, and confirmed by FACS analysis of neuronal nuclei in a subset of cases (Fig. 1e). In 23 out of 24 cases of amnestic MCI, nuclear REST levels were reduced relative to the mean of aged controls (Fig. 1e). A survey of brain regions showed significantly reduced nuclear REST levels in affected neuronal populations in AD, including prefrontal cortical neurons, and neurons of the CA1, CA3 and CA4 subfields of the hippocampus (Extended Data Fig. 1a, d). In contrast, nuclear REST was not reduced in neurons of the dentate gyrus and cerebellum, which are relatively unaffected in AD (Extended Data Fig. 1a, d). Reduced REST levels in AD were confirmed by western blotting of prefrontal cortical grey matter samples, as well as isolated cortical nuclei (Fig. 1c, upper and far right panels, respectively). Furthermore, ChIP-PCR analysis of neuronal nuclei showed significantly reduced or absent REST–RE1 site binding in AD relative to normal ageing controls (Figs 1f and 2a). Thus, REST is induced during normal brain ageing, but is markedly reduced in AD in vulnerable neuronal populations.

REST represses genes involved in cell death and AD

To obtain greater insight into the functional consequences of REST induction in the ageing brain, we explored REST targets by chromatin immunoprecipitation and deep sequencing (ChIP-seq) in the neural cell line SH-SY5Y. REST binding to a subset of the identified target genes was confirmed by ChIP-qPCR (Extended Data Fig. 2). REST ChIP-seq sites were significantly enriched among age-regulated genes in human PFC ($P = 1.65 \times 10^{-12}$), and were predominantly downregulated from the third to the tenth decades of life (Extended Data Fig. 2a). Pathway analysis of the ChIP-seq database showed a highly significant enrichment for genes in cell death pathways and genes that are associated with AD/dementia (Extended Data Fig. 2b, c). Cell death genes include p38 MAP kinase (*MAPK11*), *FAS*, *FADD*, *TRADD*, *BAX*, *BID*, *DAXX*, *PUMA* (also known as *BBC3*), the mitochondrial permeability transition pore ANT1 (also known as *SLC25A4*) and cytochrome C. REST targets related to the pathology of AD include the genes coding for gamma secretase complex members presenilin 2 and presenilin enhancer 2 implicated in amyloid β-protein (Aβ) generation, and CDK5R1 (also known as p35), CDK5R2 (also known as p39), p38 MAP kinase (also known as MAPK11, 12 and 14) and 14-3-3 zeta (also known as YWHAZ) implicated in tau (also known as MAPT) phosphorylation. REST repressed the expression of these genes in SH-SY5Y cells (Extended Data Fig. 3a, b), and strongly inhibited tau phosphorylation at the AT8 and PHF1 epitopes associated with AD (Extended Data Fig. 3c).

We next examined the regulation of these REST target genes in the normal ageing brain and AD. To distinguish effects at different stages of cognitive decline, we stratified the AD population for least severe (AD1; Mini Mental State Examination (MMSE) score > 18) and most severe (AD2; MMSE score < 8). REST targets related to cell death and AD pathology showed reduced REST binding and elevated mRNA expression at both stages of AD (Fig. 2a, b). REST target genes related to neurotransmission, however, showed elevated mRNA expression in early stage AD1, but reduced expression in later stage AD2 (Fig. 2b). Regression analysis of protein levels showed that nuclear REST is inversely correlated with levels of pro-apoptotic (BAX and DAXX) and AD-related (presenilin 2) proteins in normal ageing and AD neurons (Fig. 2c, d). This was also observed for REST targets associated with neurotransmission (SST and CALB1) during normal ageing and early stage AD1, but not in late stage AD2 (Fig. 2c). The activation-associated histone modification H3K9ac, which is typically reduced by REST, was globally downregulated in normal ageing PFC neurons (Fig. 2e) and was inversely correlated with REST levels (Fig. 2c, d). In contrast, neuronal

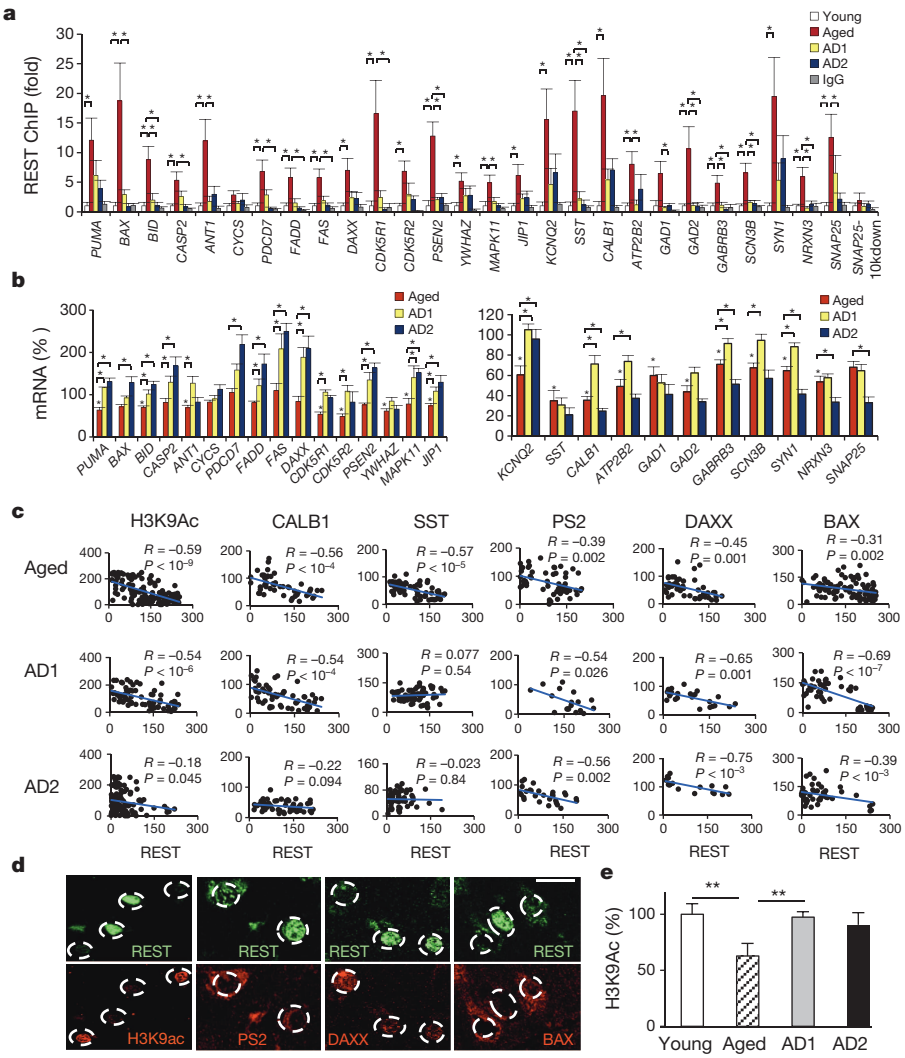


Figure 2 | Regulation of REST target genes in ageing and AD. **a**, ChIP-qPCR analysis of REST binding to target genes in isolated PFC neuronal nuclei. Shown are young adult (Young), normal aged (Aged) and AD cases, which were stratified as AD1 (MMSE score > 18) and AD2 (MMSE score < 8). Also shown are control ChIP assays with non-specific IgG, and a REST ChIP with PCR primers directed 10 kb 3′ to the *SNAP25* RE1 site (*SNAP25*-10K down). Values represent fold change relative to the mean young adult value, and represent the mean ± s.e.m., $n = 7$. $^*P < 0.05$ by the Mann–Whitney test. **b**, mRNA expression of REST target genes determined by qRT–PCR of PFC. Shown are REST target genes related to cell death pathways and AD pathology (left), and neurotransmission (right). Values are normalized to the mean young adult expression level (100%) and represent the mean ± s.e.m. Young, $n = 9$; Aged, $n = 10$; AD1, $n = 10$; AD2, $n = 10$. Asterisks above the 'Aged' bars indicate significance relative to young adults; brackets indicate significance of Aged versus AD1 or AD2. $^*P < 0.05$ by the Mann–Whitney test. **c**, Linear regression analysis of nuclear REST versus levels of protein targets or histone H3K9ac in double-labelled PFC neurons. **d**, Representative images showing inverse relationships between nuclear REST and H3K9ac, PS2, DAXX and BAX in AD1 neurons. Scale bar, 20 μm. **e**, Global epigenetic regulation in ageing and AD. Histone modification H3K9ac was determined in isolated PFC neuronal nuclei. Values are normalized to the mean of the young adult group (100%), and represent the mean ± s.e.m., $^{**}P < 0.005$ by Student's unpaired *t*-test. Young, $n = 8$; Aged, $n = 11$; AD1, $n = 4$; AD2, $n = 4$.

H3K9ac increased significantly in early AD (Fig. 2e). Thus, loss of nuclear REST in AD is associated with epigenetic derepression of potentially pathogenic genes.

REST is neuroprotective

Because REST represses pro-apoptotic genes, we explored a role in neuroprotection by assessing vulnerability to oxidative stress and oligomeric Aβ, stressors associated with brain ageing[9] and AD[11]. Primary cultures of mouse E16 cortical neurons were established from *Nestin-Cre:REST*$^{lx/lx}$ (REST cKO) and littermate control embryos. After 1–2 weeks in culture, neuronal differentiation was comparable in REST-deficient and control cultures. Upon treatment with hydrogen peroxide or oligomeric Aβ42, REST-deficient neurons showed markedly increased degeneration and cell death relative to controls (Fig. 3a and Extended Data Fig. 4a). REST-deficient neurons were rescued by lentiviral transduction of wild-type REST, leading to expression levels 2–2.5-fold that of wild type (Fig. 3a, Extended Data Fig. 4a). In contrast, a REST mutant that is deficient in nuclear translocation (ΔREST)[12,13] was unable to rescue (Fig. 3a, b). REST-deficient neurons showed increased expression of pro-apoptotic genes in response to oxidative stress and oligomeric Aβ (Extended Data Fig. 4b). Thus, endogenous REST is neuroprotective.

REST-deficient mice showed a progressive age-related neurodegenerative phenotype. At 1 month of age, neuronal numbers in the cortex and hippocampus of REST cKO mice were not significantly different from controls. Neuronal degeneration and apoptosis appeared by 8 months of age, as indicated by positive neuronal labelling for cleaved caspase-3 and TdT-mediated dUTP nick end labelling (TUNEL; Fig. 3c, d). This was accompanied by significant neuronal loss in the cerebral cortex and hippocampus, and pronounced gliosis (Fig. 3c, d). These results indicate that REST is essential for maintaining neuronal viability in the ageing brain.

To confirm the findings on REST-deficient neurons, we assessed vulnerability to oxidative stress after REST overexpression or short hairpin RNA-mediated knockdown in neural SH-SY5Y cells. Cell death induced by hydrogen peroxide was significantly reduced when REST was overexpressed by 3- to 20-fold (Extended Data Fig. 4c). Higher levels of REST overexpression increased cell death. REST knockdown by lentiviral shRNA transduction significantly increased H_2O_2-induced apoptosis in both SH-SY5Y cells and cultured human cortical neurons, which was prevented by shRNA-resistant mouse REST (Extended Data Fig. 4c, d). Furthermore, shRNA-mediated REST knockdown increased levels of reactive oxygen species (ROS), which was prevented by the antioxidant *N*-acetyl-cysteine (NAC) (Extended Data Fig. 4e, f). Moreover, REST knockdown increased oxidative DNA damage, which was reduced by REST overexpression (Extended Data Fig. 4g). To examine the relationship of REST expression to oxidative DNA damage in the human brain, we double-labelled for REST and the oxidative DNA damage marker 8-oxoguanine (8-oxoG). Linear regression analysis showed a significant inverse relationship between nuclear REST and 8-oxoG levels in normal ageing neurons and in AD (Extended Data Fig. 4h). Thus, REST expression during ageing may reduce oxidative damage.

The nematode *Caenarhobditis elegans* has served as a model system of ageing and stress resistance[4]. Two multi-zinc finger transcription factors in *C. elegans* structurally resemble mammalian REST (*spr-3* and

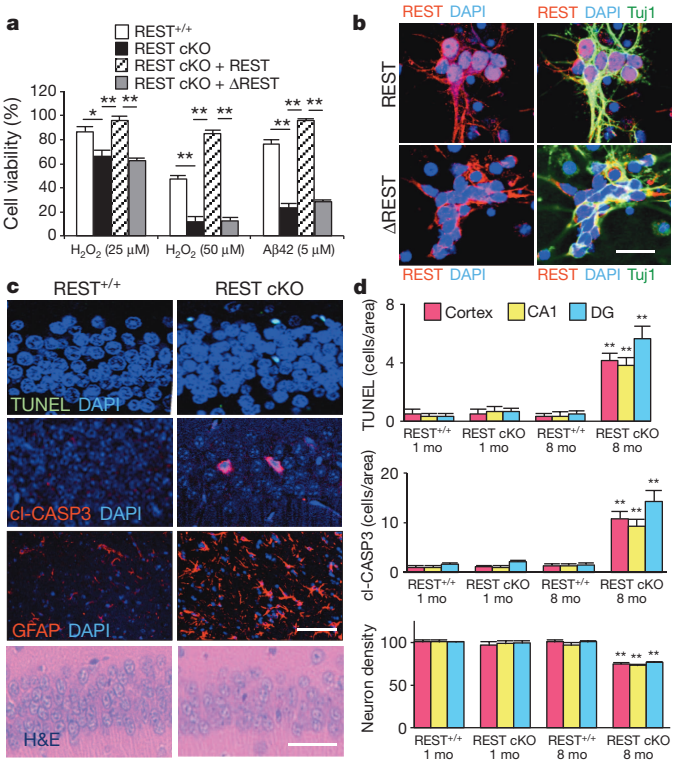


Figure 3 | REST is neuroprotective. **a**, REST-deficient cortical neurons cultured from REST cKO mouse embryos show increased vulnerability to H_2O_2 and oligomeric Aβ42 toxicity. Cell viability is expressed as percentage of the value in untreated cultures. Values represent the mean ± s.e.m., $n = 6$–8. $^*P < 0.05$, $^{**}P < 0.01$ by Student's unpaired t-test. **b**, Reduced nuclear translocation of the ΔREST mutant. **c**, REST cKO mice show age-related neurodegeneration. Shown are sections of the hippocampal CA1 subfield from 8-month-old REST cKO mice labelled for TUNEL, cleaved caspase 3 (cl-CASP3) and GFAP, or stained with haematoxylin and eosin (H&E). **d**, Neurodegeneration in 8-months but not 1-month-old REST cKO mice in the cortex, and in hippocampal CA1 and dentate gyrus (DG) regions. Values represent the mean ± s.e.m., $n = 6$ mice; $^{**}P < 0.001$ relative to control by Student's unpaired t-test. Scale bars, 20 μm.

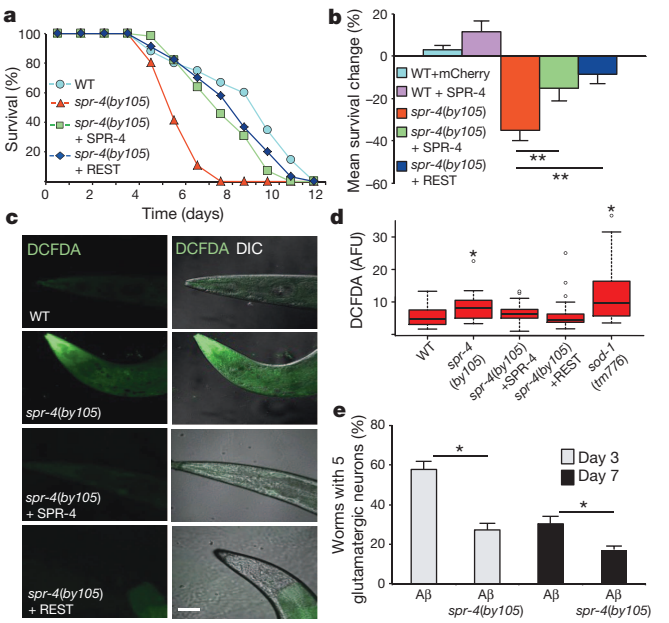


Figure 4 | *Caenorhabditis elegans* SPR-4 protects against oxidative stress and Aβ toxicity. **a**, *spr-4*(*by105*) worms incubated continuously with paraquat (5 mM) exhibit increased mortality rescued by wild-type SPR-4 or human REST. Shown is a representative experiment replicated three times. **b**, Quantitative analysis of survival in wild-type worms expressing mCherry or SPR-4 (WT+mCherry and WT+SPR-4), *spr-4*(*by105*) mutants, and *spr-4*(*by105*) mutants expressing SPR-4 or human REST (*spr4*(*by105*)+SPR4 and *spr4*(*by105*)+REST). Shown is the percentage change in mean survival relative to wild-type. Values represent the mean ± s.d., $n = 3$ independent replicates of at least 30 animals per genotype; $^{**}P < 0.01$ relative to wild type by the log-rank test. **c**, REST and SPR-4 reduce levels of reactive oxygen species (ROS). Shown are representative confocal images of paraquat-treated worms labelled with the ROS-sensitive dye DCFDA. WT, wild type. Scale bar, 30 μm. **d**, Quantitation of ROS levels by DCFDA labelling. Horizontal bars indicate the median; boxed areas represent the second and third quartiles. $^*P < 0.05$ relative to wild type by analysis of variance with post-hoc Tukey test; $n = 30$ worms. AFU, arbitrary fluorescence units. **e**, SPR-4 protects against Aβ neurotoxicity. Shown are Aβ worms (expressing a stably integrated Aβ1-42 transgene[17]) and Aβ;*spr4*(*by105*) worms. Neuronal degeneration does not occur in wild-type worms or *spr4*(*by105*) mutants in the absence of the Aβ transgene. Values represent the percentage of worms that retain five glutamatergic tail neurons at the indicated age (days), and are the mean ± s.d., $n = 3$. $^*P < 0.05$ by unpaired t-test.

spr-4)[14,15], and another gene is closely homologous to CoREST (*spr-1*)[16]. To test the role of these genes in oxidative stress resistance, we placed adult worms on nematode growth medium supplemented with paraquat, a superoxide generator. *C. elegans* lines with mutations in *spr-1*, *spr-3* and *spr-4* all showed significantly reduced survival relative to wild-type controls (Fig. 4a and Extended Data Fig. 5a). The *spr-4*(*by105*) mutant showed the greatest vulnerability. Furthermore, depletion of *spr-4* by RNA interference significantly reduced survival in wild-type worms after treatment with paraquat (Extended Data Fig. 5b). SPR-4 was expressed predominantly in neurons (Extended Data Fig. 5c), and was induced by oxidative stress (Extended Data Fig. 5d).

The increased sensitivity of the *spr-4* mutant worms to oxidative stress was significantly reduced by transgenic expression of either wild-type SPR-4 or human REST (Fig. 4a, b). To ascertain whether SPR-4 and REST directly modulate ROS levels, worms were labelled with the fluorescent ROS indicator 2′,7′-dichlorodihydrofluorescein diacetate (DCFDA). *spr-4*(*by105*) mutants treated with paraquat showed significantly elevated ROS levels relative to wild-type, which was reduced by transgenic expression of either wild-type SPR-4 or human REST (Fig. 4c, d). To determine whether REST can substitute for SPR-4 as a transcriptional repressor, we assessed the expression of the presenilin gene *hop*-1, which is repressed by SPR-4. The expression of *hop-1* was elevated in adult *spr-4* mutants relative to wild-type worms, and was repressed by both wild-type SPR-4 and human REST (Extended Data Fig. 5e). These results indicate that human REST can functionally substitute for SPR-4 in *C. elegans*.

SPR-4 also modulated the neurotoxicity of Aβ in a *C. elegans* line that expresses Aβ42 in glutamatergic neurons and undergoes age-dependent neuronal loss[17]. When this Aβ-expressing line was crossed with the *spr-4*(*by105*) mutant, the resulting line expressed Aβ42 in an *spr-4* loss-of-function background and showed significantly accelerated neurodegeneration (Fig. 4e). Thus, SPR-4 protects against both oxidative stress and Aβ toxicity, consistent with a general role in stress resistance.

Induction of REST by stress and Wnt signalling

To determine whether REST might be induced by oxidative and other forms of stress in the ageing brain, we subjected primary human cortical neuronal cultures to a variety of stressors. Incubation with redox-active iron (Fe^{+2}), hydrogen peroxide, the glutathione synthesis inhibitor buthionine sulphoxide (BSO), or Aβ42-induced REST mRNA and protein and increased REST–RE1 site binding (Extended Data Fig. 6a–f). To determine whether secreted factors might contribute to REST induction, cortical neurons were incubated with H_2O_2 followed by removal

of H_2O_2 and generation of conditioned medium. Addition of stress-related conditioned medium to naive neuronal cultures resulted in a marked induction of REST mRNA expression (Extended Data Fig. 6g). Thus, a variety of stressors induce REST expression, which is mediated at least in part through cell non-autonomous signalling.

To explore the role of cell non-autonomous signalling in the ageing brain, we generated extracts of human PFC from young adult and aged samples and assessed their effects on SH-SY5Y cells. REST was robustly induced by extracts of aged human cortex, but minimally by extracts of young adult cortex (Extended Data Fig. 6h). Furthermore, extracts of AD cortex showed reduced REST-inducing activity relative to age-matched controls (Extended Data Fig. 6h). Because REST has been identified as a target of canonical Wnt-β-catenin signalling[18], we examined the role of this cell non-autonomous signalling pathway. Extracts of aged cortex and conditioned medium from stressed cells treated with H_2O_2, tunicamycin or wortmannin induced β-catenin in parallel with REST (Extended Data Figs 6h and 7a, b). Induction of REST and β-catenin was partially inhibited by the Wnt signalling antagonist Dickkopf (Extended Data Fig. 7a, b). REST was also induced by addition of purified Wnt-3a and Wnt-7a. Moreover, the GSK3-β inhibitors lithium chloride and Chiron 99021 that activate Wnt signalling induced REST, increased REST–RE1 site binding, and augmented nuclear translocation (Extended Data Fig. 7c–f). To assess Wnt signalling in the human brain, we measured nuclear β-catenin levels in the PFC of young adult and aged cases. Nuclear β-catenin was significantly elevated in ageing PFC neurons and colocalized with REST (Extended Data Fig. 7g, h). These results indicate that Wnt-β-catenin signalling may contribute to the induction of REST in the ageing brain.

Autophagy of REST in neurodegenerative disorders

In AD, loss of REST from the nucleus was often accompanied by localization to punctate cytoplasmic structures that were labelled with the autophagosome markers LC3, ATG7 and ATG12, but not the lysosomal markers LAMP1 and LAMP2 (Extended Data Fig. 8a). To determine if autophagy could alter the nuclear:cytoplasmic distribution of REST, autophagy was activated in SH-SY5Y cells by serum deprivation. This resulted in significantly reduced nuclear REST, which was reversed by treatment with 3-methyladenine or bafilomycin, both inhibitors of autophagy (Extended Data Fig. 8b, c). Because activation of autophagy is characteristic of many age-related neurodegenerative disorders, we asked whether REST is also depleted in other dementing diseases. Examination of cortical sections from patients with frontotemporal dementia (FTD) or dementia with Lewy bodies (DLB) showed marked depletion of nuclear REST (Fig. 5a).

Induction of autophagy in neurodegenerative disorders has been associated with pathogenic misfolded proteins such as Aβ (ref. 19). In AD, many cortical and hippocampal neurons showed a notable colocalization of REST with Aβ in LC3-positive autophagosomes (Fig. 5b). In contrast, REST did not colocalize with Aβ in amyloid plaques. Phosphorylated tau and TDP-43 have been implicated as abnormally folded pathogenic proteins in FTD, whereas α-synuclein has been implicated in DLB. Every case (8) of tau-positive FTD examined showed colocalization of REST with phosphorylated tau (PHF1 epitope) in LC3-positive autophagosomes (Fig. 5b). Similarly, every case (6) of TDP-43-positive FTD showed colocalization of REST with TDP-43 and LC3 (Fig. 5b). All ten cases of DLB examined showed partial colocalization of REST with α-synuclein in autophagosomes (Fig. 5b). Thus, autophagy of REST,

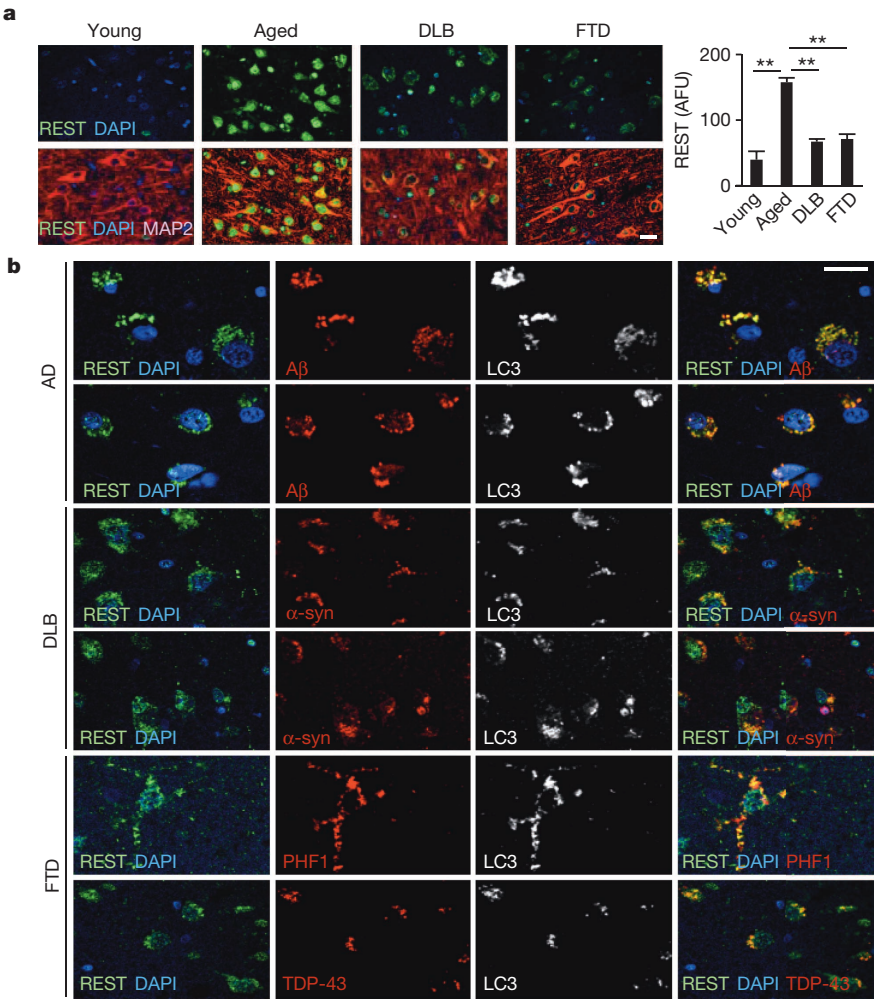


Figure 5 | REST, autophagy and proteostasis in AD, DLB and FTD. **a**, Loss of nuclear REST in DLB and FTD. Left, immunofluorescence microscopy with labelling for REST (green), the neuronal marker MAP2 (red) and DNA (DAPI, blue). Right, quantitative analysis of immunofluorescence. Values (AFU) represent the mean ± s.e.m.; Young, $n = 11$; Aged, $n = 21$; DLB, $n = 18$; FTD, $n = 16$. $^{**}P < 0.001$ by Student's unpaired t-test. **b**, Colocalization of REST with disease-associated misfolded proteins in autophagosomes. Cortical sections were triple-labelled for REST, LC3, and either Aβ in AD, α-synuclein (α-syn) in DLB, or phosphorylated tau (PHF1 epitope) or TDP-43 in FTD. Two different representative cases are shown for each disease. Scale bars, 20 μm.

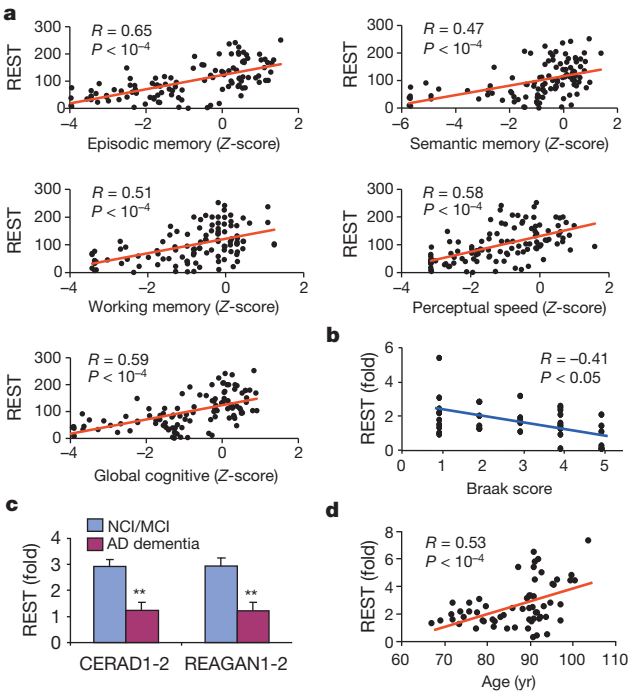


Figure 6 | Nuclear REST is positively correlated with cognitive function and longevity. **a**, Linear regression analysis of nuclear REST levels in PFC neurons and cognitive test scores. Nuclear REST was imaged by anti-REST immunofluorescence and quantified by MetaMorph. Each point represents an individual case. $n = 111$ cases (59 females, 52 males), age range 71–90 years. **b**, Nuclear REST levels are inversely correlated with Braak stage (extent of neurofibrillary tangle formation). **c**, Nuclear REST levels in PFC from cases with neuropathological AD (moderate/frequent plaques by CERAD score and intermediate/high likelihood AD by NIA-Reagan criteria) that had mild or no cognitive impairment (NCI/MCI), or AD dementia. Values are the mean ± s.e.m. $*P < 0.01$ by Student's unpaired t-test; NCI/MCI $n = 30$, AD dementia $n = 21$. **d**, Nuclear REST levels in ageing neurons correlate positively with longevity. REST levels were quantified by FACS analysis of isolated PFC neuronal nuclei in 61 individuals without AD (age range 67–104 years).

together with pathogenic misfolded proteins, may reduce nuclear translocation in AD and other neurodegenerative disorders.

REST, cognitive preservation and longevity

A central question is whether loss of REST in AD contributes to cognitive decline. We explored this question by using measures of cognitive function derived from neuropsychiatric assessments performed longitudinally in the Religious Orders Study and the Rush Memory and Ageing Project (Methods). Linear regression analysis showed that nuclear REST levels in prefrontal cortical neurons are positively correlated with a measure of global cognition, and separate measures of episodic, semantic and working memory with high levels of statistical significance (Fig. 6a and Extended Data Fig. 9a). Similar results were obtained in pyramidal neurons in the hippocampal CA1 region. For both brain regions, REST levels were most robustly correlated with episodic memory, which typically declines early in the course of AD. Elevated REST levels also predicted reduced AD neuropathology (Fig. 6b and Extended Data Fig. 9b).

We then asked if elevated REST levels might confer resistance to AD pathology. Cases with National Institutes on Aging and Reagan Institute (NIA-Reagan) and Consortium to Establish a Registry for Alzheimer's disease (CERAD) scores indicative of pathological AD, but with mild or no cognitive impairment, were compared to cases with pathological AD together with severe cognitive impairment. Nuclear REST levels were significantly elevated in cognitively preserved cases relative to cognitively impaired cases, despite the fact that both groups met pathological criteria for AD (Fig. 6c). Thus, elevated REST levels are associated with cognitive preservation in individuals who meet pathological criteria for AD.

Because REST is associated with stress resistance and preservation of cognitive function, we explored the relationship to human lifespan. Nuclear REST levels in PFC neurons showed a significant positive correlation with longevity ($R = 0.53$; $P < 10^{-4}$) (Fig. 6d). This relationship remained significant after multivariate adjustment for cognitive test scores and indices of neuropathology. Furthermore, the expression of REST target genes showed a highly significant inverse correlation with longevity (Extended Data Fig. 2a). Examination of the neuroanatomical distribution of REST in individuals with extreme longevity showed marked induction in neurons of the PFC and hippocampus, but not in the Purkinje or granule cell neurons of the cerebellum (Extended Data Fig. 10a). Thus, REST expression in specific neuronal populations correlates with longevity.

We then asked if REST regulates genes that have been implicated in stress resistance and the ageing process. In addition to repressing cell death genes, REST also promoted expression of the anti-apoptotic gene *BCL2*, the antioxidants catalase and *SOD1*, and the FOXO transcription factor *FOXO1* (Extended Data Figs 3a and 10b). As *FOXO1* has been implicated in oxidative stress resistance[20] and human longevity[21], we assessed its regulation by REST at the protein level. REST knockdown by shRNA transduction in SH-SY5Y cells almost completely abolished FOXO1 expression, which was reversed by an shRNA-resistant mouse REST vector (Extended Data Fig. 10c). Thus, REST may broadly confer stress resistance during ageing.

Discussion

We have demonstrated a striking induction of the developmental transcriptional repressor REST in specific neuronal populations of the ageing human brain together with epigenetic repression of REST target genes. REST has been investigated extensively as a repressor of neuronal genes during embryonic development, a function that persists in adult non-neuronal cells[22]. Our findings suggest that REST additionally plays a role as a neuroprotective modulator, in part by repressing genes that promote cell death and the pathology of AD. Moreover, REST increases the expression of FOXO transcription factors that mediate oxidative stress resistance[20], as well as the antioxidant enzymes catalase and SOD1, possibly through indirect mechanisms such as repression of micro RNAs. Consistent with these findings, REST confers oxidative stress resistance and protects against toxic insults associated with AD, including Aβ oligomers and tau phosphorylation. Furthermore, REST appears to be essential for maintaining neuronal viability in the normal ageing cortex and hippocampus. This protective function is conserved in *C.elegans*. Previous ChIP-SACO[23] and ChIP-chip[24] studies also identified REST target genes involved in cell death. In addition, a recent study suggests that REST protects against the toxicity of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a model of Parkinson's disease[25]. Thus, REST may coordinate a stress response that is broadly neuroprotective in the ageing brain.

REST dysfunction may contribute to the pathogenesis of a number of different neurodegenerative disorders. In addition to AD, REST was also significantly depleted in frontotemporal dementia and dementia with Lewy bodies. In each of these disorders, REST was lost from the nucleus and appeared in autophagosomes together with pathological misfolded proteins, including Aβ, phosphorylated tau, TDP-43 and α-synuclein. This may represent a common pathogenic mechanism that links altered proteostasis to aberrant gene expression.

Recent studies suggest that epigenetic regulation of chromatin may modulate the cognitive outcome of a variety of neuropathological states[26–28]. It is intriguing that ageing individuals who harbour substantial AD pathology do not appear to progress to dementia when neuronal REST levels are high. This raises the possibility that structural pathology, such as Aβ deposition and neurofibrillary tangles, may not be sufficient to cause dementia. Rather, failure of the brain's stress response system may also be necessary, suggesting new possibilities for therapeutic intervention.

METHODS SUMMARY

Post-mortem human brain material was procured in accordance with institutional guidelines governed by approved protocols. Tissue samples were procured from the Rush University Medical Center, the University of Maryland, Duke University, the Brigham and Women's Hospital, and the Massachusetts General Hospital. Mouse protocols were approved by the Institutional Animal Care and Use Committee of Harvard Medical School.

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

Received 27 June 2013; accepted 21 February 2014.
Published online 19 March 2014.

1. Gómez-Isla, T. *et al.* Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *J. Neurosci.* **16,** 4491–4500 (1996).
2. Peters, A., Sethares, C. & Moss, M. B. The effects of aging on layer 1 in area 46 of prefrontal cortex in the rhesus monkey. *Cereb. Cortex* **8,** 671–684 (1998).
3. Yankner, B. A., Lu, T. & Loerch, P. The aging brain. *Annu. Rev. Pathol.* **3,** 41–66 (2008).
4. Kenyon, C. J. The genetics of ageing. *Nature* **464,** 504–512 (2010).
5. Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H. & Frisen, J. Retrospective birth dating of cells in humans. *Cell* **122,** 133–143 (2005).
6. Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80,** 949–957 (1995).
7. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267,** 1360–1363 (1995).
8. Ballas, N., Grunseich, C., Lu, D. D., Speh, J. C. & Mandel, G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121,** 645–657 (2005).
9. Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429,** 883–891 (2004).
10. Loerch, P. M. *et al.* Evolution of the aging brain transcriptome and synaptic regulation. *PLoS ONE* **3,** e3329 (2008).
11. Holtzman, D. M., Morris, J. C. & Goate, A. M. Alzheimer's disease: the challenge of the second century. *Sci. Transl. Med.* **3,** 77sr1 (2011).
12. Grimes, J. A. *et al.* The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex. *J. Biol. Chem.* **275,** 9461–9467 (2000).
13. Shimojo, M. Characterization of the nuclear targeting signal of REST/NRSF. *Neurosci. Lett.* **398,** 161–166 (2006).
14. Wen, C., Levitan, D., Li, X. & Greenwald, I. *spr-2*, a suppressor of the egg-laying defect caused by loss of *sel-12* presenilin in *Caenorhabditis elegans*, is a member of the SET protein subfamily. *Proc. Natl Acad. Sci. USA* **97,** 14524–14529 (2000).
15. Lakowski, B. *et al.* Two suppressors of *sel-12* encode C_2H_2 zinc-finger proteins that regulate presenilin transcription in *Caenorhabditis elegans*. *Development* **130,** 2117–2128 (2003).
16. Jarriault, S. & Greenwald, I. Suppressors of the egg-laying defective phenotype of *sel-12* presenilin mutants implicate the CoREST corepressor complex in LIN-12/Notch signaling in *C. elegans*. *Genes Dev.* **16,** 2713–2728 (2002).
17. Treusch, S. *et al.* Functional links between Aβ toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. *Science* **334,** 1241–1245 (2011).
18. Willert, J., Epping, M., Pollack, J. R., Brown, P. O. & Nusse, R. A transcriptional response to Wnt protein in human embryonic carcinoma cells. *BMC Dev. Biol.* **2,** 8 (2002).
19. Lipinski, M. M. *et al.* Genome-wide analysis reveals mechanisms modulating autophagy in normal brain aging and in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **107,** 14164–14169 (2010).
20. Tothova, Z. *et al.* FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128,** 325–339 (2007).
21. Li, Y. *et al.* Genetic association of *FOXO1A* and *FOXO3A* with longevity trait in Han Chinese populations. *Hum. Mol. Genet.* **18,** 4897–4904 (2009).
22. Ballas, N. & Mandel, G. The many faces of REST oversee epigenetic programming of neuronal genes. *Curr. Opin. Neurobiol.* **15,** 500–506 (2005).
23. Otto, S. J. *et al.* A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions. *J. Neurosci.* **27,** 6729–6739 (2007).
24. Abrajano, J. J. *et al.* REST and CoREST modulate neuronal subtype specification, maturation and maintenance. *PLoS ONE* **4,** e7936 (2009).
25. Yu, M. *et al.* NRSF/REST neuronal deficient mice are more vulnerable to the neurotoxin MPTP. *Neurobiol. Aging* **34,** 916–927 (2013).
26. Fischer, A., Sananbenesi, F., Wang, X., Dobbin, M. & Tsai, L. H. Recovery of learning and memory is associated with chromatin remodelling. *Nature* **447,** 178–182 (2007).
27. Gräff, J. & Tsai, L. H. Histone acetylation: molecular mnemonics on the chromatin. *Nature Rev. Neurosci.* **14,** 97–111 (2013).
28. Ronan, J. L., Wu, W. & Crabtree, G. R. From neural development to cognition: unexpected roles for chromatin. *Nature Rev. Genet.* **14,** 347–359 (2013).
29. Bruce, A. W. *et al.* Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc. Natl Acad. Sci. USA* **101,** 10458–10463 (2004).

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

Acknowledgements We thank members of the Yankner laboratory for suggestions and discussion, Monlan Yuan, Allison Harwick, Kelly Dakin and Gregory Klein for assistance, and Cheng Li and Dana Gabuzda for helpful discussion. We also acknowledge the Rush Alzheimer's Disease Center, the Brigham and Women's Hospital Brain Bank, the Massachusetts General Hospital ADRC Brain Bank, and the Kathleen Price Bryan Brain Bank at Duke University for providing tissue samples. This work was supported by an NIH Director's Pioneer Award (DP1OD006849) and NIH grants PO1AG27916 and RO1AG26651 to B.A.Y., RO1GM072551 to M.P.C., P30AG10161, R01AG15819 and R01AG17917 to D.A.B., and a grant from the Glenn Foundation for Medical Research to B.A.Y. J.Z. is a Molecular Biology of Neurodegeneration fellow at Harvard Medical School.

Author Contributions T.L., L.A., J.Z., Y.P., H.K., and H.-M.K. performed experiments. T.L., L.A., J.Z., M.C. and B.A.Y. contributed to the overall study design. T.L., Y.C., T.-H.Y., D.D. and X.L. performed informatics analysis. D.A.B. contributed tissue samples, cognitive test data and analysis. B.A.Y. directed the study and B.A.Y., T.L., L.A. and J.Z. wrote the manuscript, which was edited by all the coauthors.

Author Information Data are available in the Gene Expression Omnibus (GEO) under accession GSE53890. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.A.Y. (Bruce_Yankner@hms.harvard.edu).

ARTICLE

doi:10.1038/nature12787

An atlas of active enhancers across human cell types and tissues

Robin Andersson[1]*, Claudia Gebhard[2,3]*, Irene Miguel-Escalada[4], Ilka Hoof[1], Jette Bornholdt[1], Mette Boyd[1], Yun Chen[1], Xiaobei Zhao[1,5], Christian Schmidl[2], Takahiro Suzuki[6,7], Evgenia Ntini[8], Erik Arner[6,7], Eivind Valen[1,9], Kang Li[1], Lucia Schwarzfischer[2], Dagmar Glatz[2], Johanna Raithel[2], Berit Lilje[1], Nicolas Rapin[1,10], Frederik Otzen Bagger[1,10], Mette Jørgensen[1], Peter Refsing Andersen[8], Nicolas Bertin[6,7], Owen Rackham[6,7], A. Maxwell Burroughs[6,7], J. Kenneth Baillie[11], Yuri Ishizu[6,7], Yuri Shimizu[6,7], Erina Furuhata[6,7], Shiori Maeda[6,7], Yutaka Negishi[6,7], Christopher J. Mungall[12], Terrence F. Meehan[13], Timo Lassmann[6,7], Masayoshi Itoh[6,7,14], Hideya Kawaji[6,14], Naoto Kondo[6,14], Jun Kawai[6,14], Andreas Lennartsson[15], Carsten O. Daub[6,7,15], Peter Heutink[16], David A. Hume[11], Torben Heick Jensen[8], Harukazu Suzuki[6,7], Yoshihide Hayashizaki[6,14], Ferenc Müller[4], The FANTOM Consortium†, Alistair R. R. Forrest[6,7], Piero Carninci[6,7], Michael Rehli[2,3] & Albin Sandelin[1]

Enhancers control the correct temporal and cell-type-specific activation of gene expression in multicellular eukaryotes. Knowing their properties, regulatory activity and targets is crucial to understand the regulation of differentiation and homeostasis. Here we use the FANTOM5 panel of samples, covering the majority of human tissues and cell types, to produce an atlas of active, *in vivo*-transcribed enhancers. We show that enhancers share properties with CpG-poor messenger RNA promoters but produce bidirectional, exosome-sensitive, relatively short unspliced RNAs, the generation of which is strongly related to enhancer activity. The atlas is used to compare regulatory programs between different cells at unprecedented depth, to identify disease-associated regulatory single nucleotide polymorphisms, and to classify cell-type-specific and ubiquitous enhancers. We further explore the utility of enhancer redundancy, which explains gene expression strength rather than expression patterns. The online FANTOM5 enhancer atlas represents a unique resource for studies on cell-type-specific enhancers and gene regulation.

Precise regulation of gene expression in time and space is required for development, differentiation and homeostasis[1]. Sequence elements within or near core promoter regions contribute to regulation[2], but promoter-distal regulatory regions like enhancers are essential in the control of cell-type specificity[1]. Enhancers were originally defined as remote elements that increase transcription independently of their orientation, position and distance to a promoter[3]. They were only recently found to initiate RNA polymerase II (RNAPII) transcription, producing so-called eRNAs[4]. Genomic locations of enhancers can be detected by mapping of chromatin marks and transcription factor binding sites from chromatin immunoprecipitation (ChIP) assays and DNase I hypersensitive sites (DHSs) (reviewed in ref. 1), but there has been no systematic analysis of enhancer usage in the large variety of cell types and tissues present in the human body.

Using cap analysis of gene expression[5] (CAGE), we show that enhancer activity can be detected through the presence of balanced bidirectional capped transcripts, enabling the identification of enhancers from small primary cell populations. Based upon the FANTOM5 CAGE expression atlas encompassing 432 primary cell, 135 tissue and 241 cell line samples from human[6], we identify 43,011 enhancer candidates and characterize their activity across the majority of human cell types and tissues. The resulting catalogue of transcribed enhancers enables classification of ubiquitous and cell-type-specific enhancers, modelling of physical interactions between multiple enhancers and TSSs, and identification of potential disease-associated regulatory single nucleotide polymorphisms (SNPs).

Bidirectional capped RNAs identify active enhancers

The FANTOM5 project has generated a CAGE-based transcription start site (TSS) atlas across a broad panel of primary cells, tissues and cell lines covering the vast majority of human cell types[6]. Within that data set, well-studied enhancers often have CAGE peaks delineating nucleosome-deficient regions (NDRs) (Supplementary Fig. 1). To determine whether this is a general enhancer feature, FANTOM5 CAGE (Supplementary Table 1) was superimposed on active (H3K27ac-marked) enhancers defined by HeLa-S3 ENCODE ChIP-seq data[7]. CAGE tags showed a bimodal distribution flanking the central P300 peak, with divergent transcription from the enhancer (Fig. 1a). Similar patterns

[1]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre, University of Copenhagen, Ole Maaloes Vej 5, DK-2200 Copenhagen, Denmark. [2]Department of Internal Medicine III, University Hospital Regensburg, Franz-Josef-Strauss-Allee 11, 93042 Regensburg, Germany. [3]Regensburg Centre for Interventional Immunology (RCI), D-93042 Regensburg, Germany. [4]School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [5]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [6]RIKEN OMICS Science Centre, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan. [7]RIKEN Center for Life Science Technologies (Division of Genomic Technologies), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan. [8]Centre for mRNP Biogenesis and Metabolism, Department of Molecular Biology and Genetics, C.F. Møllers Alle 3, Building 1130, DK-8000 Aarhus, Denmark. [9]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [10]The Finsen Laboratory, Rigshospitalet and Danish Stem Cell Centre (DanStem), University of Copenhagen, Ole Maaloes Vej 5, DK-2200, Denmark. [11]Roslin Institute, Edinburgh University, Easter Bush, Midlothian, Edinburgh EH25 9RG, UK. [12]Genomics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road MS 64-121, Berkeley, California 94720, USA. [13]EMBL Outstation - Hinxton, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [14]RIKEN Preventive Medicine and Diagnosis Innovation Program, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan. [15]Department of Biosciences and Nutrition, Karolinska Institutet, Hälsovägen 7, SE-4183 Huddinge, Stockholm, Sweden. [16]Department of Clinical Genetics, VU University Medical Center, van der Boechorststraat 7, 1081 BT Amsterdam, Netherlands.
*These authors contributed equally to this work.
†A list of authors and affiliations appears in the Supplementary Information.

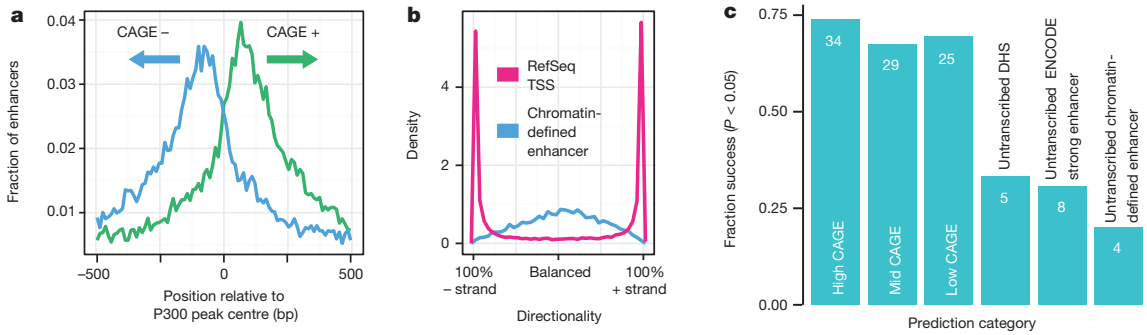


Figure 1 | Bidirectional capped RNAs is a signature feature of active enhancers. **a**, Enhancers identified by co-occurrence of H3K27ac and H3K4me1 ChIP-seq data[7], centred on P300 binding sites, in HeLa cells were overlaid with HeLa CAGE data (unique positions of CAGE tag 5′ ends, smoothed by a 5-bp window), revealing a bidirectional transcription pattern. Horizontal axis shows the ± 500 bp region around enhancer midpoints. **b**, Density plot illustrating the difference in directionality of transcription according to FANTOM5-pooled CAGE tags mapped within ± 300 bp of 22,486 TSSs of RefSeq protein-coding genes and centre positions of 10,138 HeLa enhancers defined as above. **c**, Success rates of 184 *in vitro* enhancer assays in HeLa cells. Vertical axis shows the fraction of active enhancers (success defined by Student's *t*-test, $P < 0.05$ versus random regions; also see Supplementary Fig. 9). Numbers of successful assays are shown on the respective bar. See main text for details.

were observed in other cell lines (Supplementary Fig. 2a). Enhancer-associated reverse and forward strand transcription initiation events were, on average, separated by 180 base pairs (bp) and corresponded to nucleosome boundaries (Supplementary Figs 3 and 4). As a class, active HeLa-S3 enhancers had 231-fold more CAGE tags than polycomb-repressed enhancers, indicating that transcription is a marker for active usage. Indeed, ENCODE-predicted enhancers[7] with significant reporter activity[8] had greater CAGE expression levels than those lacking reporter activity ($P < 4 \times 10^{-22}$, Mann–Whitney *U* test). A lenient threshold on enhancer expression increased the validation rate of ENCODE enhancers from 27% to 57% (Supplementary Fig. 5).

Although capped RNAs of protein-coding gene promoters were strongly biased towards the sense direction, similar levels of capped RNA in both directions were detected at enhancers (Fig. 1b and Supplementary Fig. 2b, c). Thus, bidirectional capped RNAs is a signature feature of active enhancers. On this basis, we identified 43,011 enhancer candidates across 808 human CAGE libraries (see Supplementary Text and Supplementary Figs 6–8). Interestingly, the candidates were depleted of CpG islands (CGI) and repeats (with the exception of neural stem cells, see ref. 9).

To confirm the activity of newly identified candidate enhancers, we randomly selected 46 strong, 41 moderate and 36 low activity enhancers (as defined by CAGE tag frequency in HeLa cells) and examined their activity using enhancer reporter assays compared to randomly selected untranscribed loci with regulatory potential in HeLa-S3 cells: 15 DHSs[10], 26 ENCODE-predicted 'strong enhancers'[7] and 20 enhancers defined as in Fig. 1a (Supplementary Tables 2 and 3). Whereas 67.4–73.9% of the CAGE-defined enhancers showed significant reporter activity, only 20–33.3% of the untranscribed candidate regulatory regions were active (Fig. 1c and Supplementary Fig. 9a). The same trend was observed in HepG2 cells (Supplementary Fig. 10a, b). Corresponding promoter-less constructs showed that the enhancer transcription read-through is negligible (Supplementary Fig. 9b, c). Many CAGE-defined enhancers overlapped predicted ENCODE 'strong enhancers' or 'TSS' states (25% and 62%, respectively, for HeLa-S3), but there was no substantial difference in validation rates between these classes (Supplementary Fig. 10c, d). In summary, active CAGE-defined enhancers were much more likely to be validated in functional assays than untranscribed candidate enhancers defined by histone modifications or DHSs.

Initiation and fate of enhancer RNAs versus mRNAs

RNA-seq data from matching primary cells and tissues showed that ~95% of RNAs originating from enhancers were unspliced and typically short (median 346 nucleotides)—a striking difference to mRNAs

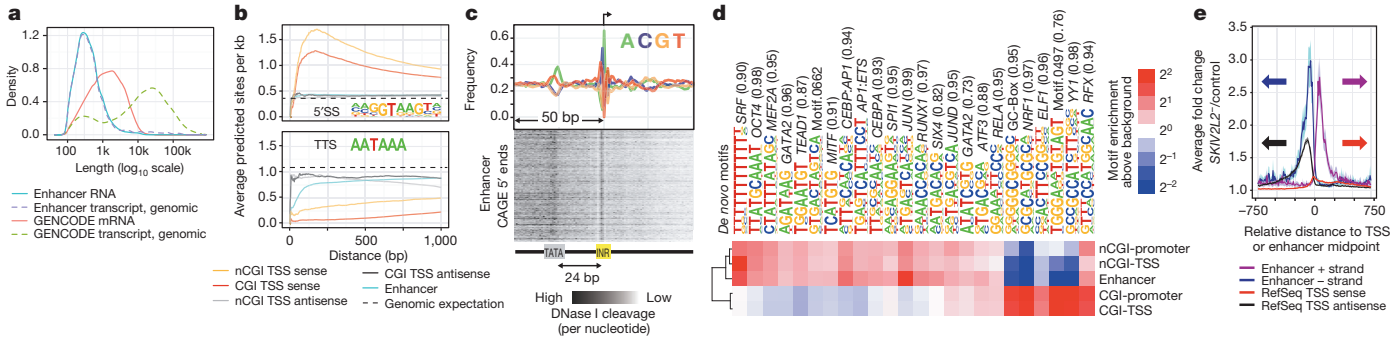


Figure 2 | Features distinguishing enhancer TSSs from mRNA TSSs. **a**, Densities of the genomic and processed RNA lengths of transcripts starting from enhancer TSSs and mRNA TSSs using assembled RNA-seq reads from 13 pooled FANTOM5 libraries. **b**, Frequencies of RNA processing motifs (5′ splice motif (5′SS, upper panel) and the transcription termination site hexamer (TTS, lower panel)) around enhancer and mRNA TSSs. Vertical axis shows the average number of predicted sites per kb within a certain window size from the TSS (horizontal axis) in which the motif search was done. Dashed lines indicate expected hit density from random genomic background. The window always starts at the gene or enhancer CAGE summits and expands in the sense direction. nCGI, non-CGI. **c**, Average nucleotide frequencies (top panel) and DNase I cleavage patterns (lower panel) of enhancer CAGE peaks (arrow at +1 indicates position of the main enhancer CAGE peaks; direction of transcription goes left to right) reveal distinct cleavage patterns at sequences resembling the INR and TATA elements. **d**, *De novo* motif enrichment analyses around enhancers and non-enhancer FANTOM5 CAGE-defined TSSs (CAGE TSSs matching annotated TSSs are referred to as 'promoters'), contingent on CGI overlap. Top enriched/depleted motifs are shown along with their best-known motif match name. Enrichment versus random background is presented as a heat map. **e**, Vertical axis shows average HeLa CAGE expression fold change versus control at enhancers and RefSeq TSSs after exosome depletion. Horizontal axis shows position relative to the TSS or the centre of the enhancer. Translucent colours indicate the 95% confidence interval of the mean.

(19% unspliced, median 1,256 nucleotides) (Fig. 2a and Supplementary Fig. 11a–c). Unlike TSSs of mRNAs, which are enriched for predicted 5′ splice sites but depleted of downstream polyadenylation signals[11,12], enhancers showed no evidence of associated downstream RNA processing motifs, and thus resemble antisense promoter upstream transcripts (PROMPTs)[11] (Fig. 2b and Supplementary Fig. 11d). Most CAGE-defined enhancers gave rise to nuclear (>80%) and non-polyadenylated (~90%) RNAs[13] (Supplementary Fig. 11e). Based on RNA-seq, few enhancer RNAs overlap exons of known protein-coding genes or large intergenic noncoding RNAs (9 and 1 out of 4,208 enhancers detected, respectively), indicating that they are not a substantial source of alternative promoters for known genes (as in ref. 14).

TSS-associated, uncapped small RNAs (TSSa-RNAs), attributed to RNAPII protection and found immediately downstream of mRNA TSSs[15,16], were detectable in the same positions downstream of enhancer TSSs (Supplementary Fig. 12), indicating that RNAPII initiation at enhancer and mRNA TSSs is similar. Indeed, CAGE-defined enhancer TSSs resembled the proximal position-specific sequence patterns of non-CGI RefSeq TSSs (Fig. 2c and Supplementary Fig. 13a). Furthermore, *de novo* motif analysis revealed sequence signatures in CAGE-defined enhancers closely resembling non-CGI promoters (Fig. 2d and Supplementary Fig. 13b).

Because of the similarity with PROMPTs, we reasoned that capped enhancer RNAs might be rapidly degraded by the exosome. Indeed, small interfering RNA-mediated depletion of the *SKIV2L2* (also known as MTR4) co-factor of the exosome complex resulted in a median 3.14-fold increase of capped enhancer-RNA abundance (Fig. 2e and Supplementary Fig. 14a, b), but only a negligible increase at mRNA TSSs. This increasing trend is similar to that of PROMPT regions upstream of TSSs, although the increase of enhancer RNAs was significantly higher ($P < 4.6 \times 10^{-67}$, Mann–Whitney *U* test; Fig. 2e and Supplementary Fig. 14b, c). Thus, the bidirectional transcriptional activity observed at enhancers is also present at promoters, as suggested previously[17], but in promoters only the antisense RNA is degraded. Furthermore, the CAGE expression of enhancers in control and *SKIV2L2* -depleted cells was proportional (Supplementary Fig. 14d), indicating that virtually all identified enhancers produce exosome-sensitive RNAs. The number of detectable bidirectional CAGE peaks increased 1.7-fold upon *SKIV2L2* depletion and novel enhancer candidates had on average similar, but weaker, chromatin modification signals compared to control HeLa cells (Supplementary Fig. 14e).

CAGE identifies cell-specific enhancer usage

To test whether CAGE expression can identify cell-type-specific enhancer usage *in vivo*, ChIP-seq (H3K27ac and H3K4me1), DNA methylation and triplicate CAGE analyses were performed in five primary blood cell types, and compared to published DHS data (http://www.roadmapepigenomics.org/, Supplementary Table 4). CAGE-defined enhancers were strongly supported by proximal H3K4me1/H3K27ac peaks (71%) and DHSs (87%) from the same cell type. Conversely, H3K4me1 and H3K27ac supported only 24% of DHSs distal to promoters and exons and only 4% of DHSs overlapped CAGE-defined enhancers (Supplementary Fig. 15), indicating that a minority of promoter-distal DHSs identify enhancers. From the opposite perspective, only 11% of H3K4me1/H3K27ac loci overlapped CAGE-defined enhancers and untranscribed loci showed weaker ChIP-seq signals than transcribed ones (Supplementary Fig. 16). Moreover, there was a clear correlation between CAGE, DNase I hypersensitivity, H3K4me1 and H3K27ac for CAGE-defined enhancers expressed in blood cells (Fig. 3a). Accordingly, cell-type-specific enhancer expression corresponds to cell-type-specific histone modifications (Fig. 3b). The majority of selected cell-type-specific enhancers could be validated in corresponding cell lines and were associated with cell-type-specific DNA demethylation (Supplementary Text, Supplementary Fig. 17 and Supplementary Tables 5–8, see also ref. 18). Thus, bidirectional CAGE pairs are robust predictors for cell-type-specific enhancer activity.

An atlas of transcribed enhancers across human cells

The FANTOM5 CAGE library collection[6] enables the dissection of enhancer usage across cell types and tissues comprehensively sampled across the human body. Clustering based on enhancer expression clearly grouped functionally related samples together (Fig. 3c and Supplementary Figs 18 and 19). Although fetal and adult tissue often grouped together, two large fetal-specific clusters were identified: one brain-specific (pink) and one with diverse tissues (green). The fetal-brain cluster is associated with enhancers that are located close to known neural developmental genes, including *NEUROG2*, *SCRT2*, *POU3F2* and *MEF2C* (Supplementary Fig. 18b), for which gene expression patterns correlate with enhancer RNA abundance across libraries, suggesting regulatory interaction (see below). The results corroborate the functional relevance of these enhancers for tissue-specific gene expression and indicate that they are an important part of the regulatory programs of cellular differentiation and organogenesis.

To confirm that candidate enhancers can drive tissue-specific gene expression *in vivo*, five evolutionarily conserved CAGE-defined human enhancers (including the *POU3F2* and *MEF2C*-proximal enhancers

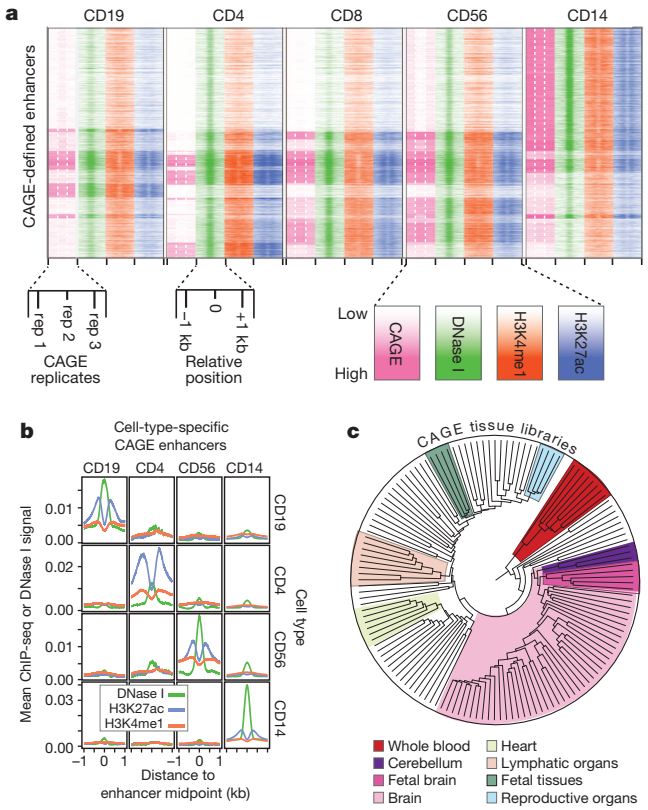


Figure 3 | CAGE expression identifies cell-type-specific enhancer usage. **a**, Relationship between CAGE and histone modifications in blood cells. Rows represent CAGE-defined enhancers that are ordered based on hierarchical clustering of CAGE expression. Columns for the CAGE tags (pink) represent the expression intensity for three biological replicates. DNase I hypersensitivity and H3K27ac and H3K4me1 ChIP-seq signals ± 1kb around the enhancer midpoints are shown in green, blue and orange, respectively. **b**, Mean signal of DNase-seq as well as ChIP-seq for H3K27ac and H3K4me1 (vertical axes) per cell type (rows) in ± 1kb regions (horizontal axes) around enhancer midpoints, for enhancers with blood-cell type-specific CAGE expression (columns). **c**, Dendrogram resulting from agglomerative hierarchical clustering of tissue samples based on their enhancer expression: each leaf of the tree represents one CAGE tissue sample (for a labelled tree and the corresponding results on primary cell samples, see Supplementary Figs 18 and 19). Sub-trees dominated by one tissue/organ type or morphology are highlighted. Some of the enhancers responsible for the fetal-specific subgroup in the larger brain sub-tree are validated *in vivo* (Fig. 4).

identified above) were tested via Tol2-mediated transgenesis in zebrafish embryos. We observed tissue-specific enhancer activity with 3 of 5 fragments, which corresponded to the human enhancer tissue expression (Fig. 4). None of three control fragments without CAGE signal activated the *GATA2* promoter (Supplementary Table 9). Although the sample size is not high enough to reliably estimate the validation rates in zebrafish, the correlation between the enhancer usage profiles in zebrafish to those defined in human by CAGE is notable.

We grouped the primary cell and tissue samples into larger, mutually exclusive cell type and organ/tissue groups (referred to as facets), respectively, with similar function or morphology (Supplementary Tables 10 and 11). Figure 5 summarizes how many enhancers were detected in each facet and the degree of facet-specific CAGE expression (see also Supplementary Fig. 21). From the data we can draw several conclusions:

First, the majority of detected enhancers within any facet are not restricted to that facet. Exceptions, where facets use a higher fraction of specific enhancers, include immune cells, neurons, neural stem cells and hepatocytes amongst the cell-type facets, and brain, blood, liver and testis amongst the organ/tissue facets.

Second, despite their apparent promiscuity, enhancers are more generally detected in a much smaller subset of samples than mRNA transcripts (Supplementary Figs 21 and 22a, b), consistent with cell-line studies[7] and the higher specificity of ncRNAs in general[13]. Facets in which we detect many enhancers typically also have a higher fraction of facet-specific enhancers (Supplementary Fig. 22c, d).

Third, the number of detected expressed enhancers and mRNA transcripts is correlated (Supplementary Fig. 21b), but the number of detected expressed gene transcripts (>1 tag per million mapped reads (TPM)) is 19–34 fold larger than the number of detected enhancers

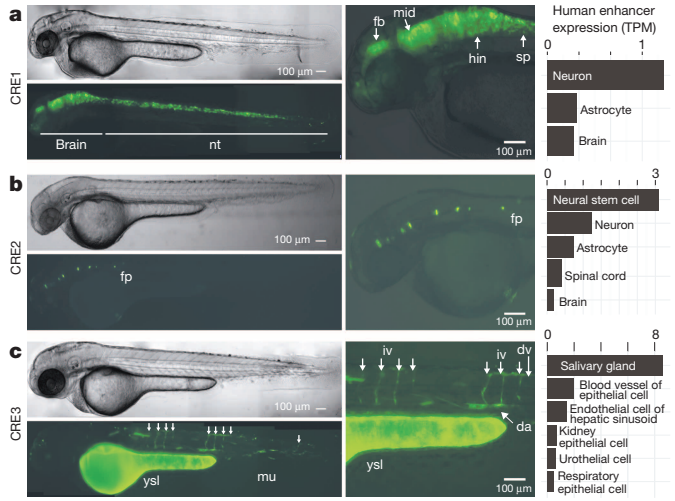


Figure 4 | *In vivo* validation in zebrafish of tissue-specific enhancers. Validations of *in vivo* activity of CAGE-defined human enhancers CRE1, CRE2 and CRE3 in zebrafish embryos at long-pec stage. Each panel shows, from left to right, representative yellow fluorescent protein (YFP) and bright field images of embryos injected with the human enhancer *gata2* promoter reporter gene construct (left), YFP zoom-ins (middle) and CAGE expression in TPM in human tissues/cell types for the enhancer (right). Muscle (mu) and yolk syncytial layer (ysl) activities are background expression coming from the *gata2* promoter-containing reporter construct. All images are lateral, head to the left. Note the correspondence between zebrafish and human enhancer usage/expression. Supplementary Fig. 20 shows UCSC browser images of each selected enhancer. **a**, CRE1, ~230 kb upstream of the *MEFC2* gene, drives highly robust expression in the brain (brain) and neural tube (nt). Right panel gives zoom-in overlay image showing expression in the forebrain (fb), midbrain (mid), hindbrain (hin) and spinal cord (sp). **b**, CRE2, 5 kb upstream of the *POU3F2* gene, is active in the floor plate (fp). **c**, CRE3, 10 kb upstream of the *SOX7* gene TSS, shows specific expression in the vasculature (including intersegmental vessels (iv), dorsal vein (dv) and dorsal aorta (da).

with the cut-offs used. Noteworthy exceptions include blood and immune cells, testis, thymus and spleen, which have high enhancer/gene ratios. Conversely, smooth and skeletal muscle and skin, bone and epithelia-related cells have low ratios. Differential exosome activity between cell types might affect these results, but there was no correlation between *SKIV2L2* mRNA expression and the number of enhancers detected (Supplementary Fig. 22e, f).

As expected, consensus motifs of known key regulators are over-represented in corresponding facet-specific enhancers, for instance ETS, C/EBP and NF-κB in monocyte-specific enhancers, RFX and SOX in neurons, and HNF1 and HNF4a in hepatocytes (Supplementary Fig. 23). Notably, the AP1 motif appears to be enriched across all facets, perhaps associated with a general role for AP1 in regulating open chromatin[19].

Expression clustering reveals ubiquitous enhancers

Hierarchical clustering of enhancers by facet expression revealed a small subset of enhancers (200 or 247, defined by primary cell or tissue facets, respectively) expressed in the large majority of facets (Supplementary Text, Supplementary Figs 24 and 25, and Supplementary Tables 12 and 13). Compared to other enhancers, these ubiquitous (u-) enhancers are 8 times more likely to overlap CGIs and they are twice as conserved (Supplementary Fig. 26a–c). U-enhancers overlap typical chromatin enhancer marks but have higher H3K4me3 signal (Supplementary Fig. 26d). Although they produce significantly longer ncRNAs than other enhancers (median 530 nucleotides, $P < 1.5 \times 10^{-8}$, Mann–Whitney *U* test), the transcripts remain predominantly (~78%) unspliced and significantly shorter ($P < 4.2 \times 10^{-18}$, Mann–Whitney *U* test) than mRNAs (Supplementary Figs 27 and 28), do not share exons with known genes, and are exosome-sensitive (Supplementary Fig. 14b). Therefore, it is unlikely that these are novel mRNA promoters. They are also highly enriched for P300 and cohesin ChIP-seq peaks[20] and RNAPII-mediated ChIA-PET signal[21] compared to other enhancers (Supplementary Fig. 26d). These results indicate that u-enhancers comprise a small but distinct subset of enhancers, which probably has specific regulatory functions used by virtually every human cell.

Linking enhancer usage with TSS expression

A major challenge is to link enhancers to their target genes[21,22]. Uniquely, FANTOM5 CAGE allows for direct comparison between transcriptional activity of the enhancer and of putative target gene TSSs across a diverse set of human cells. Based on pairwise expression correlation, nearly half (40%) of the inferred TSS-associated enhancers (Methods) were linked with the nearest TSS, and 64% of enhancers have at least one correlated TSS within 500 kilobases. Several associations (10,260; 15.3%) are supported by ChIA-PET (RNAPII-mediated) interaction data[21], and the supported fraction increases with the correlation threshold (Supplementary Fig. 29a). The fraction of supported associations is 4.8-fold higher than that of associations predicted from DNase I hypersensitivity correlations[10] (20.6% versus 4.3%, at the same correlation threshold), indicating that transcription is a better predictor of regulatory targets than chromatin accessibility. Conserved sequence motifs and ChIP-seq peaks also co-occurred significantly in associated enhancer-promoter pairs (Benjamini–Hochberg false discovery rate (FDR) < 0.05, binomial test), suggesting an additive or synergistic cooperation between enhancers and promoters at RNAPII foci.

On average, a RefSeq TSS was associated with 4.9 enhancers and an enhancer with 2.4 TSSs and we observed different regulatory architectures around genes (Supplementary Fig. 30). For example, at the beta-globin locus the CAGE expression patterns of four locus control region hypersensitive sites are highly correlated (Pearson's *r* between 0.88 and 0.98) with the expression of known target genes[23,24] *HBG2* and *HBD*, and to some extent *HBG1*.

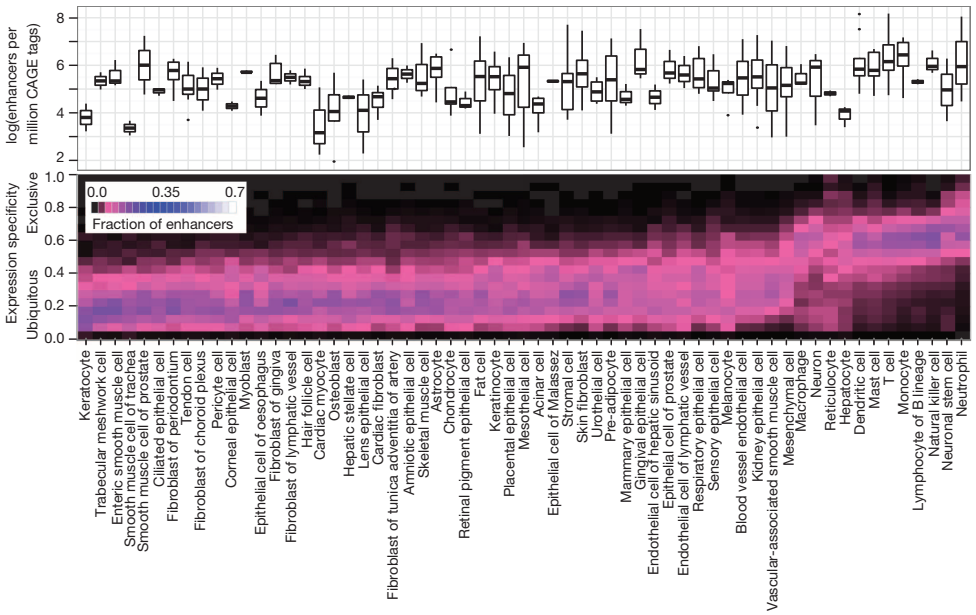


Figure 5 | Enhancer usage and specificity in groups of cells. The upper panel gives the number of detected enhancers per million CAGE tags within each group (facet) of related cell type libraries. The expression specificity of the enhancers is shown as a heat map in the panel below. Colours show the fraction of expressed enhancers in each facet (columns) that are in each specificity range (rows). For corresponding plots on organ/tissue facets and genes, see Supplementary Fig. 21.

These observations call for computational models of enhancer regulation, in which multiple enhancers may work in concert to enhance the expression of a gene. To this end, we focused on 2,206 RefSeq TSSs for which the joint expression of nearby enhancers (the closest ten enhancers within 500 kb) is highly predictive of the gene expression. Model shrinkage showed that in most cases, only one to three enhancers are necessary to explain the expression variance observed in the linked gene, and generally proximal enhancers are more predictive than distal ones (Fig. 6a, Supplementary Fig. 29b–d and Supplementary Text). One hypothesis explaining the function of multiple enhancers driving the

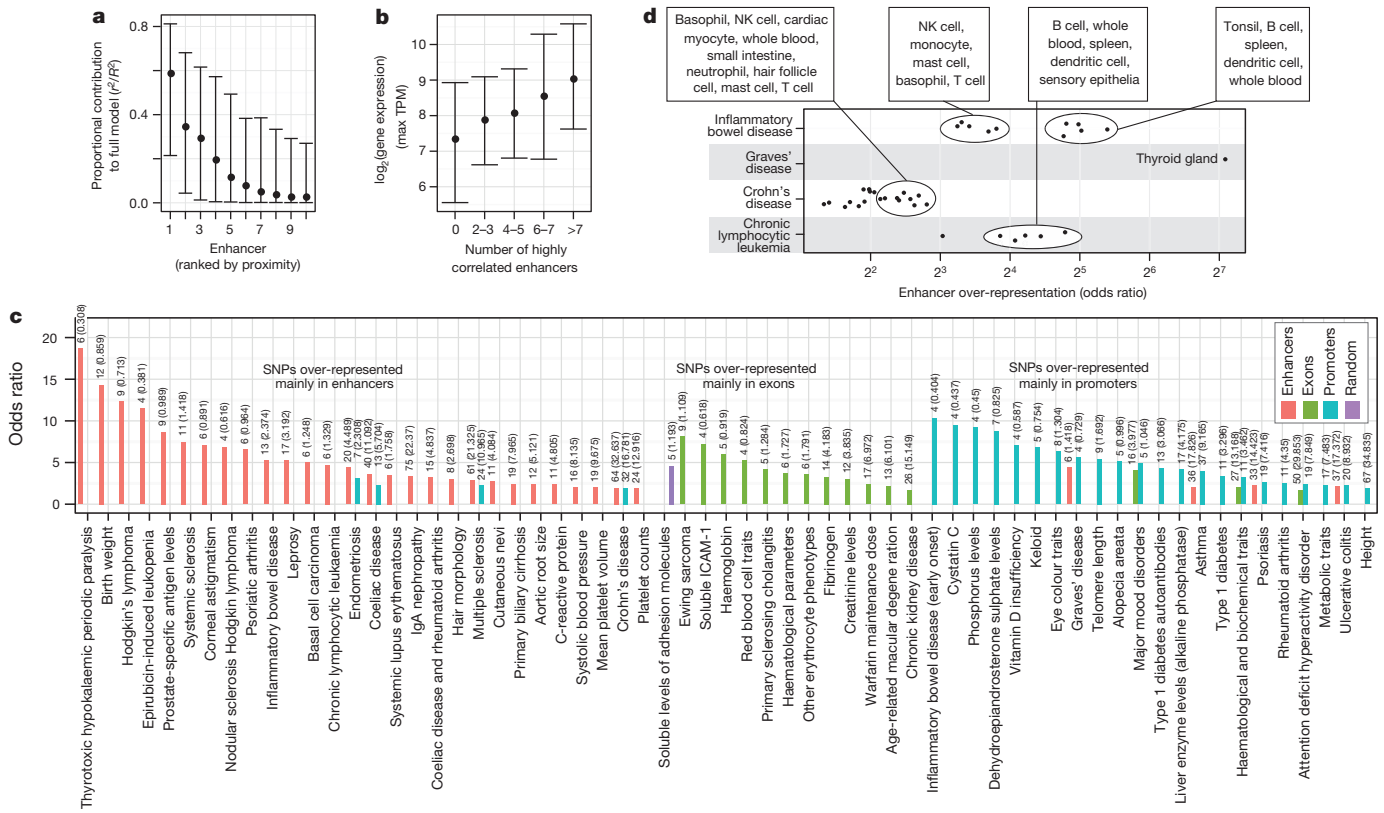


Figure 6 | Linking enhancers to TSSs and disease-associated SNPs. **a**, The proportional contribution (see Methods) of the 10 most proximal enhancers within 500 kb of a TSS in a model explaining gene expression variance (vertical axis) as a function of enhancer expression. *x* axis indicates the position of the enhancer relative to the TSS: 1 the closest, etc. Bars indicate interquartile ranges and dots medians. **b**, Relationship between the number of highly correlated ('redundant') enhancers per locus (horizontal axis) and the maximal expression (TPM) of the associated TSS in the same model over all CAGE libraries (vertical axis). Error bars as in **a**. **c**, GWAS SNP sets preferentially overrepresented within enhancers, exons and mRNA promoters. Observed and expected overlaps are shown above bars. The vertical axis gives enrichment odds ratios. The horizontal axis shows GWAS traits or diseases. **d**, Diseases with GWAS-associated SNPs over-represented in enhancers of certain expression facets. The horizontal axis gives the odds ratio as in panel **c**, broken up by expression facets: each point represents the odds ratio of GWAS SNP enrichment for a disease (vertical axis) in a specific expression facet. Summary annotations of point clouds are shown. See also Supplementary Fig. 31.

same expression pattern is that they might confer higher transcriptional output of a gene[25,26]. Indeed, the number of highly correlated (redundant) enhancers close to TSSs (Supplementary Methods) increased with the observed maximal TSS expression over all libraries (Fig. 6b), implying that these enhancers are redundant in terms of transcription patterns but additive in terms of expression strength. Expression redundancy is also common in genomic clusters of closely spaced enhancers (24% of 815 identified genomic clusters, Supplementary Table 15). These are associated with TSSs of genes involved in immune and defence responses and, as suggested by a previous study[27], have a higher expression than other enhancer-associated genes (eightfold increase on average).

Disease–associated SNPs are enriched in enhancers

Many disease-associated SNPs are located outside of protein-coding exons and a large proportion of human genes display expression polymorphism[28]. Using the NHGRI genome-wide association studies (GWAS) catalogue[29] and extending the compilation of lead SNPs with proxy SNPs in strong linkage disequilibrium (similar to refs 30, 31), we identified diseases/traits whose associated SNPs overlapped enhancers, promoters, exons and random regions significantly more than expected by chance (Fisher's exact test $P < 0.01$, Supplementary Table 16). Disease-associated SNPs were over-represented in regulatory regions to a greater extent than in exons (Fig. 6c). For many traits where enriched disease-associated SNPs were within enhancers, enhancer activity was detected in pathologically relevant cell types (Fig. 6d and Supplementary Figs 31 and 32). Examples include Graves' disease-associated SNPs enriched in enhancers that are expressed predominantly in thyroid tissue, and similarly lymphocytes for chronic lymphocytic leukaemia. As a proof of concept, we validated the impact of two disease-associated regulatory SNPs within enhancers (Supplementary Fig. 33).

Conclusions

The data presented here demonstrate that bidirectional capped RNAs, as measured by CAGE, are robust predictors of enhancer activity in a cell. Transcription is only measured at a fraction of chromatin-defined enhancers and few untranscribed enhancers show potential enhancer activity. This implies that many chromatin-defined enhancers are not regulatory active in that particular cellular state, but may be active in other cells of the same lineage[32] or are pre-marked for fast regulatory activity upon stimulation[33]. Of course, given the relative instability of enhancer RNAs some chromatin-defined sites may be active but fall below the limits of detection of CAGE.

Our results show that position-specific sequence signals upstream of the transcription initiation sites and the production of small, uncapped RNAs immediately downstream is present at both enhancers and mRNA promoters, suggesting similar mechanisms of initiation. Previous studies (for example refs 10, 34, 35) suggested that promoters and enhancers differ in motif composition. This view is not supported by the larger FANTOM5 data set. Instead, the differences reflect the local G+C content because transcribed enhancers tend to harbour low G+C content motifs like non-CGI promoters. Features distinguishing enhancers from mRNA promoters are (1) enhancer RNAs are exosome-sensitive regardless of direction whereas (sense) mRNAs have a longer half-life than their antisense counterpart; (2) enhancer RNAs are short, unspliced, nuclear and non-polyadenylated and (3) enhancers have downstream polyadenylation and 5′ splice motif frequencies at genomic background level similar to antisense PROMPTs, whereas mRNAs are depleted of termination signals and enriched for 5′ splice sites[11,12].

The collection of active enhancers presented here provides a resource that complements the activity of the ENCODE consortium[7] across a much greater diversity of tissues and cellular states. It has clear applications in human genetics, to narrow the search windows for functional association, and for the definition of regulatory networks that underpin the processes of cellular differentiation and organogenesis in human development.

METHODS SUMMARY

Single-molecule HeliScopeCAGE data was generated as described elsewhere[6]. Sequencing and processing of ribosomal RNA-depleted RNAs, short RNAs and H3K27ac or H3K4me1 ChIPs as well as the processing of publicly available DNase-seq data are described in the Methods.

Putative enhancers were identified from bidirectionally transcribed loci having divergent CAGE tag clusters separated by at most 400 bp (described in Supplementary Fig. 6a). We required loci to be divergently transcribed in at least one FANTOM5 sample, defined by CAGE tag 5′ ends within 200 bp divergent strand-specific windows immediately flanking the loci midpoints. The expression of each enhancer in each FANTOM5 sample was quantified as the normalized sum of strand-specific sums of CAGE tags in these windows. A sample-set wide directionality score, D, for each locus over aggregated normalized reverse, R, and forward, F, strand window-expression values across all samples, $D = (F - R)/(F + R)$, were then used to filter putative enhancers to have low, non-promoter-like, directionality scores ($|D| < 0.8$). Further filtering ensured enhancers to be located distant to TSSs and exons of protein- and noncoding genes.

Motif enrichment analyses were done using HOMER[36]. Regulatory targets of enhancers were predicted by correlation tests using the sample-set wide expression profiles of all enhancer-promoter pairs within 500 kb. The regulatory effects of multiple enhancers were modelled using linear regression followed by lasso-based model-shrinkage[37].

Enhancer activity was tested *in vivo* in zebrafish embryos using Tol2-mediated transgenesis[38]. Expression patterns were documented at 48 h post fertilization using >200 eggs per construct. Large-scale *in vitro* validations on randomly selected enhancers were performed using firefly/*Renilla* luciferase reporter plasmids with enhancer sequences cloned upstream of an EF1α basal promoter separated by a synthetic polyA signal/transcriptional pause site in a modified pGL4.10 (Promega) vector (Supplementary Fig. 9d). Full details are 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 4 January; accepted 16 October 2013.

1. Bulger, M. & Groudine, M. Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev. Biol.* **339,** 250–257 (2010).
2. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Rev. Genet.* **13,** 233–245 (2012).
3. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308 (1981).
4. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
5. Kodzius, R. *et al.* CAGE: cap analysis of gene expression. *Nature Methods* **3,** 211–222 (2006).
6. The FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature* http://dx.doi.org/10.1038/nature13182 (this issue).
7. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
8. Kheradpour, P. *et al.* Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* **23,** 800–811 (2013).
9. Fort, A. *et al.* Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genet.* (in the press).
10. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
11. Ntini, E. *et al.* Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nature Struct. Mol. Biol.* **20,** 923–928 (2013).
12. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499,** 360–363 (2013).
13. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).
14. Kowalczyk, M. S. *et al.* Intragenic enhancers act as alternative promoters. *Mol. Cell* **45,** 447–458 (2012).
15. Valen, E. *et al.* Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nature Struct. Mol. Biol.* **18,** 1075–1082 (2011).
16. Taft, R. J. *et al.* Tiny RNAs associated with transcription start sites in animals. *Nature Genet.* **41,** 572–578 (2009).
17. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).
18. Rönnerblad, M. *et al.* Analysis of the DNA methylome and transcriptome in granulopoiesis reveal timed changes and dynamic enhancer methylation. *Blood* http://dx.doi.org/10.1182/blood-2013-02-482893 (in the press).

19. Biddie, S. C. *et al.* Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol. Cell* **43,** 145–155 (2011).
20. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20,** 578–588 (2010).
21. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
22. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. & Zhao, K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* **22,** 490–503 (2012).
23. Fraser, P., Pruzina, S., Antoniou, M. & Grosveld, F. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev.* **7,** 106–113 (1993).
24. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16,** 1299–1309 (2006).
25. Barolo, S. Shadow enhancers: frequently asked questions about distributed *cis*-regulatory information and enhancer redundancy. *Bioessays* **34,** 135–141 (2012).
26. Schaffner, G., Schirm, S., Müller-Baden, B., Weber, F. & Schaffner, W. Redundancy of information in enhancers as a principle of mammalian transcription control. *J. Mol. Biol.* **201,** 81–90 (1988).
27. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
28. Göring, H. H. H. *et al.* Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nature Genet.* **39,** 1208–1216 (2007).
29. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
30. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40,** D930–D934 (2012).
31. Maurano, M. T., Wang, H., Kutyavin, T. & Stamatoyannopoulos, J. A. Widespread site-dependent buffering of human regulatory polymorphism. *PLoS Genet.* **8,** e1002599 (2012).
32. Mercer, E. M. *et al.* Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors. *Immunity* **35,** 413–425 (2011).
33. Ostuni, R. *et al.* Latent enhancers activated by stimulation in differentiated cells. *Cell* **152,** 157–171 (2013).
34. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–283 (2011).
35. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).
36. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).
37. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33,** 1–22 (2010).
38. Gehrig, J. *et al.* Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. *Nature Methods* **6,** 911–916 (2009).

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

Acknowledgements FANTOM5 was made possible by a Research Grant for RIKEN Omics Science Center from MEXT to Y.H. and a Grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT, Japan, to Y.H. The A.S. group was supported by funds from the European Research Council FP7/2007-2013/ERC no. 204135, the Novo Nordisk and Lundbeck foundations. Work in the M.R. group was funded by grants from the Deutsche Forschungsgemeinschaft (RE 1310/7, 11, 13) and Rudolf Bartling Stiftung. F.M. and I.M.E. were supported by "BOLD" Marie Curie ITN and "ZF- Health" Integrated project of the European Commission. We thank S. Noma, M. Sakai and H. Tarui for RNA-seq and sRNA-seq preparation, RIKEN GeNAS for generation and sequencing of the Heliscope CAGE libraries, Illumina RNA-seq and sRNA-seq, the Copenhagen National High-throughput DNA Sequencing Center for Illumina CAGE-seq, M. Edinger, P. Hoffmann and R. Eder for cell sorting, A. Albrechtsen, I. Moltke, W. Wasserman for advice, and the Netherlands Brain Bank for post-mortem human brain material.

Author Contributions R.A., I.H., E.A., E.V., K.L., Y.C., B.L., X.Z., M.J., H.K., T.F.M., T.L., N.B., O.R., A.M.B. , J.K.B, C.J.M, N.R., F.O.B., M.R., A.S. made the computational analysis. J.B., M.B., T.L., H.K., N.K., J.K., H.S., M.I., C.O.D, A.R.R.F., P.C., Y.H. prepared and pre-processed CAGE and/or RNA-seq libraries. E.N., P.R.A., T.H.J., J.B., M.B. made the knockdown experiments followed by CAGE. C.G., C.S., L.S., J.R., D.G., M.R. made the blood cell ChIP experiments, methylation assays and *in vitro* blood cell validations. T.S., C.G., Y.I., Y.S., E.F., S.M., Y.N., A.R.R.F., P.C. and H.S. made the HeLa/HepG2 *in vitro* validations. I.M.E., R.A., A.S., F.M. designed and carried out zebrafish *in vivo* tests. R.A., C.G., I.H., C.S., E.A., E.V., F.M., I.M.E., P.C., A.R.R.F, M.B., J.B., A.L., C.D., D.A.H., P.H., M.R., A.S. interpreted results. R.A., C.G., I.H., E.V., I.M.E., J.B., F.M., D.A.H., M.R., A.S. wrote the paper with input from all authors. M.R. and A.S. coordinated and supervised the project.

Author Information The FANTOM5 atlas is accessible from http://fantom.gsc.riken.jp/5. FANTOM5 CAGE, RNA-seq and sRNA data have been deposited in DDBJ/EMBL/GenBank (accession codes DRA000991, DRA001101)**.** Genome browser tracks for enhancers with user-definable expression specificity-constraints can be generated at http://enhancer.binf.ku.dk. Here, processed enhancer expression data, predefined enhancer tracks and motif finding results are also deposited. Blood-cell ChIP-seq data and CAGE data on exosome-depleted HeLa cells have been deposited in the NCBI GEO database (accession codes GSE40668, GSE49834). 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.R.R.F. (alistair.forrest@gmail.com), P.C. (carninci@riken.jp), M.R. (michael.rehli@ukr.de) or A.S. (albin@binf.ku.dk).

ARTICLE

doi:10.1038/nature13182

A promoter-level mammalian expression atlas

The FANTOM Consortium and the RIKEN PMI and CLST (DGT)*

Regulated transcription controls the diversity, developmental pathways and spatial organization of the hundreds of cell types that make up a mammal. Using single-molecule cDNA sequencing, we mapped transcription start sites (TSSs) and their usage in human and mouse primary cells, cell lines and tissues to produce a comprehensive overview of mammalian gene expression across the human body. We find that few genes are truly 'housekeeping', whereas many mammalian promoters are composite entities composed of several closely separated TSSs, with independent cell-type-specific expression profiles. TSSs specific to different cell types evolve at different rates, whereas promoters of broadly expressed genes are the most conserved. Promoter-based expression analysis reveals key transcription factors defining cell states and links them to binding-site motifs. The functions of identified novel transcripts can be predicted by coexpression and sample ontology enrichment analyses. The functional annotation of the mammalian genome 5 (FANTOM5) project provides comprehensive expression profiles and functional annotation of mammalian cell-type-specific transcriptomes with wide applications in biomedical research.

The mammalian genome encodes the instructions to specify development from the zygote through gastrulation, implantation and generation of the full set of organs necessary to become an adult, to respond to environmental influences, and eventually to reproduce. Although the genome information is the same in almost all cells of an individual, at least 400 distinct cell types[1] have their own regulatory repertoire of active and inactive genes. Each cell type responds acutely to alterations in its environment with changes in gene expression, and interacts with other cells to generate complex activities such as movement, vision, memory and immune response.

Identities of cell types are determined by transcriptional cascades that start initially in the fertilised egg. In each cell lineage, specific sets of transcription factors are induced or repressed. These factors together provide proximal and distal regulatory inputs that are integrated at transcription start sites (TSSs) to control the transcription of target genes. Most genes have more than one TSS, and the regulatory inputs that determine TSS choice and activity are diverse and complex (reviewed in ref. 2).

Unbiased annotation of the regulation, expression and function of mammalian genes requires systematic sampling of the distinct mammalian cell types and methods that can identify the set of TSSs and transcription factors that regulate their utilization. To this end, the FANTOM5 project has performed cap analysis of gene expression (CAGE)[3] across 975 human and 399 mouse samples, including primary cells, tissues and cancer cell lines, using single-molecule sequencing[3] (Fig. 1; see the full sample list in Supplementary Table 1).

CAGE libraries were sequenced to a median depth of 4 million mapped tags per sample (Supplementary Methods) to produce a unique gene expression profile, focused specifically on promoter utilization. CAGE has advantages over RNA-seq or microarrays for this purpose, because it permits separate analysis of multiple promoters linked to the same gene[13]. Moreover, we show in an accompanying manuscript[4] that the data can be used to locate active enhancers, and to provide numerous insights into cell-type-specific transcriptional regulatory networks (see the FANTOM5 website http://fantom.gsc.riken.jp/5). The data extend and complement the recently published ENCODE[5] data, and microarray-based gene expression atlases[6] to provide a major resource for functional genome annotation and for understanding the transcriptional networks underpinning mammalian cellular differentiation.

The FANTOM5 promoter atlas

Single molecule CAGE profiles were generated across a collection of 573 human primary cell samples (~ 3 donors for most cell types) and 128 mouse primary cell samples, covering most mammalian cell steady states. This data set is complemented with profiles of 250 different cancer cell lines (all available through public repositories and representing 154 distinct cancer subtypes), 152 human post-mortem tissues and 271 mouse developmental tissue samples (Fig. 1a; see the full sample list in Supplementary Table 1). To facilitate data mining all samples were annotated using structured ontologies (Cell Ontology[7], Uberon[8], Disease Ontology[9]). The results of all analyses are summarized in the FANTOM5 online resource (http://fantom.gsc.riken.jp/5). We also developed two specialized tools for exploration of the data. ZENBU, based on the genome browser concept, allows users to interactively explore the relationship between genomic distribution of CAGE tags and expression profiles[10]. SSTAR, an interconnected semantic tool, allows users to explore the relationships between genes, promoters, samples, transcription factors, transcription factor binding sites and coexpressed sets of promoters. These and other ways to access the data are described in more detail in Supplementary Note 1.

CAGE peak identification and thresholding

To identify CAGE peaks across the genome we developed decomposition-based peak identification (DPI; described in Supplementary Methods; Extended Data Fig. 1). This method first clusters CAGE tags based on proximity. For clusters wider than 49 base pairs (bp) it attempts to decompose the signal into non-overlapping sub-regions with different expression profiles using independent component analysis[11]. Sample- and genome-wide, DPI identified 3,492,729 peaks in human and 2,088,255 peaks in mouse. To minimize the fraction of peaks[3] that map to internal exons (which could exist due to post-transcriptional cleavage and recapping of RNAs[12]), and enrich for TSSs, we applied tag evidence thresholds

*Lists of participants and their affiliations appear at the end of the paper.

Figure 1 | Promoter discovery and definition in FANTOM5. **a**, Samples profiled in FANTOM5. **b**, Reproducible cell-type-specific CAGE patterns observed for the 266 base CpG island associated *B4GALT1* locus transcription initiation region hg19:chr9:33167138..33167403. CAGE profiles for $CD4^+$ T cells (blue), $CD14^+$ monocytes (gold), aortic smooth muscle cells (green) and the adrenal cortex adenocarcinoma cell line SW-13 (red) are shown. A combined pooled profile showing TSS distribution across the entire human collection is shown in black. Values on the *y* axis correspond to maximum normalized TPM for a single base in each track. **c**, Decomposition-based peak identification (DPI) finds 6 differentially used peaks within this composite transcription initiation region (note: peaks are labelled from p1@B4GALT1 with most tag support through to p7@B4GALT1 with the least tag support; p4@B4GALT1 is not shown and is in the 3′ UTR of the locus at position hg19::chr9:33111241..33111254−). Note in particular one large broad region on the left used in all samples and a sharp peak to the right, preferentially used in the aortic smooth muscle cells. **d**, Venn diagram showing DPI defined peaks expressed at ≥ 10 TPM in primary cells (red), tissues (blue) and cell lines (green). **e**, Fraction of unannotated peaks observed in subsets of **d**. P, primary cells, T, tissues, C, cell lines, PT, TC, PC and PTC correspond to peaks found in multiple sample types, for example, PT, found in primary cells and tissue samples.

to define robust and permissive subsets (described in more detail in Supplementary Methods and summarized in Table 1). Specifically the robust threshold, which is used for most of the analyses presented here, enriched for peaks at known 5′ ends compared to known internal exons by twofold (that is, two-thirds of the peaks hitting known full-length transcript models hit the 5′ end). A flow diagram showing the relationship between samples, peaks, thresholding and subsets used in each analysis is provided in the Supplementary Figure 1. Supporting evidence that the peaks are genuine TSSs, based upon support from expressed sequence tags (ESTs), histone H3 lysine 4 trimethylation (H3K4Me3) marks and DNase hypersensitive sites is provided in Supplementary Note 2.

Figure 1b illustrates the 266 bp spanning transcription initiation region of *B4GALT1,* where 6 independent robust peaks were identified by DPI, each with a unique regulatory pattern (Fig. 1c). A total of 58% of human and 56% of mouse robust peaks occur in such composite transcription initiation regions, defined as clusters of robust peaks within 100 bases of each other. More than half of these contain peaks with statistically significant differences in expression profiles (63% of human and 54% of mouse composite transcription initiation regions; likelihood ratio test, false discovery rate (FDR) <1%, Extended Data Fig. 1d). Supplementary Tables 2 and 3 summarize public domain EST evidence that these independent peaks contained within composite transcription initiation regions give rise to long RNAs.

Known gene coverage in FANTOM5

To provide annotation of the CAGE peaks, the distance between individual peaks and the 5′ ends of known full-length transcripts was determined and then peaks within 500 bases of the 5′ end of known transcript models were assigned to that gene (see Supplementary Methods, Table 1). To provide names for each TSS region, peaks identified at the permissive threshold were ranked by the total number of tags supporting each and then sequentially numbered (for example, p1@GFAP corresponds to the promoter of *GFAP* which has the highest tag support). From these annotations, TSS for 91% of human protein coding genes (as defined by the HUGO Gene Nomenclature Committee) were supported by robust CAGE peaks, and 94% at the permissive threshold (Supplementary Note 3). The atlas also detected signals from the promoters of short RNA primary transcripts, and long non-coding RNAs. In comparison to the previous FANTOM3 and 4 projects, FANTOM5 measured expression at an additional 4,721 human and 5,127 mouse RefSeq genes. The inclusion of primary cells, cell lines and tissues in the atlas provided greater coverage than any of the sample types alone (Fig. 1d) and the primary cell samples in particular were a rich source of unannotated peaks (Fig. 1e).

Mammalian promoter architectures

Mammalian promoters can be classified as broad or sharp types, based upon local spread of TSSs along the genome[13]. The FANTOM5 data

Table 1 | Summary of peaks, coverage and genes hit in FANTOM5

	Human							Mouse						
	Peaks	Stranded genome coverage (bp)		Number of aligned reads		Genes hit	Peaks per gene	Peaks	Stranded genome coverage (bp)		Number of aligned reads		Genes hit	Peaks per gene
The whole genome	—	6.2×109	100%	4.5×10^9	100%	—	—	—	5.3×10^9	100%	1.9×10^9	100%	—	—
'Permissive' CAGE peaks	1,048,124	1.4×10^7	0.22%	3.6×10^9	80%	20,808	—	652,860	8.4×10^6	0.16%	1.5×10^9	79%	20,480	—
(A) Within 500 bp of annotated 5′	245,514	4.3×10^6	0.07%	3.0×10^9	68%	20,808	11.8	146,185	2.5×10^6	0.05%	1.3×10^9	69%	20,480	7.1
(B) TSS classifier positive	217,572	4.0×10^6	0.06%	2.9×10^9	64%	18,503	—	129,466	2.4×10^6	0.05%	1.0×10^9	52%	17,088	—
(A or B) Likely TSS	308,214	5.3×10^6	0.09%	3.2×10^9	72%	20,808	—	173,564	3.0×10^6	0.06%	1.4×10^9	70%	20,480	—
'Robust' CAGE peaks	184,827	3.9×10^6	0.06%	3.5×10^9	77%	18,961	—	116,277	2.5×10^6	0.05%	1.4×10^9	75%	19,001	—
(A) Within 500bp of annotated 5′	82,150	2.2×10^6	0.04%	3.0×10^9	66%	18,961	4.3	61,134	1.6×10^6	0.03%	1.3×10^9	68%	19,001	3.2
(B) TSS classifier positive	76,445	2.1×10^6	0.03%	2.9×10^9	63%	17,285	—	51,611	1.4×10^6	0.03%	9.9×10^8	51%	16,028	—
(A or B) Likely TSS	92,783	2.4×10^6	0.04%	3.2×10^9	70%	18,961	—	77674	1.7×10^6	0.03%	1.3×10^9	69%	19,001	—
Cross-species projected robust peaks	70,351	1.6×10^6	0.03%	—	—	—	—	105,157	2.4×10^6	0.04%	—	—	—	—
'Homologous' robust peaks	34,041	1.0×10^6	0.02%	—	—	—	—	42,423	1.3×10^6	0.02%	—	—	—	—

confirmed this general observation (Extended Data Fig. 2), however, for the first time the greater depth of sequencing enabled identification of the preferred TSS within broad promoters. Taking each library in turn, using the location of the dominant TSS (that is, the TSS with the highest number of tags), we searched for phased WW dinucleotides (AA/AT/TA/TT) associated with nucleosome location[14] (Extended Data Fig. 2). Remarkably, on a genome-wide scale, there was a periodic spacing of WW motifs with a 10.5 bp repeat downstream of the dominant TSS, exactly as shown previously for well-phased H2A.Z nucleosomes[14] (Extended Data Fig. 2d). The precise phasing was supported further by the pattern of H2A.Z and H3K4me3 chromatin immunoprecipitation sequencing (ChIP-seq) signal seen around TSS in $CD14^+$ monocytes and frontal lobe respectively (Extended Data Fig. 2e, f). This observation indicates that the positioned nucleosome is a key indicator of start site preference in broad promoters.

Expression levels and tissue specificity

The raw tag counts under the DPI peak coordinates were used to generate an expression table across the entire collection. Normalized tags per million (TPM) were then calculated using the relative log expression (RLE) method in edgeR[15]. Almost all peaks (96%) were reproducibly detected above 1 TPM in at least two samples, but most were detected in less than half the samples. Examining the distribution of expression level and breadth across the collection, we classified the 185K robust human peak expression profiles as non-ubiquitous (cell-type-restricted, 80%), ubiquitous-uniform ('housekeeping', 6%) or ubiquitous-non-uniform (14%) (Fig. 2a, b). We define ubiquitous as detected in more than 50% of samples (median >0.2 TPM) and uniform as a less than tenfold difference between maximum and median expression. Estimation using the smaller mouse expression data set or human primary cell, cell line or tissue data subsets resulted in different fractions, yet in all cases ubiquitous-uniform expression profiles were in the minority (Extended Data Fig. 3a–e). Alternative measures such as richness index and Shannon entropy confirm that only a minor fraction of transcripts can be considered as genuine housekeeping genes with broad and uniform expression (Supplementary Note 4 and Supplementary Table 4 for a list of housekeeping genes). In addition many of the 1,225 known genes that were missed in the collection are known to be specifically expressed in cell types that are not easily procured; indicating that even more of the mammalian transcriptome has a cell-type-restricted expression

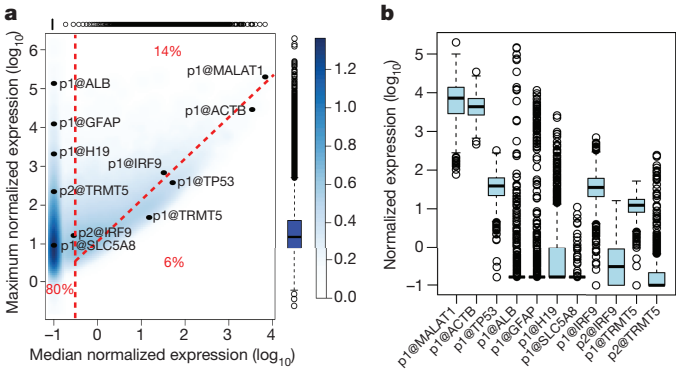


Figure 2 | Cell-type-restricted and housekeeping transcripts encoded in the mammalian genome. a, Density plot summarizing the distribution of relative log expression (RLE) normalized maximum and median TPM expression values for the 185K robustly detected human peaks identified by FANTOM5 (colour bar on right indicates relative density). Box and whiskers plots above and to right show distribution of median and maximum values in the data set (box shows the interquartile range). Promoters of named genes are highlighted to show extremes of expression level and expression breadth, note the alternative promoters of *IRF9* and *TRMT5* have different maximums and breadths of expression (see Extended Data Fig. 10). Fraction on left of the red vertical dashed line corresponds to peaks detected in less than 50% of samples with non-ubiquitous (cell-type-restricted) expression patterns (median <0.2 TPM). Fraction below the red diagonal dashed line corresponds to ubiquitous-uniform (housekeeping) expression profiles (maximum < 10× median). Fraction above diagonal and to the right of the vertical dashed lines corresponds to ubiquitous-non-uniform expression profiles (maximum > 10× median). **b**, Box and whisker plots showing the distribution of expression levels for the same peaks as in **a** across the 889 samples (box shows the interquartile range).

pattern (Supplementary Note 3). In overview, the data confirm the argument that most genes are regulated in a tissue-dependent manner[16]. According to Gene Ontology enrichment analysis[17] of genes within each of the three classes (Supplementary Table 5), the non-ubiquitous genes were enriched for proteins involved in cell–cell signalling, plasma membrane receptors, cell adhesion molecules and signal transduction, whereas genes in the housekeeping set were enriched for components of the ribonucleoprotein complex and RNA processing. The ubiquitous-non-uniform set was enriched for cell cycle genes, with 204 of the 268 human genes annotated with the 'mitotic cell cycle' term, a reflection of the fact that the fraction of actively proliferating cells inevitably varies greatly across the collection.

Finally, of the 104,859 peaks expressed at 10 TPM (~3 copies per cell[18]) or greater, an average primary cell sample expressed a median of 8,757 including peaks for 430 transcription factor mRNAs (Extended Data Fig. 3f, g).

Promoter conservation between human and mouse

Regulatory regions such as transcription factor binding sites are often, but not always, located in conserved and orthologous regions[19]. Overall human TSSs were significantly enriched in evolutionarily conserved regions compared to the genome-wide null expectation, with 38% overlapping previously defined mammalian constrained elements (Fisher's exact test, odds ratio 10.2, P value $< 2.2 \times 10^{-16}$; see Supplementary Methods). Despite this general level of conservation, there is evidence of extensive evolutionary remodelling of transcription initiation. For example, 43% (79,670 out of 184,476) of human TSSs could not be aligned to the mouse genome, and 39% (45,926 out of 116,277) of mouse TSSs could not be aligned to the human genome (Supplementary Methods). Alignment between species decayed as a function of neutral sequence divergence (Fig. 3). Housekeeping TSSs showed highest TSS conservation, whereas the TSSs of non-coding RNAs were less conserved than those of protein-coding TSSs. Indeed, the alignment of promoters of broadly expressed non-coding transcripts was not greatly different from randomly selected genomic sites (Fig. 3a). However, it is important to note that the random permutations inevitably overlap constrained elements, so cannot be considered representative of neutral evolution.

TSSs that were highly-restricted or biased in their expression to a single cell type or tissue were more likely to be gained or lost through evolution (Fig. 3a). TSSs preferentially expressed in fibroblasts, chondrocytes and pre-adipocytes were among the most conserved, whereas those enriched in T-cells, macrophages, dendritic cells, whole blood and endothelial cells were the most likely to be gained or lost (Fig. 3b). This suggests a more rapidly evolving immune system. It also suggests contributions of relaxed constraint and positive selection to the remodelling of transcription initiation through the insertion and deletion of promoter sequences.

To enable comparative analysis, we projected the expression patterns from one species to the other (Extended Data Fig. 4) and provide the peak position and orthologous expression profile through a cross-species track in ZENBU[10]. Only 54% and 61% of human and mouse conserved TSSs (of protein coding genes) had an orthologous peak in the other species. This increased to 61% and 63% respectively for TSSs from well matched samples (for example, human and mouse hepatocytes), however, surprisingly, almost 40% of conserved TSS do not appear to be used even in the matched cells (Supplementary Table 6).

Features of cell-type-specific promoters

Carrying out a systematic *de novo* motif discovery analysis in cell-type-specific promoters, recovered motifs similar to the binding motifs of transcription factors known to be relevant to the corresponding cellular states (Extended Data Fig. 5a–c and described in Supplementary Note 5). Examining general promoter features many CpG island (CGI) based promoters (54%) and most non-CGI-non-TATA promoters (92%) had non-ubiquitous expression profiles (Extended Data Fig. 3k–n). Although CGI promoters are generally associated with housekeeping

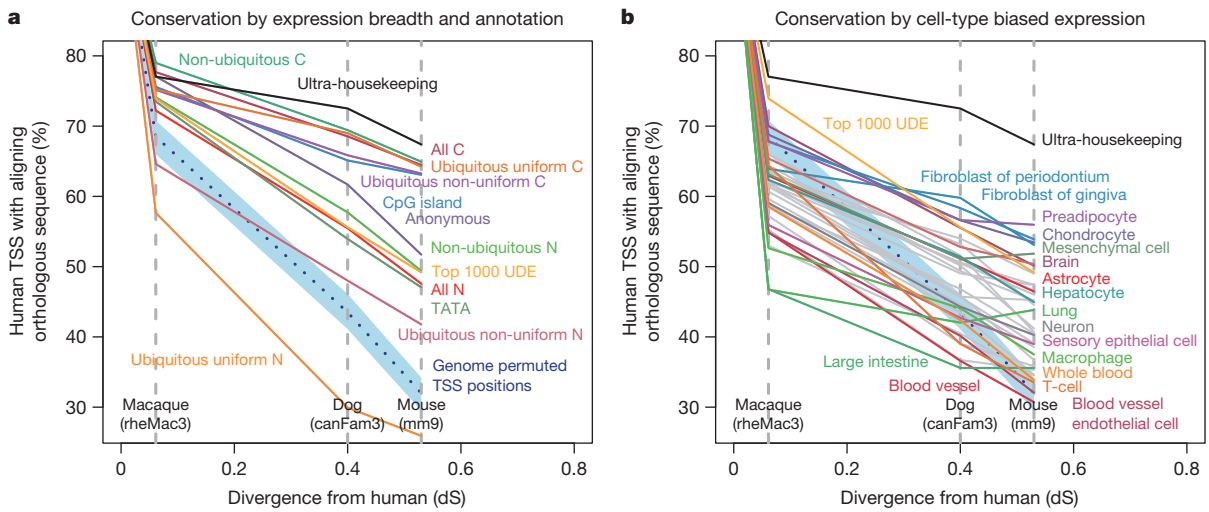


Figure 3 | TSS conservation as a function of expression properties and functional annotation. **a**, **b**, Human robust TSS coordinates were projected through EPO12 whole genome multiple sequence alignments (Supplementary Methods). The y-axis values show the fraction of human TSSs that align to an orthologous position in the indicated species. The x axis shows the relative divergence of macaque, dog and mouse genomes as the substitution rate at fourfold degenerate sites in protein coding sequence. The TSS locations were genome permuted (Supplementary Methods) and then projected through EPO12 alignments to give the null expectation (dashed blue line). The 95% confidence intervals of 1,000 samples of 1,000 TSS are shown (blue shading). **a**, TSS mapped to the 5′ ends of protein coding and non-coding transcripts are labelled (C and N, respectively), those that do not map to a known transcript 5′ end are shown as the 'anonymous' category. With the exception of anonymous, all robust TSSs represented in both panels are associated with the 5′ ends of previously annotated transcripts. Non-ubiquitous (cell-type-restricted), ubiquitous-uniform (housekeeping) and non-uniform-ubiquitous were defined as in Fig. 2. Ultra-housekeeping TSSs were defined as those with less than fivefold difference between maximum and median. The category top 1000 UDE represents the 1,000 ubiquitous TSSs that are most differentially expressed[4]. There are 1,016 ultra-housekeeping TSSs, 276 ubiquitous-uniform non-coding TSSs and all other categories contain over 2,000 TSSs. **b**, Same axes as panel **a** showing TSSs with expression that is biased towards a single expression facet (larger mutually exclusive grouping of the primary cell and tissue samples based on the sample ontologies CO and UBERON, defined in ref. 4). Only expression facets with greater than 250 enriched TSSs are shown. For clarity, only a subset of expression facets are coloured and labelled.

genes, we observed a subset with highly cell-type-restricted expression profiles (right tail of Extended Data Fig. 6a). Examining CGI and non-CGI promoters separately we find that cell-type-specific promoters of both classes were enriched for binding of cell-type-specific transcription factors (evidenced by over-representation of motifs and bound sites in public ChIP-seq data sets). For the human hepatocellular carcinoma cell line HepG2 we observed enrichment of liver-specific transcription factors (HNF4, FOXA2, and TCF7L2) at both CGI and non-CGI HepG2 specific promoters (Extended Data Fig. 6b, c; similar examples are shown in Extended Data Figs 5d and 7). As noted in the accompanying analysis[4], both cell-type-specific CGI and non-CGI promoters tend to have proximal high-specificity enhancers (Extended Data Fig. 6d). This indicates that specific expression at CGI promoters uses the same type of signals as non-CGI promoters: proximal transcription factor motifs and high-specificity enhancers.

Of note, a small number of highly abundant RNAs account for 20% or more of the reads in some libraries: HBB, SMR3B, STATH, PRB4, CLPS, HTN3, SERPINA1, CTRB2, CPB1, CPA1 and MALAT1. Although the abundance of these transcripts is a function of their relatively stability as well as rate of initiation, a modest but significant over representation of ETS and YY1 sites was found in highly expressed promoters compared to weakly expressed ones (Extended Data Fig. 5g). Although the different motif composition may contribute to expression levels, the accompanying manuscript[4] shows that arrays of enhancers with similar usage[20] probably contribute to the higher maximal expression rate.

Key cell-type-specific transcription factors

Among 1,762 human and 1,516 mouse transcription factors compiled from the literature[21–23], promoter level expression profiles for 1,665 human transcription factors (94%) and 1,382 mouse transcription factors (91%) were obtained (Supplementary Tables 7, 8 and 9 and Supplementary Note 6). The distribution of expression levels and cell-type or tissue-specificity of transcription factors (Extended Data Fig. 3f–j) and the number of robust promoter peaks per transcription factor gene was similar to coding genes in general (4.8 compared to 4.6). In any given primary cell type, a median of 430 (306 to 722) transcription factors were expressed at 10 TPM or above (~3 copies per cell based on 300,000 mRNAs per cell[18]) (Extended Data Fig. 3g).

Clustering transcription factors by expression profile revealed sets of transcription factors specifically enriched in each cell type (Extended Data Fig. 8). For each primary cell sample we have made available ranked lists of transcription factors based on their promoter expression in the sample relative to the median across the collection (http://fantom.gsc.riken.jp/5/sstar/Browse_samples). For most cell types we found one transcription factor that was very highly enriched (≥100-fold), 23 highly enriched transcription factors (≥ tenfold) and 82 moderately enriched transcription factors (≥ fivefold) (numbers of transcription factors are based on median number of transcription factors observed at each enrichment threshold across the primary cell samples). To demonstrate their likely relevance we systematically reviewed phenotypes of transcription factor knockout mice at the MGI (see Supplementary Note 7). The clear connection between tissue-specific expression profiles and relevant knockout phenotypes is summarized in Supplementary Table 10. For example, in mouse inner ear hair cells, knockout of six of the top 20 most enriched transcription factor genes in mouse (*Pou3f4* (ref. 24), *Sox2* (ref. 25), *Egr2*, *Six1* (ref. 26), *Fos*[27], *Tbx18* (ref. 28)) as well as patient mutations in a further four top transcription factor genes (*POU4F3* (ref. 29), *ZIC2* (ref. 30), *SOX10* (ref. 31), *FOXF2* (ref. 32)) resulted in hearing-related defects. Similarly, mouse knockouts or patients with mutations in the transcription factors enriched in osteoblasts (*CREB3L1* (ref. 33), *DLX5* (ref. 34), *EBF2* (ref. 35), *HAND2* (ref. 36), *HOXC5* (ref. 37), *NFIX*[38], *PRRX1* (ref. 39), *PRRX2* (ref. 40), *SIX1* (ref. 41), *TWIST1* (ref. 42), *SHOX*[43], *Six2* (ref. 44)) had bone and osteoblast phenotypes. A substantial fraction of top transcription factors (61% of mouse and 40% of human transcription factors) have relevant phenotypes recorded in knockout mice (Supplementary Table 10).

Inferring function from expression profiles

Taking a pair-wise Pearson correlation matrix of the promoter expression profiles we carried out MCL clustering[45] (Supplementary Methods) to group promoters that share similar expression profiles across the atlas. Figure 4 shows a graphical overview of the structure of the data (and the mouse counterpart is shown in Extended Data Fig. 9). We find 6,030 cases of named genes with alternative promoters participating in two or more coexpression clusters (Extended Data Fig. 10). To evaluate and annotate these coexpressed groups, we tested for enrichment in specific Gene Ontology terms and in a curated database of 489 biological pathways. Of these, 356 pathways (174 KEGG, 114 WikiPathways, 46 Reactome, 22 Netpath) were significantly enriched in at least one human coexpression group (FDR <0.05). Using this approach, 38% of the unannotated robust peaks (35,082 out of 91,269) were within a cluster with a significant association to a pathway. The annotated coexpression groups are summarized in the website (http://fantom.gsc.riken.jp/5/sstar/Browse_coexpression_clusters) and a detailed example identifying genes putatively involved in influenza A pathogenesis is shown in Extended Data Fig. 10a.

Introducing sample ontology enrichment analysis (SOEA), we show that expression profiles can also be associated with cell, anatomical and disease ontology terms by testing for overrepresentation of terms in ranked lists of systematically annotated samples expressing each peak (Extended Data Fig. 11 and Supplementary Methods). Novel peaks can be annotated in this way. For example, an un-annotated DPI peak at hg19::chr18:3659943..3659972,+ is linked to the terms classical monocyte (CL:0000860; P value $= 6.35 \times 10^{-124}$, Extended Data Fig. 11h) and bone marrow (UBERON:0002371; P value $= 2.7 \times 10^{-80}$). Manual examination of the profile confirms the transcript is predominantly expressed in myeloid cells with higher levels in $CD14^{+}$ monocytes. Applied to all CAGE peaks, 127,645 human and 44,449 mouse robust peaks were annotated as enriched in at least one CL, DOID or UBERON term (Extended Data Fig. 11i, j). The most commonly-enriched terms at a P value threshold of 10^{-20} were classical monocyte (CL:0000860; 26,634 peaks, 14%), bone marrow (UBERON:0002371; 22,387 peaks, 12%) and neural tube (UBERON:0001049; 20,484 peaks, 11%) (Supplementary Table 13). This is consistent with the coexpression clustering in Fig. 4 (green and purple spheres correspond to leukocyte and central nervous system enriched expression profiles) and indicates that a large fraction of the mammalian genome is dedicated to immune and nervous system specific functions.

Conclusion

The FANTOM5 promoter atlas is a natural extension of earlier maps of active transcripts and promoters complementing the sequencing of mammalian genomes[46,47]. It represents an advance in an order of magnitude in the wide range of cell types and the amount of data produced per sample, and using single-molecule sequencing avoided polymerase chain reaction (PCR), digestion and cloning bias[48]. We have identified and quantified the activity of at least one promoter for more than 95% of annotated protein-coding genes in the human reference genome; only the activity of 1,225 promoters remains uncharacterized. Some of these may not actually be expressed. Some cannot be unambiguously measured with CAGE due to copy number variants or closely related multigene families. The remaining promoters are probably expressed in rare cell types or during windows of development or states of cellular activation that are not readily accessible and remain to be sampled. A continued effort to add profiles from these cells will make it possible to integrate them with the FANTOM5 data, and to extract metadata to identify those regulatory elements that are new and lineage-specific.

The FANTOM5 data highlights the value in profiling primary cells as opposed to whole tissues. It also highlights the weakness of using cancer cell lines. The cancer cell lines generally fail to cluster in a sample-to-sample correlation graph with their supposed cell type or tissue of origin (Extended Data Fig. 12) and express more transcription factors than primary cells (Extended Data Fig. 3g). The mutations and

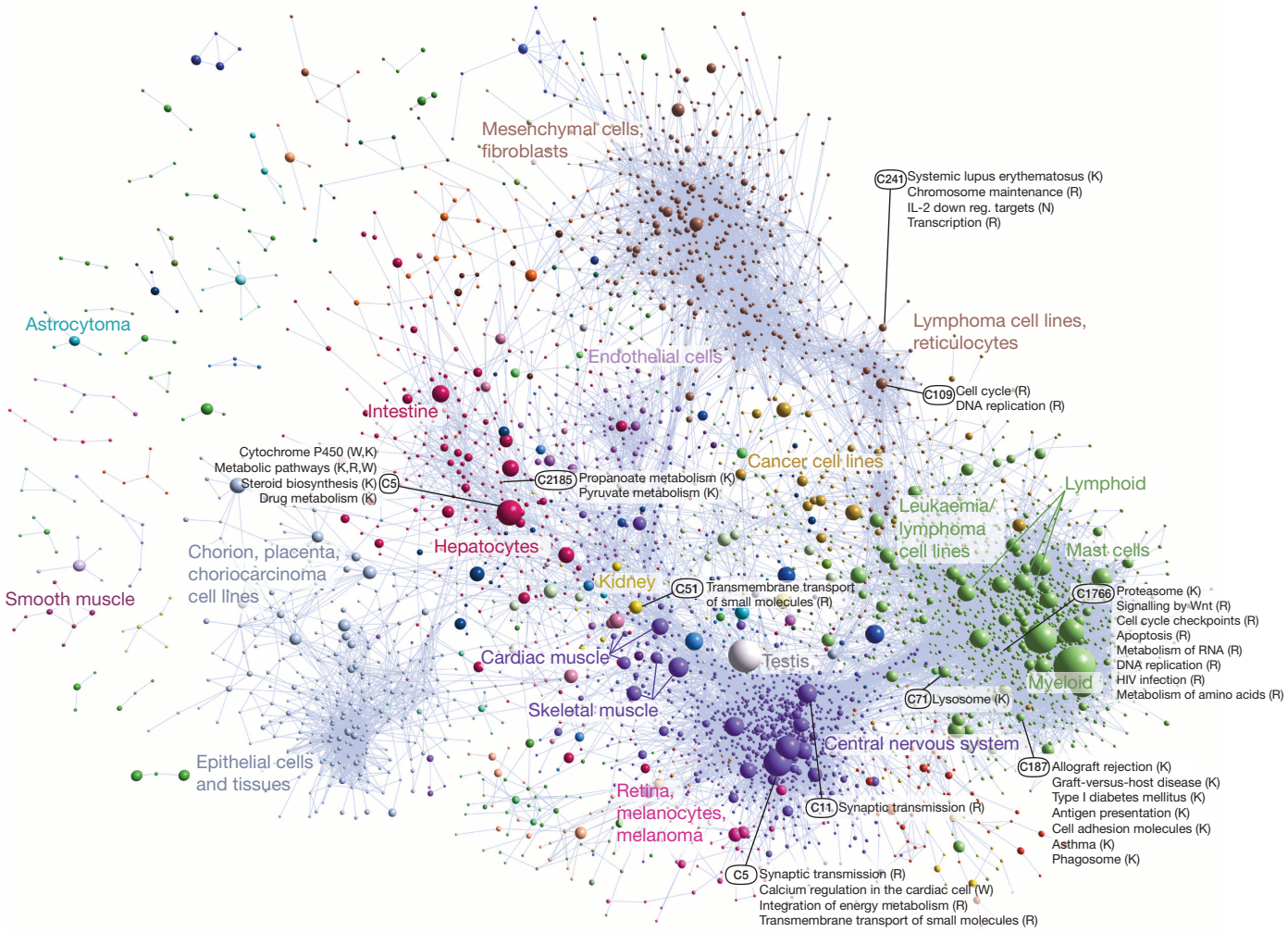


Figure 4 | Coexpression clustering of human promoters in FANTOM5. Collapsed coexpression network derived from 4,882 coexpression groups (one node is one group of promoters; 4,664 groups are shown here) derived from expression profiles of 124,090 promoters across all primary cell types, tissues and cell lines (visualized using Biolayout Express[3D] (ref. 45), $r > 0.75$, MCLi = 2.2). For display, each group of promoters is collapsed into a sphere, the radius of which is proportional to the cube root of the number of promoters in that group. Edges indicate $r > 0.6$ between the average expression profiles of each cluster. Colours indicate loosely-associated collections of coexpression groups (MCLi = 1.2). Labels show representative descriptions of the dominant cell type in coexpression groups in each region of the network, and a selection of highly-enriched pathways (FDR $<10^{-4}$) from KEGG (K), WikiPathways (W), Netpath (N) and Reactome (R). Promoters and genes in the coexpression groups are available online at (http://fantom.gsc.riken.jp/5/data/).

chromosomal rearrangements that occur in cancer result in unique transcriptional networks that do not exist in the untransformed state and do not necessarily generalize across multiple tumours of the same type. In terms of building mammalian transcriptional regulatory network models that reflect the normal untransformed state, primary cells are the logical choice. They have normal genomes, and express in the order of 430 transcription factors at appreciable levels, ranking of which can be used to reduce the complexity further and identify key known regulators of cellular phenotypes. Focusing on these key regulators and motif searching in the corresponding cell-type-specific promoters provides the data to build cell-type-specific regulatory network models and support a rational approach to identification of drivers required to reprogram cells from one lineage to another. Promoter-based expression data also has direct practical applications in the interpretation (and re-interpretation) of the function of single nucleotide polymorphisms (SNPs) in genome-wide association studies (GWAS), which commonly occur in non-coding sequences. In accompanying manuscripts, reanalysis of several GWAS data sets uncovered new disease associations in FANTOM5 promoters and identification of regulatory SNPs within enhancers that were active in medically relevant samples (ref. 4 and manuscript in preparation). Accordingly, the data will enable the design of genotyping arrays and sequence-capture systems to target regulatory variation, and the design of promoter constructs allowing researchers to specify the cell-type-specificity and absolute expression levels of their constructs (particularly for Cre-conditional knockouts[49] and gene therapy vectors[50]). In all these respects, the FANTOM5 data set greatly extends the data generated by ENCODE[5] to further our knowledge of genome function.

METHODS SUMMARY

All Methods are described in full in the Supplementary Information.

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

Received 4 January 2013; accepted 26 February 2014.

1. Vickaryous, M. K. & Hall, B. K. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev. Camb. Philos. Soc.* **81,** 425–455 (2006).
2. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Rev. Genet.* **13,** 233–245 (2012).
3. Kanamori-Katayama, M. *et al.* Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* **21,** 1150–1159 (2011).

4. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* http://dx.doi.org/10.1038/nature12787 (this issue).
5. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
6. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101,** 6062–6067 (2004).
7. Meehan, T. F. *et al.* Logical development of the cell ontology. *BMC Bioinformatics* **12,** 6 (2011).
8. Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E. & Haendel, M. A. Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* **13,** R5 (2012).
9. Osborne, J. D. *et al.* Annotating the human genome with Disease Ontology. *BMC Genomics* **10** (Suppl 1), S6 (2009).
10. Severin, J. *et al.* Interactive visualization and analysis of large-scale NGS data-sets using ZENBU. *Nature Biotechnol.* http://dx.doi.org/10.1038/nbt.2840 (2014).
11. Oja, E., Hyvarinen, A. & Karhunen, J. *Independent Component Analysis* (John Wiley & Sons, 2001).
12. Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature* **457,** 1028–1032 (2009).
13. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet.* **38,** 626–635 (2006).
14. Ioshikhes, I., Hosid, S. & Pugh, B. F. Variety of genomic DNA patterns for nucleosome positioning. *Genome Res.* **21,** 1863–1871 (2011).
15. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
16. Schug, J. *et al.* Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6,** R33 (2005).
17. Beissbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20,** 1464–1465 (2004).
18. Velculescu, V. E. *et al.* Analysis of human transcriptomes. *Nature Genet.* **23,** 387–388 (1999).
19. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328,** 1036–1040 (2010).
20. Barolo, S. Shadow enhancers: frequently asked questions about distributed *cis*-regulatory information and enhancer redundancy. *Bioessays* **34,** 135–141 (2012).
21. Roach, J. C. *et al.* Transcription factor expression in lipopolysaccharide-activated peripheral-blood-derived mononuclear cells. *Proc. Natl Acad. Sci. USA* **104,** 16245–16250 (2007).
22. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Rev. Genet.* **10,** 252–263 (2009).
23. Wingender, E., Schoeps, T. & Dönitz, J. TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Res.* **41,** D165–D170 (2013).
24. de Kok, Y. J. *et al.* Association between X-linked mixed deafness and mutations in the POU domain gene *POU3F4*. *Science* **267,** 685–688 (1995).
25. Kiernan, A. E. *et al. Sox2* is required for sensory organ development in the mammalian inner ear. *Nature* **434,** 1031–1035 (2005).
26. Zheng, W. *et al.* The role of *Six1* in mammalian auditory system development. *Development* **130,** 3989–4000 (2003).
27. Paylor, R., Johnson, R. S., Papaioannou, V., Spiegelman, B. M. & Wehner, J. M. Behavioral assessment of *c-fos* mutant mice. *Brain Res.* **651,** 275–282 (1994).
28. Trowe, M. O., Maier, H., Schweizer, M. & Kispert, A. Deafness in mice lacking the T-box transcription factor Tbx18 in otic fibrocytes. *Development* **135,** 1725–1734 (2008).
29. Vahava, O. *et al.* Mutation in transcription factor POU4F3 associated with inherited progressive hearing loss in humans. *Science* **279,** 1950–1954 (1998).
30. Chabchoub, E., Willekens, D., Vermeesch, J. R. & Fryns, J. P. Holoprosencephaly and *ZIC2* microdeletions: novel clinical and epidemiological specificities delineated. *Clin. Genet.* **81,** 584–589 (2012).
31. Pingault, V. *et al. SOX10* mutations in patients with Waardenburg-Hirschsprung disease. *Nature Genet.* **18,** 171–173 (1998).
32. Kapoor, S., Mukherjee, S. B., Shroff, D. & Arora, R. Dysmyelination of the cerebral white matter with microdeletion at 6p25. *Indian Pediatr.* **48,** 727–729 (2011).
33. Murakami, T. *et al.* Signalling mediated by the endoplasmic reticulum stress transducer OASIS is involved in bone formation. *Nature Cell Biol.* **11,** 1205–1211 (2009).
34. Acampora, D. *et al.* Craniofacial, vestibular and bone defects in mice lacking the *Distal-less*-related gene *Dlx5*. *Development* **126,** 3795–3809 (1999).
35. Kieslinger, M. *et al.* EBF2 regulates osteoblast-dependent differentiation of osteoclasts. *Dev. Cell* **9,** 757–767 (2005).
36. Funato, N. *et al.* Hand2 controls osteoblast differentiation in the branchial arch by inhibiting DNA binding of Runx2. *Development* **136,** 615–625 (2009).
37. McIntyre, D. C. *et al.* Hox patterning of the vertebrate rib cage. *Development* **134,** 2981–2989 (2007).
38. Driller, K. *et al.* Nuclear factor I X deficiency causes brain malformation and severe skeletal defects. *Mol. Cell. Biol.* **27,** 3855–3867 (2007).
39. Lu, M. F. *et al. prx-1* functions cooperatively with another paired-related homeobox gene, *prx-2*, to maintain cell fates within the craniofacial mesenchyme. *Development* **126,** 495–504 (1999).
40. Ten Berge, D., Brouwer, A., Korving, J., Martin, J. F. & Meijlink, F. *Prx1* and *Prx2* in skeletogenesis: roles in the craniofacial region, inner ear and limbs. *Development* **125,** 3831–3842 (1998).
41. Laclef, C. *et al.* Altered myogenesis in *Six1*-deficient mice. *Development* **130,** 2239–2252 (2003).
42. Lee, M. S., Lowe, G. N., Strong, D. D., Wergedal, J. E. & Glackin, C. A. *TWIST*, a basic helix-loop-helix transcription factor, can regulate the human osteogenic lineage. *J. Cell. Biochem.* **75,** 566–577 (1999).
43. Clement-Jones, M. *et al.* The short stature homeobox gene *SHOX* is involved in skeletal abnormalities in Turner syndrome. *Hum. Mol. Genet.* **9,** 695–702 (2000).
44. He, G. *et al.* Inactivation of *Six2* in mouse identifies a novel genetic mechanism controlling development and growth of the cranial base. *Dev. Biol.* **344,** 720–730 (2010).
45. Freeman, T. C. *et al.* Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput. Biol.* **3,** e206 (2007).
46. The FANTOM Consortium. The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–1563 (2005).
47. Suzuki, H. *et al.* The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genet.* **41,** 553–562 (2009).
48. Kawaji, H. *et al.* Comparison of CAGE and RNA-seq transcriptome profiling using a clonally amplified and single molecule next generation sequencing. *Genome Res.* http://dx.doi.org/10.1101/gr.156232.113 (2014).
49. Heffner, C. S. *et al.* Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource. *Nature Commun.* **3,** 1218 (2012).
50. Pringle, I. A. *et al.* Rapid identification of novel functional promoters for gene therapy. *J. Mol. Med.* **90,** 1487–1496 (2012).

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

Acknowledgements FANTOM5 was made possible by a Research Grant for RIKEN Omics Science Center from MEXT to Y. Hayashizaki and a grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT, Japan to Y. Hayashizaki. It was also supported by Research Grants for RIKEN Preventive Medicine and Diagnosis Innovation Program (RIKEN PMI) to Y. Hayashizaki and RIKEN Centre for Life Science Technologies, Division of Genomic Technologies (RIKEN CLST (DGT)) from the MEXT, Japan. Extended acknowledgements are provided in the Supplementary Information.

Author Contributions The core members of FANTOM5 phase 1 were Alistair R. R. Forrest, Hideya Kawaji, Michael Rehli, J. Kenneth Baillie, Michiel J. L. de Hoon, Timo Lassmann, Masayoshi Itoh, Kim M. Summers, Harukazu Suzuki, Carsten O. Daub, Jun Kawai, Peter Heutink, Winston Hide, Tom C. Freeman, Boris Lenhard, Vladimir B. Bajic, Martin S. Taylor, Vsevolod J. Makeev, Albin Sandelin, David A. Hume, Piero Carninci and Yoshihide Hayashizaki. Samples were provided by: A. Blumenthal, A. Bonetti, A. Mackay-sim, A. Sajantila, A. Saxena, A. Schwegmann, A.G.B., A.J.K., A.L., A.R.R.F., A.S.B.E., B.B., C. Schmidl, C. Schneider, C.A.D., C.A.W., C.K., C.L.M., D.A.H., D.A.O., D.G., D.S., D.V., E.W., F.B., F.N., G.G.S., G.J.F., G.S., H. Kawamoto, H. Koseki, H. Morikawa, H. Motohashi, H. Ohno, H. Sato, H. Satoh, H. Tanaka, H. Tatsukawa, H. Toyoda, H.C.C., H.E., J. Kere, J.B., J.F., J.K.B., J.S.K., J.T., J.W.S., K.E., K.J.H., K.M., K.M.S., L.F., L.M.K., L.M.vdB., L.N.W., M. Edinger, M. Endoh, M. Fagiolini, M. Hamaguchi, M. Hara, M. Herlyn, M. Morimoto, M. Rehli, M. Yamamoto, M. Yoneda, M.B., M.C.F.C., M.D., M.E.F., M.O., M.O.H., M.P., M.vdW., N.M., N.O., N.T., P.A., P.G.Z., P.H., P.R., R.F., R.G., R.K.S., R.P., R.V., S. Guhl, S. Gustincich, S. Kojima, S. Koyasu, S. Krampitz, S. Sakaguchi, S. Savvi, S.E.Z., S.O., S.P.B., S.P.K., S. Roy., S.Z., T. Kitamura, T. Nakamura, T. Nozaki, T. Sugiyama, T.B.G., T.D., T.G., T.I., T.J.H., T.J.K., V.O., W.L., Y. Hasegawa, Y. Nakachi, Y. Nakamura, Y. Yamaguchi, Y. Yonekura, Y.I., Y.I.K., Y.M. and Y.O. Analyses were carried out by: A. Mathelier, A. Meynert, A. Sandelin, A.C., A.D.D., A.P.G., A.H., A.J., A.M.B., A.P., A.R.R.F., A.S.K., A.T.K., A.V.F., B. Lenhard, B. Lilje, B.D., B.K., B.M., B.R.J., C. Schmidl, C. Schneider, C.A.S., C.F., C.J.M., C.O.D., C.P., C.V.C., D.A., D.A.M., D.C., E. Dalla, E. Dimont, E.A., E.A.S., E.J.W., E.M., E.V., Ev.N., F.D., G.J., G.J.F., G.M.A., H. Kawaji, H. Ohmiya, H. Shimoji, H.F., H.J., H.P., I.A., I.E.V., I.H., I.V.K., J.A.B., J.A.C.A., J.A.R., J.C.M., J.F.J.L., J.G., J.G.D.P., J.H., J.K.B., J.S., K. Kajiyama, K.I., K.L., L.H., L.L., M. Francescatto, M. Rashid, M. Rehli, M. Roncador, M. Thompson, M.B.R., M.C., M.C.F., M.J., M.J.L.dH., M.L., M.S.T., M.V., N.B., O.J.L.R., O.M.H., P.A.C.tH., P.J.B, R.A., R.S.Y., S. Katayama, S. Kawaguchi, S. Schmeier, S. Rennie, S.F., S.J.H.S., S.P., T. Sengstag, T.C.F., T.F.M., T.H., T.K., T.L., T.R., T.T., U.S., V.B.B., V.H., V.J.M., W.H., W.W.W., X.Z., Y. Chen, Y. Ciani, Y.A.M., Y.S., Z.T. Libraries were generated by: A. Kaiho, A. Kubosaki, A. Saka, C. Simon, E.S., F.H., H.N., J. Kawai, K. Kaida, K.N., M. Furuno, M. Murata, M. Sakai, M. Tagami, M.I., M.K., M.K.K., N.K., N.N., N.S., P.C., R.M., S. Kato, S.N., S.N.-S., S.W., S.Y., T.A., T. Kawashima. The manuscript was written by A.R.R.F. and D.A.H. with help from A. Sandelin, J.K.B., M. Rehli, H.K., M.J.L.dH., V.H., I.V.K., M.T. and K.M.S. with contributions, edits and comments from all authors. The project was managed by Y. Hayashizaki, A.R.R.F., P.C., M.I., M.S., J. Kawai, C.O.D., H. Suzuki, T.L. and N.K. The scientific coordinator was A.R.R.F and the general organizer was Y. Hayashizaki.

Author Information All CAGE data has been deposited at DDBJ DRA under accession number DRA000991. 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.R.R.F (alistair.forrest@gmail.com), P.C. (carninci@riken.jp) or Y.H. (yosihide@gsc.riken.jp).

The FANTOM Consortium and the RIKEN PMI and CLST (DGT)

Alistair R. R. Forrest[1,2]*, Hideya Kawaji[1,2,3]*, Michael Rehli[4,5]*, J. Kenneth Baillie[6]*, Michiel J. L. de Hoon[1,2], Vanja Haberle[7,8], Timo Lassmann[1,2], Ivan V. Kulakovskiy[9,10], Marina Lizio[1,2], Masayoshi Itoh[1,2,3], Robin Andersson[11], Christopher J. Mungall[12], Terrence F. Meehan[13], Sebastian Schmeier[14,15], Nicolas Bertin[1,2], Mette Jørgensen[11], Emmanuel Dimont[16], Erik Arner[1,2], Christian Schmidl[4]†, Ulf Schaefer[14], Yulia A. Medvedeva[10,14]†, Charles Plessy[1,2], Morana Vitezic[1,17], Jessica Severin[1,2], Colin A. Semple[18], Yuri Ishizu[1,2], Robert S. Young[18], Margherita Francescatto[19,20], Intikhab Alam[14], Davide Albanese[21], Gabriel M. Altschuler[16], Takahiro Arakawa[1,2], John A. C.

Archer[14], Peter Arner[22], Magda Babina[23], Sarah Rennie[18], Piotr J. Balwierz[24], Anthony G. Beckhouse[25,26], Swati Pradhan-Bhatt[27], Judith A. Blake[28], Antje Blumenthal[26,29], Beatrice Bodega[30], Alessandro Bonetti[1,2], James Briggs[25]†, Frank Brombacher[31,32], A. Maxwell Burroughs[1], Andrea Califano[33,34,35,36], Carlo V. Cannistraci[37,38]†, Daniel Carbajo[39], Yun Chen[11], Marco Chierici[21], Yari Ciani[40], Hans C. Clevers[41,42,43], Emiliano Dalla[40], Carrie A. Davis[44], Michael Detmar[45], Alexander D. Diehl[46], Taeko Dohi[47], Finn Drabløs[48], Albert S. B. Edge[49], Matthias Edinger[4,5], Karl Ekwall[50], Mitsuhiro Endoh[51,52], Hideki Enomoto[53], Michela Fagiolini[54], Lynsey Fairbairn[6], Hai Fang[55], Mary C. Farach-Carson[56], Geoffrey J. Faulkner[57], Alexander V. Favorov[10,58,59], Malcolm E. Fisher[6], Martin C. Frith[60], Rie Fujita[61], Shiro Fukuda[1], Cesare Furlanello[21], Masaaki Furuno[1,2], Jun-ichi Furusawa[51,52,62], Teunis B. Geijtenbeek[63], Andrew P. Gibson[64], Thomas Gingeras[44], Daniel Goldowitz[65], Julian Gough[55], Sven Guhl[23], Reto Guler[31,32], Stefano Gustincich[66], Thomas J. Ha[65], Masahide Hamaguchi[67], Mitsuko Hara[68], Matthias Harbers[1], Jayson Harshbarger[1,2], Akira Hasegawa[1,2], Yuki Hasegawa[1,2], Takehiro Hashimoto[1], Meenhard Herlyn[69], Kelly J. Hitchens[25,26], Shannan J. Ho Sui[16], Oliver M. Hofmann[16], Ilka Hoof[11], Fumi Hori[1,2], Lukasz Huminiecki[17], Kei Iida[70], Tomokatsu Ikawa[51,52], Boris R. Jankovic[14], Hui Jia[71], Anagha Joshi[6], Giuseppe Jurman[21], Bogumil Kaczkowski[1,2], Chieko Kai[72], Kaoru Kaida[1,2], Ai Kaiho[1], Kazuhiro Kajiyama[1,2], Mutsumi Kanamori-Katayama[1], Artem S. Kasianov[10], Takeya Kasukawa[2], Shintaro Katayama[1], Sachi Kato[1,2], Shuji Kawaguchi[70], Hiroshi Kawamoto[51], Yuki I. Kawamura[47], Tsugumi Kawashima[1,2], Judith S. Kempfle[49], Tony J. Kenna[29], Juha Kere[50,73], Levon M. Khachigian[74], Toshio Kitamura[75], S. Peter Klinken[76], Alan J. Knox[77], Miki Kojima[1,2], Soichi Kojima[68], Naoto Kondo[1,2], Haruhiko Koseki[51,52], Shigeo Koyasu[51,52,62], Sarah Krampitz[45], Atsutaka Kubosaki[1], Andrew T. Kwon[1,2], Jeroen F. J. Laros[64], Weonju Lee[78], Andreas Lennartsson[50], Kang Li[11], Berit Lilje[11], Leonard Lipovich[71], Alan Mackay-sim[79], Ri-ichiroh Manabe[1,2], Jessica C. Mar[39], Benoit Marchand[14], Anthony Mathelier[65], Niklas Mejhert[22], Alison Meynert[18], Yosuke Mizuno[80], David A. de Lima Morais[81], Hiromasa Morikawa[67], Mitsuru Morimoto[53], Kazuyo Moro[51,52,62,82], Efthymios Motakis[1,2], Hozumi Motohashi[83], Christine L. Mummery[84], Mitsuyoshi Murata[1,2], Sayaka Nagao-Sato[1], Yutaka Nakachi[80,85], Fumio Nakahara[75], Toshiyuki Nakamura[72], Yukio Nakamura[86], Kenichi Nakazato[1], Erik van Nimwegen[24], Noriko Ninomiya[1], Hiromi Nishiyori[1,2], Shohei Noma[1,2], Tadasuke Nozaki[87], Soichi Ogishima[88]†, Naganari Ohkura[67], Hiroko Ohmiya[1,2]†, Hiroshi Ohno[51,52], Mitsuhiro Ohshima[89], Mariko Okada-Hatakeyama[51,52], Yasushi Okazaki[80,85], Valerio Orlando[30,37], Dmitry A. Ovchinnikov[25], Arnab Pain[14,37], Robert Passier[84], Margaret Patrikakis[74], Helena Persson[50], Silvano Piazza[40], James G. D. Prendergast[18], Owen J. L. Rackham[55], Jordan A. Ramilowski[1,2], Mamoon Rashid[14,37], Timothy Ravasi[37,38], Patrizia Rizzu[19], Marco Roncador[21], Sugata Roy[1,2], Morten B. Rye[48], Eri Saijyo[1], Antti Sajantila[90], Akiko Saka[1], Shimon Sakaguchi[67], Mizuho Sakai[1,2], Hiroki Sato[72], Hironori Satoh[61], Suzana Savvi[31,32], Alka Saxena[1]†, Claudio Schneider[40,91], Erik A. Schultes[64], Gundula G. Schulze-Tanzil[92], Anita Schwegmann[31,32], Thierry Sengstag[1], Guojun Sheng[53], Hisashi Shimoji[1], Yishai Shimoni[36], Jay W. Shin[1,2], Christophe Simon[1,2], Daisuke Sugiyama[93], Takaaki Sugiyama[72], Masanori Suzuki[1], Naoko Suzuki[1,2], Rolf K. Swoboda[69], Peter A. C. 't Hoen[64], Michihira Tagami[1,2], Naoko Takahashi[1,2], Jun Takai[61], Hiroshi Tanaka[88], Hideki Tatsukawa[94], Zuotian Tatum[64], Mark Thompson[64], Hiroo Toyoda[87], Tetsuro Toyoda[70], Eivind Valen[95], Marc van de Wetering[41], Linda M. van den Berg[63], Roberto Verardo[40], Dipti Vijayan[25,26], Ilya E. Vorontsov[10], Wyeth W. Wasserman[65], Shoko Watanabe[1], Christine A. Wells[25,26], Louise N. Winteringham[76], Ernst Wolvetang[25], Emily J. Wood[71], Yoko Yamaguchi[96], Masayuki Yamamoto[61], Misako Yoneda[72], Yohei Yonekura[53], Shigehiro Yoshida[1,2], Susan E. Zabierowski[69], Peter G. Zhang[65], Xiaobei Zhao[11], Silvia Zucchelli[66], Kim M. Summers[6], Harukazu Suzuki[1,2], Carsten O. Daub[1], Jun Kawai[1,3], Peter Heutink[19], Winston Hide[16], Tom C. Freeman[6], Boris Lenhard[8,97], Vladimir B. Bajic[14], Martin S. Taylor[18], Vsevolod J. Makeev[9,10,98], Albin Sandelin[11], David A. Hume[6], Piero Carninci[1,2], Yoshihide Hayashizaki[1,3]

[1]RIKEN Omics Science Center (OSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [2]RIKEN Center for Life Science Technologies (Division of Genomic Technologies) (CLST (DGT)), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [3]RIKEN Preventive Medicine and Diagnosis Innovation Program (PMI), 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan. [4]Department of Internal Medicine III, University Hospital Regensburg, F.-J.-Strauss Allee 11, D-93042 Regensburg, Germany. [5]Regensburg Centre for Interventional Immunology (RCI), D-93042 Regensburg, Germany. [6]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Edinburgh, Midlothian EH25 9RG, UK. [7]Department of Biology, University of Bergen, Thormøhlensgate 53, NO-5006 Bergen, Norway. [8]Faculty of Medicine, Institute of Clinical Sciences, MRC Clinical Sciences Centre, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, UK. [9]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilov str. 32, Moscow 119991, Russia. [10]Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkin str. 3, Moscow 119991, Russia. [11]The Bioinformatics Centre, Department of Biology and BRIC, University of Copenhagen, Ole Maaloes Vej 5, DK 2200 Copenhagen, Denmark. [12]Genomics Division, Lawrence Berkeley National Laboratory, 84R01, 1 Cyclotron Road, Berkeley, California 94720, USA. [13]Mouse Informatics, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [14]Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Ibn Al-Haytham Building -2, Thuwal 23955-6900, Kingdom of Saudi Arabia. [15]Institute of Natural and Mathematical Sciences, Massey University, Private Bag 102-904, North Shore Mail Centre, 0745 Auckland, New Zealand. [16]Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave, Boston, Massachusetts 02115, USA. [17]Department of Cell and Molecular Biology, Karolinska Institutet, P.O. Box 285, SE-171 77 Stockholm, Sweden. [18]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine (MRC-IGMM), University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK. [19]Department of Clinical Genetics, VU University Medical Center Amsterdam, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands. [20]Graduate Program in Areas of Basic and Applied Biology, Abel Salazar Biomedical Sciences Institute, University of Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal. [21]Predictive Models for Biomedicine and Environment, Fondazione Bruno Kessler, via Sommarive 18, 38123 Trento, Italy. [22]Department of Medicine, Karolinska Institutet at Karolinska University Hospital, Huddinge, SE-141 86 Huddinge, Sweden. [23]Department of Dermatology and Allergy, Charité Campus Mitte, Universitätsmedizin Berlin, Chariteplatz 1, 10117 Berlin, Germany. [24]Biozentrum, University of Basel, Klingelbergstrasse 50-70, 4056 Basel, Switzerland. [25]Australian Institute for Bioengineering and Nanotechnology (AIBN), University of Queensland, Brisbane St Lucia, Queensland 4072, Australia. [26]Australian Infectious Diseases Research Centre (AID), University of Queensland, Brisbane St Lucia, Queensland 4072, Australia. [27]Department of Biological Sciences, University of Delaware, Newark, Delaware 19713, USA. [28]Bioinformatics and Computational Biology, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA. [29]Diamantina Institute, University of Queensland, Brisbane St Lucia, Queensland 4072, Australia. [30]IRCCS Fondazione Santa Lucia, via del Fosso di Fiorano 64, 00143 Rome, Italy. [31]Immunology and Infectious Disease, International Centre for Genetic Engineering & Biotechnology (ICGEB) Cape Town component, Anzio Road, Observatory 7925, Cape Town, South Africa. [32]Division of Immunology, Institute of Infectious Diseases and Molecular Medicine (IDM), University of Cape Town, Anzio Road, Observatory 7925, Cape Town, South Africa. [33]Department of Systems Biology, Columbia University Medical Center, 1130 St. Nicholas Avenue, New York, New York 10032, USA. [34]Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, 701 West 168th Street, New York, New York 10032, USA. [35]Department of Biomedical Informatics, Columbia University Medical Center, 622 West 168th Street, VC5, New York, New York 10032, USA. [36]Institute of Cancer Genetics, Columbia University Medical Center, Herbert Irving Comprehensive Cancer Center, 1130 St. Nicholas Avenue, New York, New York 10032, USA. [37]Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Ibn Al-Haytham Building -2, Thuwal 23955-6900, Kingdom of Saudi Arabia. [38]Applied Mathematics and Computational Science Program, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia. [39]Department of Systems and Computational Biology, Albert Einstein College of Medicine, The Bronx, New York, New York 10461, USA. [40]Laboratorio Nazionale del Consorzio Interuniversitario per le Biotecnologie (LNCIB), Padriciano 99, 34149 Trieste, Italy. [41]Hubrecht Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. [42]The Royal Netherlands Academy of Arts and Sciences, P.O. Box 19121, NL-1000 GC Amsterdam, The Netherlands. [43]University Medical Centre Utrecht, Postbus 85500, 3508 GA Utrecht, The Netherlands. [44]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11797, USA. [45]Institute of Pharmaceutical Sciences, ETH Zurich, Vladimir-Prelog-Weg 3, HCI H 303, 8093 Zurich, Switzerland. [46]Department of Neurology, University at Buffalo School of Medicine and Biomedical Sciences, New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, New York 14203, USA. [47]Gastroenterology, Research Center for Hepatitis and Immunology Research Institute, National Center for Global Health and Medicine, 1-7-1 Kohnodai, Ichikawa, Chiba 272-8516, Japan. [48]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology (NTNU), P.O. Box 8905, NO-7491 Trondheim, Norway. [49]Department of Otology and Laryngology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Eaton-Peabody Lab, 243 Charles Street, Boston, Massachusetts 02114, USA. [50]Department of Biosciences and Nutrition, Center for Biosciences, Karolinska Institutet, Hälsovägen 7-9, SE-141 83 Huddinge, Sweden. [51]RIKEN Research Center for Allergy and Immunology (RCAI), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [52]RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [53]RIKEN Center for Developmental Biology (CDB), 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. [54]FM Kirby Neurobiology Center, Children's Hospital Boston, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. [55]Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK. [56]Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77251-1892, USA. [57]Cancer Biology Program, Mater Medical Research Institute, Raymond Terrace, South Brisbane, Queensland 4101, Australia. [58]Department of Oncology, Division of Oncology, Biostatistics and Bioinformatics, Johns Hopkins University School of Medicine, 550 North Broadway, Baltimore, Maryland 21205, USA. [59]State Research Institute of Genetics and Selection of Industrial Microorganisms GosNIIgenetika, 1-st Dorozhniy pr., 1, 117545 Moscow, Russia. [60]Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan. [61]Department of Medical Biochemistry, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan. [62]Department of Microbiology and Immunology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku, Tokyo 160-8582, Japan. [63]Experimental Immunology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. [64]Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. [65]Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, 950 West 28th Avenue, Vancouver, British Columbia V5Z 4H4, Canada. [66]Neuroscience, SISSA, via Bonomea 265, 34136 Trieste, Italy. [67]Experimental Immunology, Immunology Frontier Research Center, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan. [68]RIKEN Advanced Science Institute (ASI), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. [69]Melanoma Research Center, The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA. [70]RIKEN Bioinformatics And Systems Engineering Division (BASE), 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan. [71]Center for Molecular Medicine and Genetics, Wayne State University, 3228 Scott Hall, 540 East Canfield Street, Detroit, Michigan 48201-1928, USA. [72]Laboratory Animal Research Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [73]Science for Life Laboratory, Box 1031, SE-171 21 Solna, Sweden. [74]Centre for Vascular Research, University of New South

Wales, Sydney, New South Wales 2052, Australia. [75]Division of Cellular Therapy and Division of Stem Cell Signaling, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [76]Harry Perkins Institute of Medical Research, and the Centre for Medical Research, University of Western Australia, QQ Block, QEII Medical Centre, Nedlands, Perth, Western Australia 6009, Australia. [77]Respiratory Medicine, University of Nottingham, Clinical Sciences Building, City Hospital, Hucknall Road, Nottingham NG5 1PB, UK. [78]Department of Dermatology, Kyungpook National University School of Medicine, 130 Dongdeok-ro Jung-gu, Daegu 700-721, South Korea. [79]National Centre for Adult Stem Cell Research, Eskitis Institute for Cell and Molecular Therapies, Griffith University, Brisbane, Queensland 4111, Australia. [80]Division of Functional Genomics and Systems Medicine, Research Center for Genomic Medicine, Saitama Medical University, 1397-1 Yamane, Hidaka, Saitama 350-1241, Japan. [81]Faculty of Engineering, University of Bristol, Merchant Venturers Building, Woodland Road, Clifton BS8 1UB, UK. [82]PRESTO, Japanese Science and Technology Agency (JST), 7 Gobancho, Chiyodaku, Tokyo 102-0076, Japan. [83]Center for Radioisotope Sciences, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan. [84]Anatomy and Embryology, Leiden University Medical Center, Einthovenweg 20, P.O. Box 9600, 2300 RC Leiden, The Netherlands. [85]Division of Translational Research, Research Center for Genomic Medicine, Saitama Medical University, 1397-1 Yamane, Hidaka, Saitama 350-1241, Japan. [86]RIKEN BioResource Center (BRC), Koyadai 3-1-1, Tsukuba, Ibaraki 305-0074, Japan. [87]Department of Clinical Molecular Genetics, School of Pharmacy, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan. [88]Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan. [89]Department of Biochemistry, Ohu University School of Pharmaceutical Sciences, Misumido 31-1, Tomitamachi, Koriyama, Fukushima 963-8611, Japan. [90]Hjelt Institute, Department of Forensic Medicine, University of Helsinki, Kytosuontie 11, 003000 Helsinki, Finland. [91]DSMB Dipartimento Scienze Mediche e Biologiche University of Udine, P.le Kolbe 3, 33100 Udine, Italy. [92]Department of Orthopedic, Trauma and Reconstructive Surgery, Charité Universitätsmedizin Berlin, Garystrasse 5, 14195 Berlin, Germany. [93]Center for Clinical and Translational Reseach, Kyushu University Hospital, Station for Collaborative Research1 4F, 3-1-1 Maidashi, Higashi-Ku, Fukuoka 812-8582, Japan. [94]Graduate School of Pharmaceutical Sciences, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan. [95]Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [96]Department of Biochemistry, Nihon University School of Dentistry, 1-8-13, Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8310, Japan. [97]Department of Informatics, University of Bergen, Høgteknologisenteret, Thormøhlensgate 53, NO-5008 Bergen, Norway. [98]Department of Biological and Medical Physics, Moscow Institute of Physics and Technology (MIPT) 9, Institutsky Per., Dolgoprudny, Moscow Region 141700, Russia.

†Present addresses: Institute of Predictive and Personalized Medicine of Cancer, Ctra. de Can Roti, cami de les escoles, s/n, 08916 Badalona (Barcelona), Spain (Y.A.M.); Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Technische Universität Dresden, Dresden, Germany (C.V.C.); Genomics Core Facility, Biomedical Research Centre, Guy's Hospital, London SE1 9RT, UK (A. Saxena); RIKEN Advanced Center for Computing and Communication (ACCC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045 Japan (H. Ohmiya); Research Center for Molecular Medicine of the Austrian Academy of Sciences (CeMM), 1090 Vienna, Austria (C. Schmidl); Department of Biological and Biomedical Sciences, Harvard University, Cambridge, Massachusetts 02138, USA (J.B.); Department of Bioclinical Informatics,Tohoku Medical Megabank Organization,Tohoku University. Sendai 980-8573, Japan (S.O.).

*These authors contributed equally to this work.

51. Pham, T. H. *et al.* Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood* **119,** e161–e171 (2012).
52. Shulha, H. P. *et al.* Epigenetic signatures of autism; trimethylated H3K4 landscapes in prefrontal neurons. *Arch. Gen. Psychiatry* **69,** 314–324 (2012).
53. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5,** 730–737 (2004).
54. Shapira, S. D. *et al.* A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* **139,** 1255–1267 (2009).
55. Talukder, A. H. *et al.* Phospholipid scramblase 1 regulates Toll-like receptor 9-mediated type I interferon production in plasmacytoid dendritic cells. *Cell Res.* **22,** 1129–1139 (2012).

LETTER

doi:10.1038/nature13156

A Sedna–like body with a perihelion of 80 astronomical units

Chadwick A. Trujillo[1]* & Scott S. Sheppard[2]*

The observable Solar System can be divided into three distinct regions: the rocky terrestrial planets including the asteroids at 0.39 to 4.2 astronomical units (AU) from the Sun (where 1 AU is the mean distance between Earth and the Sun), the gas giant planets at 5 to 30 AU from the Sun, and the icy Kuiper belt objects at 30 to 50 AU from the Sun. The 1,000-kilometre-diameter dwarf planet Sedna was discovered ten years ago and was unique in that its closest approach to the Sun (perihelion) is 76 AU, far greater than that of any other Solar System body[1]. Formation models indicate that Sedna could be a link between the Kuiper belt objects and the hypothesized outer Oort cloud at around 10,000 AU from the Sun[2–6]. Here we report the presence of a second Sedna-like object, 2012 VP_{113}, whose perihelion is 80 AU. The detection of 2012 VP_{113} confirms that Sedna is not an isolated object; instead, both bodies may be members of the inner Oort cloud, whose objects could outnumber all other dynamically stable populations in the Solar System.

The inner Oort cloud objects probably formed on nearly circular orbits, allowing them to accumulate mass efficiently[6–9], and were later perturbed into the eccentric orbits we see today. We define an inner Oort cloud object as a body whose orbit is not readily formed with the known mass in the Solar System. This typically means a perihelion greater than 50 AU (beyond the range of significant Neptune interaction) and a semi-major axis in the range 150 AU $< a <$ 1,500 AU. At above 1,500 AU objects may be considered to be in the outer Oort cloud, as galactic tides start to become important in the formation process[10]. Sedna and 2012 VP_{113} are the clearest examples of inner Oort cloud objects because they do not interact significantly with any of the known planets owing to their extremely distant perihelia (Fig. 1 and Table 1).

There are two main models for inner Oort cloud object formation. In the first model, planet-sized object(s) in the outer Solar System may perturb objects from the Kuiper belt outward to inner-Oort-cloud orbits. These planet-sized objects could either remain (unseen) in the Solar System[5,11] or have been ejected from the Solar System during the creation of the inner Oort cloud[12]. In the second model, close stellar encounter(s) can create the inner Oort cloud objects, possibly within the first ten million years (Myr) of the Sun's life, when it resided within its birth cluster[1,3,4,11,13–17]. The outer Oort cloud objects could be created from galactic tides[18], but the more tightly bound inner Oort cloud objects are difficult to create without a close stellar encounter[19]. A third, less explored, model is that the inner Oort cloud objects are captured extrasolar planetesimals lost from other stars that were in the Sun's birth cluster[2,3,20]. Each theory of inner Oort cloud object formation predicts different orbital configurations for the population. Therefore, as more inner Oort cloud objects are discovered, their orbits will provide strong constraints on the inner Oort cloud object formation models and thus our Solar System's evolution.

To place constraints on the inner Oort cloud object population, we constructed a simple simulation of the observational biases affecting our survey (Methods). We find three basic results from this analysis: (1) there appears to be a paucity of inner Oort cloud objects with perihelia in the range 50–75 AU, suggesting that the inner Oort cloud object population could increase with $q > 75$ AU; (2) the existence of 2012 VP_{113} means that the inner Oort cloud object population must reach down to fairly small semi-major axes of about 250 AU; and (3) there are no observational biases that can explain the clustering of the argument of perihelion (ω) near 340° for inner Oort cloud objects and all objects with semi-major axes greater than 150 AU and perihelia greater than Neptune.

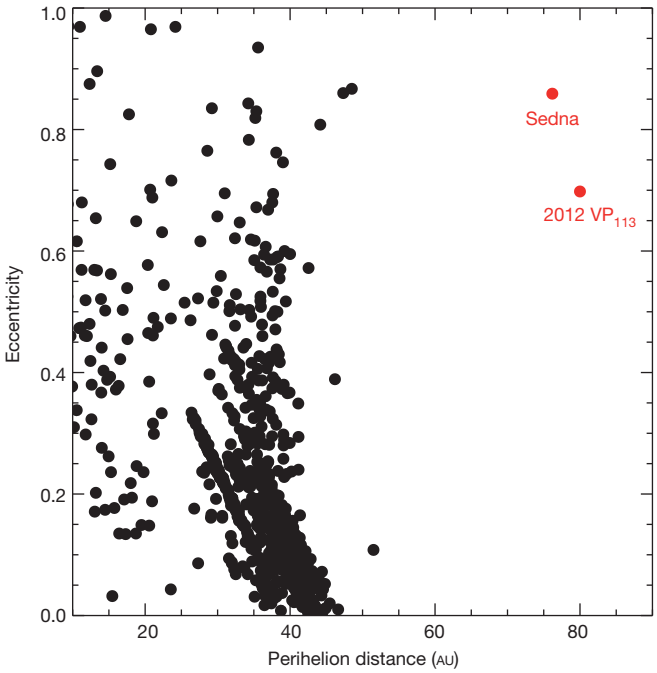


Figure 1 | Sedna and 2012 VP_{113} are clear dynamical outliers in the Solar System. Eccentricity versus perihelion distance for the approximately 1,000 minor planets with well-determined (multi-year) orbits beyond 10 AU are depicted. There are no known objects with closest-approach distances (perihelia, q) between 55 AU and 75 AU, even though such objects would be closer, and should therefore be brighter and easier to detect than Sedna or 2012 VP_{113}. This suggests there may be a paucity of inner Oort cloud objects with $q < 75$ AU. The perihelia of Sedna and 2012 VP_{113} are much too distant from Neptune (30 AU) for their existence to be explained by the known mass in the Solar System. All error bars are smaller than the data symbols.

Although our survey was sensitive to objects from 50 AU to beyond 300 AU, no objects were found with perihelion distances between 50 AU and 75 AU, where objects are brightest and easiest to detect. This was true for the original survey that found Sedna[1] and the deeper follow-up survey[21]. Figure 2 illustrates this—most simulated objects (and the real objects Sedna and 2012 VP_{113}) are found near the inner edge of the population. If the inner Oort cloud objects had a minimum perihelion of 50 AU and followed a size distribution like that of the large end of all known

[1]Gemini Observatory, 670 North A'ohoku Place, Hilo, Hawaii 96720, USA. [2]Department of Terrestrial Magnetism, Carnegie Institution for Science, 5241 Broad Branch Road NW, Washington, DC 20015, USA.
*These authors contributed equally to this work.

Table 1 | The orbital elements of the inner Oort cloud objects

Object	q (AU)	a (AU)	e	i (deg)	Ω (deg)	ω (deg)	t_p (JD)	Epoch (JD)
2012 VP_{113}	80	266	0.698	24.051	90.806	293	2,443,300	2,455,857.1
Sedna	76.23	542	0.8593	11.92840	144.466	311.01	2,479,197	2,456,400.5

The barycentric elements for 2012 VP_{113} and the heliocentric elements for Sedna (from the NASA Jet Propulsion Laboratory Horizons tool for dwarf planets http://ssd.jpl.nasa.gov/?horizons) are given as semi-major axis (a), eccentricity (e), inclination (i), longitude of ascending node (Ω), argument of perihelion (ω), and perihelion distance (q), which has been derived from a and e. Time of perihelion (t_p) and epoch are specified in terms of the Julian date (JD). The precision to which quantities are known is reflected in the number of significant digits. We define inner Oort cloud objects as having $q > 50$ AU and $150 \text{ AU} < a < 1{,}500$ AU (see main text).

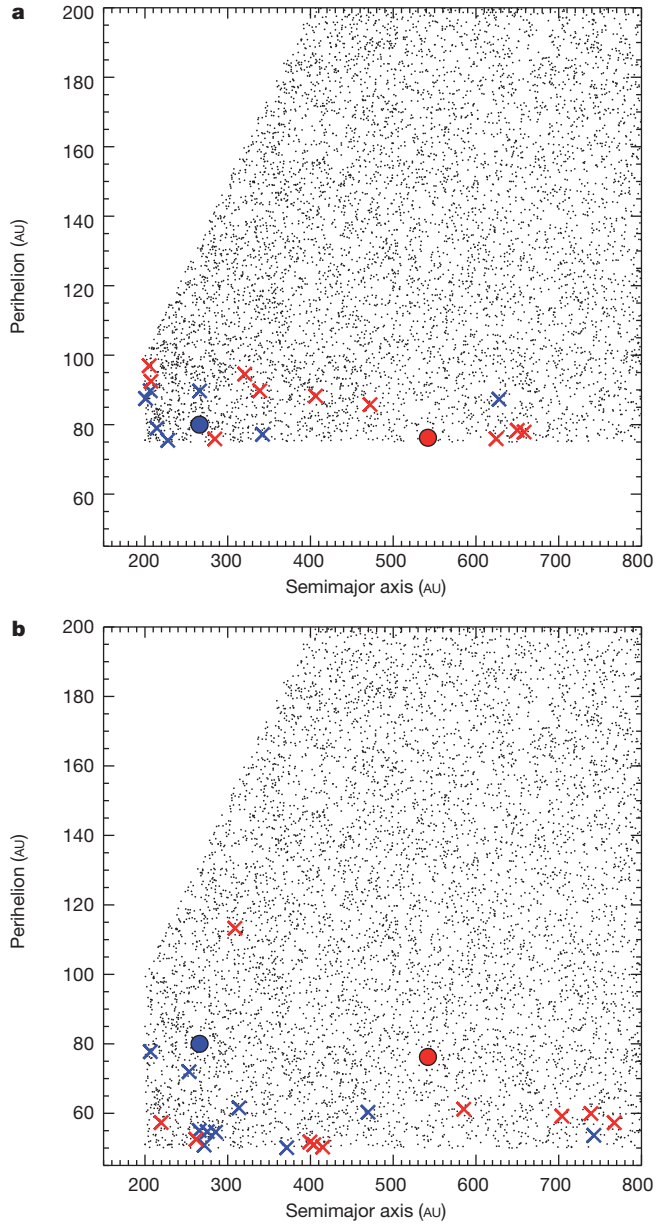


Figure 2 | Results from the observational bias simulation. The simulated inner Oort cloud objects (dots) are shown, along with the simulated detections in our survey (blue crosses) and the most sensitive all-sky survey so far, which re-detected Sedna[21] (red crosses) for two different inner Oort cloud object models: our favoured model (**a**) and a comparison model (**b**). We have displayed ten simulated survey realizations per survey. The true values for Sedna and 2012 VP_{113} are shown as red and blue circles, respectively, both with error bars much smaller than the plot symbol (there are no error bars associated with all other symbols, which describe simulated data). There is one difference between the two models, the minimum perihelion for the underlying population (75 AU in **a** versus 50 AU in **b**). Our observational results favour the model in **a** because both Sedna and 2012 VP_{113} were found within a few AU of their perihelion. This indicates that the inner Oort cloud objects may have increasing numbers with increasing distance.

small-body reservoir distributions[22,23] ($q' = 5$), there would be only a 1% chance of finding 2012 VP_{113} and Sedna with perihelion greater than 75 AU and no objects with perihelion less than 75 AU. Therefore, we conclude that there are few (although probably not zero) inner Oort cloud objects in the 50–75 AU region. Some stellar encounter models that include the capture of extrasolar material predict a strong inner edge to the perihelion distribution of objects, which is consistent with our observations[2].

Our simulation suggests that the inner Oort cloud contains about 1/80th of an Earth mass of material for our favoured $q' = 5$ model, although this can vary based on the assumed size distribution (Methods and Extended Data Table 1). This is similar to that of the Kuiper belt objects (KBOs), which have about 1/100th of an Earth mass of material[22]. We estimate that about $430^{+400}_{-240}(1\sigma)$ inner Oort cloud objects brighter than $r = 24.3$ remain undiscovered throughout the sky. In addition, $900^{+800}_{-500}(1\sigma)$ bodies are expected with diameter larger than 1,000 km (most of which are extremely distant), which is statistically consistent with detailed model estimates based only on the discovery of Sedna[21]. Our simulations assume that the inner Oort cloud objects are longitudinally symmetric, but it is interesting that both 2012 VP_{113} and Sedna were found at similar sky locations. If further inner Oort cloud object discoveries prove this effect to be statistically significant, this would strongly constrain inner Oort cloud object formation models.

Both Sedna and 2012 VP_{113} have similar arguments of perihelion ($\omega = 311^\circ$ and 293°, respectively). The orbital element ω describes the angle between the point of perihelion and where a body's orbit crosses the celestial plane. Surprisingly, this ω similarity is shared for all known objects with semi-major axes greater than 150 AU and perihelia greater than that of Neptune ($\omega \approx 340^\circ \pm 55^\circ$; see Fig. 3, Extended Data Fig. 1 and Extended Data Table 2). These extreme scattered disk objects may have a similar origin to Sedna or 2012 VP_{113} but their much lower perihelia ($q < 49$ AU) allow their existence to be explained through resonant interactions with Neptune, except for 2004 VN_{112}, 2010 GB_{174} and possibly 2000 CR_{105} (refs 10, 24–27). We find that this ω clustering cannot be due to observational bias for two reasons: (1) any bias for $\omega = 0^\circ$ is the same as the bias for $\omega = 180^\circ$; and (2) the surveys that have found the objects with clustered ω were often off-ecliptic or all-sky surveys, which would not have a bias for either $\omega = 0^\circ$ or 180° (Methods). Thus, we conclude that the ω clustering is a real effect.

The Lidov–Kozai effect is the best known dynamical mechanism for constraining the ω of a minor planet[28]. This three-body interaction can create outer Oort cloud objects with ω preferentially near 0° and 180° early in the history of the Solar System when the Sun was still among its nascent cohort of perturbing stars[29]. This cannot explain the $\omega \approx 0^\circ$ trend today, because ω circulates owing to the presence of the giant planets. By numerically simulating the effect of the known mass in the Solar System on the inner Oort cloud objects, we confirmed that inner Oort cloud objects should have random ω (Methods). This suggests that a massive outer Solar System perturber may exist and restricts ω for the inner Oort cloud objects. We numerically simulated the effect of a super-Earth-mass body at 250 AU and found that ω for inner Oort cloud objects librated around $0^\circ \pm 60^\circ$ for billions of years (see Extended Data Figs 2 and 3). This configuration is not unique and there are many possibilities for such an unseen perturber. A super-Earth-mass body at 250 AU with very low albedo would be fainter than current all-sky survey detection limits, as would larger and more distant perturbers[30].

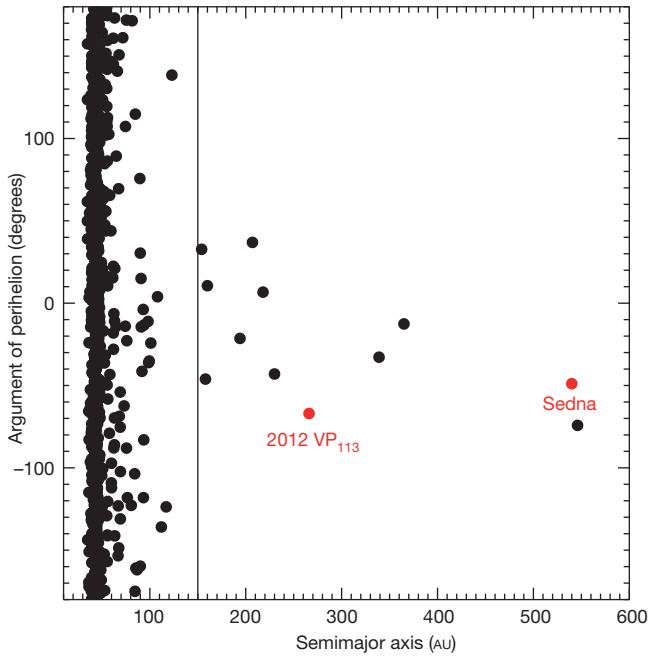


Figure 3 | The argument of perihelion for distant objects clusters about 0°. All minor planets with perihelion greater than 30 AU as a function of semi-major axis are shown. All bodies with semi-major axis greater than the line at 150 AU show a pronounced concentration near $\omega \approx 0^\circ$. Errors on these orbital elements are much smaller than the plotted symbols. This figure appears in histogram form in Extended Data Fig. 1.

Although such a perturber could keep the ω near 0°, how objects initially obtained $\omega \approx 0^\circ$ and not 180° is unknown. Stochastic events, such as a single strong stellar encounter or rogue planet could produce such asymmetric populations. A strong stochastic encounter could also produce the low semi-major axis of 2012 VP_{113}. It is possible that the ω clustering indicates that all $q > 30$ AU and $a > 150$ AU bodies formed by a similar process, although bodies on elliptical orbits with q less than approximately 45 AU can be explained by Neptune interactions. Models of galactic tides have difficulty producing inner Oort cloud objects having low semi-major axes (like Sedna) in the absence of stellar encounters[19] so the even lower semi-major axis of 2012 VP_{113} makes tides extremely unlikely. If the inner Oort cloud objects were created from the gravitational tide within our Sun's birth cluster, the low semi-major axis of 2012 VP_{113} constrains the birth cluster to have been a moderate 10,000 solar masses per cubic parsec or more, depending on the cluster model assumed[15,29]. The moderately red optical colours observed for 2012 VP_{113} (Methods and Extended Data Table 3) are consistent with formation in the gas giant region and not in the ultra-red dominated classical Kuiper belt. Assuming a moderate albedo, 2012 VP_{113} is 450 km in diameter, large enough to be considered a dwarf planet if it is composed largely of ice.

METHODS SUMMARY

For this work, we conducted two main observational investigations. First, we surveyed 52 square degrees of sky for new inner Oort cloud objects using the Dark Energy Camera (DECam) at the Cerro Tololo Inter-American Observatory (CTIO) 4-m telescope. Each patch of sky was imaged three times with 1.5 h to 2 h between images, giving us the ability to detect the motion of objects beyond 300 AU when compared to background stars. Analysis software aided our search for moving objects. After its discovery at the CTIO 4-m telescope, we measured the position and colour of 2012 VP_{113} using the Magellan 6.5-m telescope, completing the data collection portion of our project. We then created an observational bias simulation that simulated our discovery process to place constraints on the underlying number and orbital distribution of inner Oort cloud objects. This produces order-of-magnitude results for the number of inner Oort cloud objects whose basic parameters are poorly constrained because so few are known. Finally, after identifying the unusual nature of the ω distribution, we simulated the inner Oort cloud objects and the other minor planets with semi-major axes greater than 150 AU and perihelion greater than 30 AU to explore the long-term behaviour of ω for different hypothetical planetary masses in the Solar System. These methods are described in greater detail in the Methods section.

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 2013; accepted 10 February 2014.

1. Brown, M. E., Trujillo, C. & Rabinowitz, D. Discovery of a candidate inner Oort cloud planetoid. *Astrophys. J.* **617,** 645–649 (2004).
2. Morbidelli, A. & Levison, H. F. Scenarios for the origin of the orbits of the trans-Neptunian objects 2000 CR_{105} and 2003 VB_{12} (Sedna). *Astron. J.* **128,** 2564–2576 (2004).
3. Kenyon, S. J. & Bromley, B. C. Stellar encounters as the origin of distant Solar System objects in highly eccentric orbits. *Nature* **432,** 598–602 (2004).
4. Melita, M. D., Larwood, J. D. & Williams, I. P. Sculpting the outer Edgeworth Kuiper belt: stellar encounter followed by planetary perturbations. *Icarus* **173,** 559–573 (2005).
5. Gomes, R. S., Matese, J. J. & Lissauer, J. J. A distant planetary-mass solar companion may have produced distant detached objects. *Icarus* **184,** 589–601 (2006).
6. Levison, H. F. & Morbidelli, A. Models of the collisional damping scenario for ice-giant planets and Kuiper belt formation. *Icarus* **189,** 196–212 (2007).
7. Goldreich, P., Lithwick, Y. & Sari, R. Final stages of planet formation. *Astrophys. J.* **614,** 497–507 (2004).
8. Batygin, K., Brown, M. E. & Betts, H. Instability-driven dynamical evolution model of a primordially five-planet outer Solar System. *Astrophys. J.* **744,** L3 (2012).
9. Nesvorný, D. & Morbidelli, A. Statistical study of the early Solar System's instability with four, five, and six giant planets. *Astron. J.* **144,** 117 (2012).
10. Gomes, R. S., Gallardo, T., Fernández, J. A. & Brunini, A. On the origin of the high-perihelion scattered disk: the role of the Kozai mechanism and mean motion resonances. *Celestial Mech. Dyn. Astron.* **91,** 109–129 (2005).
11. Soares, J. S. & Gomes, R. S. Comparison of forming mechanisms for Sedna-type objects through an observational simulator. *Astron. Astrophys.* **553,** A110 (2013).
12. Gladman, B. & Chan, C. Production of the extended scattered disk by rogue planets. *Astrophys. J.* **643,** L135–L138 (2006).
13. Ida, S., Larwood, J. & Burkert, A. Evidence for early stellar encounters in the orbital distribution of Edgeworth-Kuiper belt objects. *Astrophys. J.* **528,** 351–356 (2000).
14. Brasser, R. A two-stage formation process for the Oort comet cloud and its implications. *Astron. Astrophys.* **492,** 251–255 (2008).
15. Brasser, R., Duncan, M. J., Levison, H. F., Schwamb, M. E. & Brown, M. E. Reassessing the formation of the inner Oort cloud in an embedded star cluster. *Icarus* **217,** 1–19 (2012).
16. Dukes, D. & Krumholz, M. R. Was the Sun born in a massive cluster? *Astrophys. J.* **754,** 56 (2012).
17. Pfalzner, S. Early evolution of the birth cluster of the solar system. *Astron. Astrophys.* **549,** A82 (2013).
18. Adams, F. C. The birth environment of the Solar System. *Annu. Rev. Astron. Astrophys.* **48,** 47–85 (2010).
19. Kaib, N. A., Roškar, R. & Quinn, T. Sedna and the Oort cloud around a migrating Sun. *Icarus* **215,** 491–507 (2011).
20. Levison, H. F., Duncan, M. J., Brasser, R. & Kaufmann, D. E. Capture of the Sun's Oort cloud from stars in its birth cluster. *Science* **329,** 187–190 (2010).
21. Schwamb, M. E., Brown, M. E., Rabinowitz, D. L. & Ragozzine, D. Properties of the distant Kuiper belt: results from the Palomar Distant Solar System Survey. *Astrophys. J.* **720,** 1691–1707 (2010).
22. Petit, J.-M. *et al.* The Canada-France Ecliptic Plane Survey—full data release: the orbital structure of the Kuiper belt. *Astron. J.* **142,** 131 (2011).
23. Schwamb, M. E., Brown, M. E. & Fraser, W. C. The small numbers of large Kuiper belt objects. *Astron. J.* **147,** 2 (2014).
24. Gladman, B. *et al.* Evidence for an extended scattered disk. *Icarus* **157,** 269–279 (2002).
25. Gomes, R. S., Fernández, J. A., Gallardo, T. & Brunini, A. in *The Solar System Beyond Neptune* (eds Barucci, M. A., Boehnhardt, H., Cruikshank, D. P., Morbidelli, A. & Dotson, R.) 259–273 (2008).
26. Becker, A. C. *et al.* Exploring the outer Solar System with the ESSENCE Supernova Survey. *Astrophys. J.* **682,** L53–L56 (2008).
27. Chen, Y.-T. *et al.* Discovery of a new member of the inner Oort cloud from the Next Generation Virgo Cluster Survey. *Astrophys. J.* **775,** L8 (2013).
28. Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.* **67,** 591 (1962).
29. Brasser, R., Duncan, M. J. & Levison, H. F. Embedded star clusters and the formation of the Oort cloud. *Icarus* **184,** 59–82 (2006).
30. Sheppard, S. S. *et al.* A southern sky and galactic plane survey for bright Kuiper belt objects. *Astron. J.* **142,** 98 (2011).

Acknowledgements We thank the Dark Energy Camera (DECam) team for obtaining observations during DECam commissioning, D. Norman for scheduling the November 2012 DECam observations, and D. Norman, A. Kunder and K. Holhjem for queue observing in November. T. Abbott and F. Valdes were very helpful during our December 2012 DECam observations. This project used data obtained with DECam, which was

constructed by the Dark Energy Survey collaborating institutions. Observations were in part obtained at the Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy, under contract with the National Science Foundation. This paper includes data gathered with the 6.5-m Magellan telescopes located at Las Campanas Observatory, Chile. This research was funded by NASA Planetary Astronomy grant NNX12AG26G and has also been supported by the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., on behalf of the international Gemini partnership of Argentina, Australia, Brazil, Canada, Chile and the USA.

Author Contributions C.T. created the moving object detection program for image analysis, developed the discovery statistic simulations, and is the principal investigator of the NASA grant supporting the project. S.S. obtained the telescope time, planned and performed the observations, analysed the data (including the colour measurements) and estimated the inner Oort cloud object orbital evolution using the Mercury integrator.

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 C.T. (trujillo@gemini.edu).

LETTER

doi:10.1038/nature13137

Quantum chaos in ultracold collisions of gas-phase erbium atoms

Albert Frisch[1], Michael Mark[1], Kiyotaka Aikawa[1], Francesca Ferlaino[1], John L. Bohn[2], Constantinos Makrides[3], Alexander Petrov[3,4,5] & Svetlana Kotochigova[3]

Atomic and molecular samples reduced to temperatures below one microkelvin, yet still in the gas phase, afford unprecedented energy resolution in probing and manipulating the interactions between their constituent particles. As a result of this resolution, atoms can be made to scatter resonantly on demand, through the precise control of a magnetic field[1]. For simple atoms, such as alkalis, scattering resonances are extremely well characterized[2]. However, ultracold physics is now poised to enter a new regime, where much more complex species can be cooled and studied, including magnetic lanthanide atoms and even molecules. For molecules, it has been speculated[3,4] that a dense set of resonances in ultracold collision cross-sections will probably exhibit essentially random fluctuations, much as the observed energy spectra of nuclear scattering do[5]. According to the Bohigas–Giannoni–Schmit conjecture, such fluctuations would imply chaotic dynamics of the underlying classical motion driving the collision[6–8]. This would necessitate new ways of looking at the fundamental interactions in ultracold atomic and molecular systems, as well as perhaps new chaos-driven states of ultracold matter. Here we describe the experimental demonstration that random spectra are indeed found at ultralow temperatures. In the experiment, an ultracold gas of erbium atoms is shown to exhibit many Fano–Feshbach resonances, of the order of three per gauss for bosons. Analysis of their statistics verifies that their distribution of nearest-neighbour spacings is what one would expect from random matrix theory[9]. The density and statistics of these resonances are explained by fully quantum mechanical scattering calculations that locate their origin in the anisotropy of the atoms' potential energy surface. Our results therefore reveal chaotic behaviour in the native interaction between ultracold atoms.

In the common perception, atoms are regarded as 'simple' systems in sharp contrast to 'complex' molecules, whose behaviour is dictated by many (rotational and vibrational) degrees of freedom. The recent realization of dipolar Bose–Einstein condensates and Fermi gases of magnetic lanthanides[10–13] made a novel class of atoms available in the ultracold regime. These exotic species, such as erbium (Er), make it possible to bridge the enormous conceptual gap between 'simple' atoms and molecules, potentially providing a natural test-bed with which to explore complex scattering dynamics in a controlled environment. The rich scattering behaviour of lanthanides has been pointed out in pioneering experiments at millikelvin temperatures[14,15] and theoretical work on cold collisions of atoms with non-zero angular momenta[16,17].

The wealth of intriguing properties of Er, which is the focus of this Letter, originates in its exotic electronic configuration. Erbium is a submerged-shell atom with electron vacancies in the inner anisotropic $4f^{12}$ shell, which lies beneath a filled $6s^2$ shell. As a consequence, it not only has a large magnetic moment, of 7 Bohr magnetons (μ_B), but also has large electronic orbital and total angular momentum quantum number, of $L = 5$ and $J = 6$, respectively; we note that for bosonic and fermionic isotopes the nuclear angular quantum numbers are $I = 0$ and $I = 7/2$, respectively. Large values of L and J are sources of anisotropy in the interatomic interaction. Moreover, the two-body scattering is controlled by as many as 91 electronic Born–Oppenheimer (B–O) interaction potentials, each potential accounting for a specific orientation of J with respect to the internuclear axis (Methods). All B–O potentials are anisotropic and include, at large internuclear separations, a strong dipole–dipole interaction and anisotropic van der Waals dispersion potentials. This situation is in contrast to that of conventional ultracold atoms, such as alkali-metal atoms, where the scattering is determined mainly by the isotropic singlet and triplet B–O potentials[2]. Recent theoretical work predicted the existence of anisotropy-induced Fano–Feshbach resonances in magnetic lanthanides[18]. This greater complexity brings significant new challenges in understanding and exploiting scattering processes.

Our experimental study is based on high-resolution trap-loss spectroscopy of Fano–Feshbach resonances in an optically trapped ultracold sample of Er atoms in the energetically lowest magnetic Zeeman sublevel. We prepare the ultracold sample by following a similar cooling and trapping approach to that described in ref. 12 for bosons and ref. 13 for fermions (Methods Summary). After the preparation procedure, the ultracold sample typically contains about 10^5 atoms at a temperature of around 400 nK. High-resolution trap-loss spectroscopy consists of many experimental cycles. In each cycle, we ramp the magnetic field to a target value, B, and hold the atoms for 400 ms in the optical dipole trap, during which time they undergo elastic and inelastic collisions. To probe the loss of atoms from the trap, we record the atom number by applying standard time-of-flight absorption imaging at zero magnetic field. In the next experimental cycle, we vary the magnetic field value from 0 to 70 G in steps of a few milligauss and repeat the measurement. Figure 1 shows the respective loss spectra for ^{168}Er and ^{166}Er. For both isotopes, we observe very many resonant loss features, which we interpret as being caused by Fano–Feshbach resonances[2]. We identify 190 resonances for ^{168}Er and 189 resonances for ^{166}Er, meaning that we observe about 3 resonances per gauss. We performed similar spectroscopic measurements with the fermionic isotope ^{167}Er, revealing a much higher density of resonances that exceeds 20 resonances per gauss (Extended Data Fig. 1). The fermionic case is complicated by its additional hyperfine structure and detailed studies will be subject of future work.

The very high density of resonances in Er is without precedent in ultracold quantum gases. For comparison, the density of resonances observed in experiments with ultracold alkali-metal atoms or even mixtures is about two orders of magnitude lower than that in Er (compare with refs 19, 20). In Er, it is unclear whether a quantitative mapping of the observed resonances is possible at all. In principle there are at least 91 unknown parameters, corresponding to the phase shifts introduced by the B–O potentials[18]. Instead, we focus our theoretical analysis on fundamental issues, such as whether the observed density of resonances can be reproduced by microscopic calculations, and whether

[1]Institut für Experimentalphysik, Universität Innsbruck, 6020 Innsbruck, Austria. [2]JILA, University of Colorado and National Institute of Standards and Technology, Boulder, Colorado 80309-0440, USA. [3]Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, USA. [4]St Petersburg Nuclear Physics Institute, Gatchina 188300, Russia. [5]Division of Quantum Mechanics, St Petersburg State University, St Petersburg 198904, Russia.

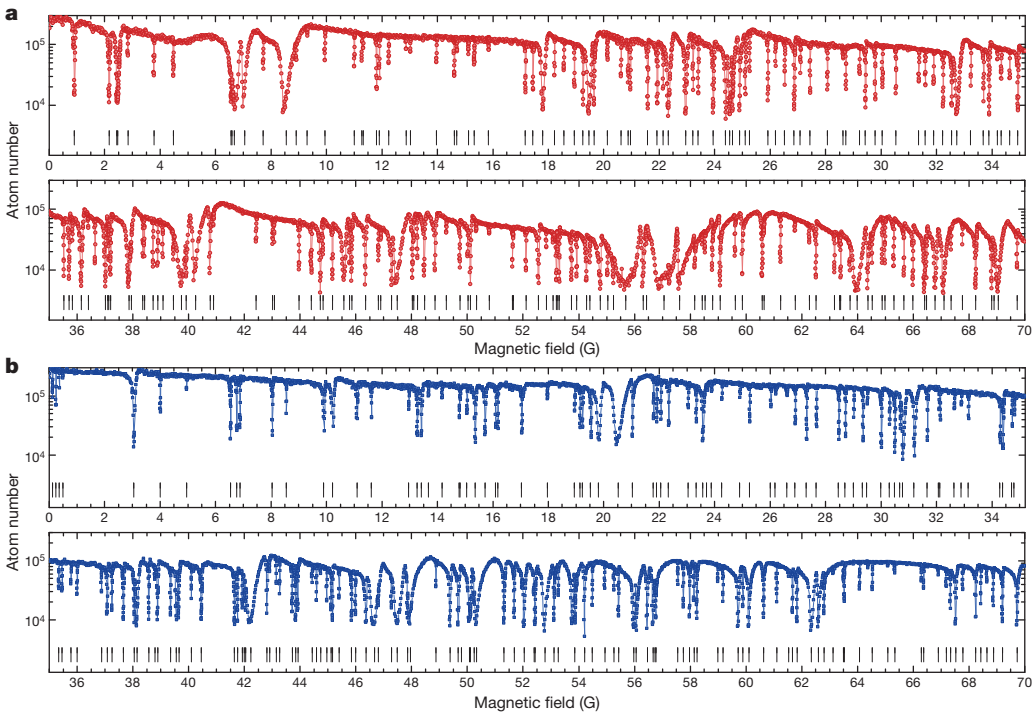


Figure 1 | Fano–Feshbach spectrum of ^{168}Er and ^{166}Er from 0 to 70 G. The trap-loss spectroscopy is performed in an optically trapped sample of Er atoms in the energetically lowest Zeeman sublevel, $m_J = -6$, at a temperature of 330 nK. The atom number is measured after a holding time of 400 ms. We observe 190 Fano–Feshbach resonances for ^{168}Er (**a**) and 189 resonances for ^{166}Er (**b**). Resonance positions are extracted by fitting a Gaussian shape to individual loss features; a full list is given in Extended Data Tables 1 and 2.

our results imply the presence of highly anisotropic interactions, which call into play resonant states of high orbital momentum. We answer these questions in the affirmative using full coupled-channel calculations, supported by an analytical model.

We construct a first-principles coupled-channel model for Er + Er scattering to calculate the spectrum of Fano–Feshbach resonances. Following ref. 18, our model uses the atomic basis set and Hamiltonian (Methods) that includes the radial kinetic and rotational energy operators, the Zeeman interaction and the 91 anisotropic B–O potentials. For small interatomic separations, R, the B–O potentials are calculated using the *ab initio* relativistic, multi-reference configuration–interaction method[21]. At intermediate to large values of R, the B–O potentials are expressed as a sum of multipolar interaction terms. The van der Waals dispersion interaction potentials ($\propto 1/R^6$) are determined from experimental data on atomic transition frequencies and oscillator strengths[22,23]. An important point is that the dispersion potentials have both isotropic and anisotropic contributions. The latter comes from the non-S state character of the Er electronic ground state. The B–O potentials thus induce either isotropic (ℓ and m_ℓ conserving) or anisotropic (ℓ or m_ℓ changing) couplings. Here, ℓ and m_ℓ are the partial-wave quantum number and its projection on the magnetic field quantization axis.

We perform coupled-channel calculations for bosonic ^{168}Er, considering s-wave ($\ell = 0$) collisions and couplings to molecular states with even ℓ up to $L_{\max} = 20$. We calculate the elastic collisional rate coefficient as a function of magnetic field to obtain the Fano–Feshbach resonance spectrum. For $L_{\max} = 20$, we observe a very dense resonance spectrum with about 1.5 resonances per gauss, which qualitatively reproduces our experimental observation (Extended Data Fig. 2). We note that resonances belonging to incident channels with $\ell > 0$ are substantially narrower and do not contribute much to the density of resonances. To get deeper insight into the role of the anisotropy of the potentials, we calculate the mean density of resonances, $\bar{\rho}$, from our coupled-channel calculations for different values of the maximum partial wave $L_{\max}$ (Fig. 2). For $L_{\max}$ up to 20, we observe that $\bar{\rho}$ increases with $L_{\max}$ in a quadratic manner. This dependence stands in stark contrast to alkali-metal atoms, where high-ℓ resonances tend to be too narrow to be observed.

Because our limited computational resources do not allow us to perform calculations for $L_{\max} > 20$, it is worth estimating the density of resonances in a simpler way, based on the separated atom quantum numbers[3]. The key ideas of our model are the following. For each channel $|j_1 m_{J,1}, j_2 m_{J,2}, \ell m_\ell\rangle$, we define the long-range potential $-C_6/R^6 + \hbar^2\ell(\ell+1)/2\mu R^2 + g\mu_B(m_{J,1} + m_{J,2})B$, where C_6 is the isotropic van der Waals coefficient of the B–O potentials and $\hbar$ is Planck's constant divided by 2π. Here μ is the reduced mass, g is the atomic g-factor and, for ground-state Er, $C_6 = 1{,}723$ atomic units (a.u.). Fano–Feshbach resonances in our open channel with $m_{J,1} = -6$ and $m_{J,2} = -6$ are due to couplings to the most weakly bound rovibrational level of closed channels. For a van der Waals potential[2,24] this bound state has a binding energy that must fall within the ℓ-dependent energy window $[-\Delta_\ell, 0]$ with $\Delta_\ell > 0$. The short-range potentials are not accurately known and, for each closed channel, there is a probability dE_b/Δ_ℓ of finding a bound state with a binding energy between E_b and $E_b + dE_b$. From ref. 24 and numerical simulations, we find $\Delta_\ell/E_{vdW} \approx 38.7 + 25.5\ell + 3.17\ell^2$, where $E_{vdW} = \hbar^2/2\mu x_{vdW}^2$ and $x_{vdW} = \sqrt[4]{2\mu C_6/\hbar^2}/2$. Each closed channel contributes $g\mu_B\delta m/\Delta_\ell$ to the mean resonance density, where $g\mu_B\delta m > 0$ is the magnetic-moment difference of the closed and open channels and δm is their difference in molecular projection quantum numbers. Adding the contributions for the closed channels gives the total mean resonance density. This simple counting technique, which we here name random quantum defect theory (RQDT), yields the mean density of resonances shown in Fig. 2. For $L_{\max} \leq 20$, the results of our analytic RQDT agrees very well with the exact coupled-channel calculations. For larger values of $L_{\max}$, the density of resonances keeps growing and eventually saturates to a value around $\rho \approx 4\,G^{-1}$, which is close to the one observed in the experiment. RQDT shows that at least 40 partial waves need to be considered to reproduce the experimental observations.

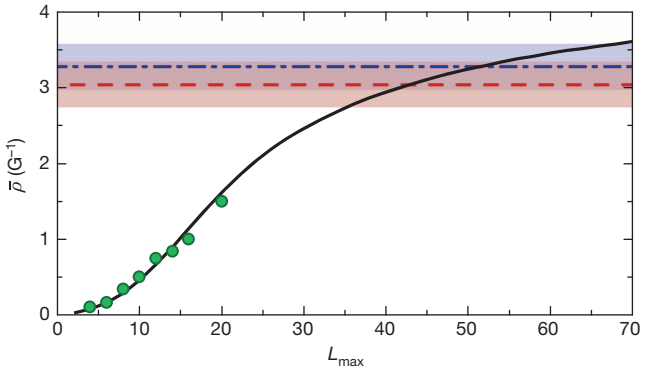


Figure 2 | Mean resonance density for bosonic Er as a function of largest included partial-wave L_{max}. Coupled-channel calculations for L_{max} up to 20 (circles) and RQDT calculation (solid line) for a magnetic field of 0 to 70 G. For calculations, a collision energy of $E = k_B \times (360\ \mathrm{nK})$ is assumed. The mean densities of resonances measured in the experiment are shown for ^{168}Er (dashed line) and for ^{166}Er (dash–dot line) with 1σ confidence bands (shaded areas).

Our microscopic models reproduce well the qualitative behaviour of the system. However, given the complexity of the scattering, the analysis of ultracold collision data can and should no longer aim to assign individual resonances, and the fundamental question of how to tackle complex scattering arises. Historically, spectra of great complexity have been understood within the framework of random matrix theory (RMT), as originally developed by Wigner to describe heavy nuclei containing a very large number of degrees of freedom[25]. This is an alternative view of the quantum mechanics of complex systems, where individual energy levels and resonances are not theoretically reproduced one by one, yet their statistics can be described[26]. RMT characterizes spectra by fluctuations of their energy levels and classifies their statistical behaviour in terms of symmetry classes; for example, the Gaussian-orthogonal ensemble (GOE) is appropriate in the case of a system with time-reversal symmetry, such as neutral atoms.

Following RMT, the distribution of spacings between neighbouring levels (or resonances) characterizes the spectral fluctuations of the system and reflects the absence or the presence of level correlations in terms of a dimensionless quantity, $s = \delta B/\bar{d}$, where δB is the space between any two adjacent levels and $\bar{d} = 1/\bar{\rho}$ is the mean level spacing. Whereas the nearest-neighbour spacing (NNS) distribution $P(s)$ of non-interacting levels is Poissonian, $P_P = \exp(-s)$, strongly interacting levels obey a totally different distribution, which, in the case of GOE statistics, is known as the Wigner–Dyson (W–D) distribution, or 'Wigner surmise'[26]

$$P_{WD} = \frac{\pi}{2} s \exp\left(-\frac{\pi}{4} s^2\right) \quad (1)$$

which shows a strong level repulsion for small s: $P_{WD}(0) = 0$. The field of application of the W–D distribution is so vast as to make it a universal feature of very complex systems, such as heavy nuclei, disordered conductors, zeros of the Riemann function in number theory, and even risk management models in finance[5,8]. Remarkably, the Bohigas–Giannoni–Schmit conjecture further enriched the field of applications of GOE statistics[6,27], showing that it applies generally to chaotic systems, such as Rydberg atoms in strong magnetic fields and Sinai billiards, where only a few degrees of freedom are relevant but where motion in these degrees of freedom occurs on a highly anisotropic potential energy surface[7]. Recently, it has been speculated that even cold and ultracold atom–molecule collisions will show essential features of GOE statistics[3,4].

Inspired by these works, we statistically analyse both the experimental and calculated Fano–Feshbach spectrum according to RMT. To obtain the NNS distribution of resonances, we first derive $\bar{\rho}$ and the mean spacing between resonances, $\bar{d}$, by constructing the 'staircase function'[7]. This step-like function counts the number of resonances below a magnetic field value B and is defined as $\mathcal{N}(B) = \int_0^B \mathrm{d}B' \sum_i \delta(B' - B_i)$, where δ is the delta function and B_i is the position of the ith resonance. For our experimental data (Fig. 3a), the staircase function shows an increase of the number of resonances with B, which is linear for large B and flattens out towards lower magnetic field values (Fig. 3b). The density of resonances is given by the derivative of the staircase function. We evaluate $\bar{\rho}$ in the region above 30 G, where the staircase function shows a linear relationship (Methods), and we obtain $\bar{\rho} = 3.0(3)\ \mathrm{G}^{-1}$ and $\bar{d} = 0.33(3)$ G. We perform a similar analysis with ^{166}Er and find that $\bar{\rho} = 3.3(3)\ \mathrm{G}^{-1}$ and that $\bar{d} = 0.31(3)$ G (Extended Data Fig. 3). For data from coupled-channel calculations, we find that $\bar{\rho} = 1.5\ \mathrm{G}^{-1}$ for $L_{max} = 20$ (Fig. 2). Finally, we derive the NNS distribution for the experimental and coupled-channel data by constructing a histogram of resonance spacings. We choose a number of bins of the order of $\sqrt{N}$, with N being the number of Fano–Feshbach resonances used for analysis[28]. We then rescale the histogram in terms of s and normalize the distribution to obtain $P(s)$.

Figure 4 is the main result of our statistical analysis for ^{168}Er. The plot shows the NNS distribution of the experimental and coupled-channel Fano–Feshbach resonances, together with the parameter-free Poisson and W–D distributions (equation (1)). We see an impressive agreement between the experimental result and the coupled-channel calculations. Remarkably, both follow a distribution much closer to the W–D one than to the Poissonian one. To quantify the agreement with the GOE statistics, we evaluate the reduced χ^2 statistic, $\tilde{\chi}^2$, between our data and the Poisson and W–D distributions. We find that $\tilde{\chi}^2_{WD} = 0.9$ and $\tilde{\chi}^2_P = 2.3$ for our experimental data and that $\tilde{\chi}^2_{WD} = 0.8$ and $\tilde{\chi}^2_P = 3.0$ for the data of the coupled-channel calculations. The fact that $\tilde{\chi}^2_{WD} \leq 1$ confirms that our data are well described by a W–D distribution. Similar results are found for ^{166}Er (Extended Data Fig. 4).

To investigate the spectral correlations further, we analyse our data in terms of other statistical quantities, such as the number variance and the two-gap distribution function[29] (Methods). The number variance, $\Sigma^2(\Delta B)$, measures the fluctuations in the number of resonances in a magnetic field interval ΔB (ref. 7; Methods). For uncorrelated (Poisson-distributed) levels, $\Sigma^2 = \Delta B$, indicating large fluctuations around a mean value. For quantum chaotic systems, the correlations are strong and the fluctuations are thus less spread out. In this case, $\Sigma^2 \propto \ln(\Delta B)$. This slower increase of the number variance is regarded as indicating the strong 'spectral rigidity' of the system[7]. Our observations clearly

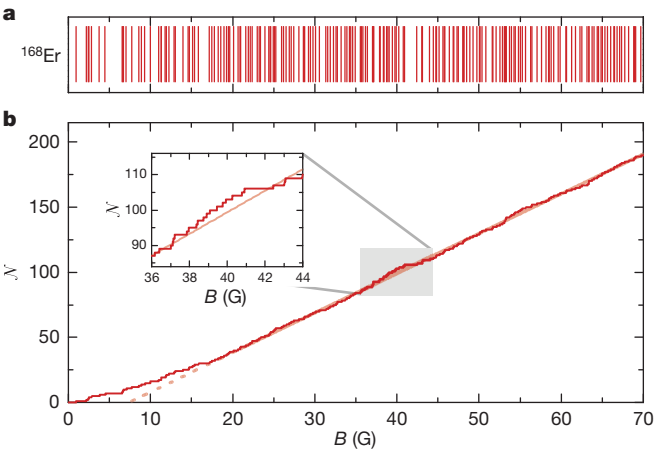


Figure 3 | Loss-maxima position and staircase function for ^{168}Er. **a**, Positions of the measured loss maxima of Fig. 1 are shown as vertical lines. **b**, The staircase function shows a linear dependence on the magnetic field at large values. A linear fit to the data above 30 G is plotted in a lighter colour. Inset, magnification of the data to emphasize the step-like nature of the staircase function.

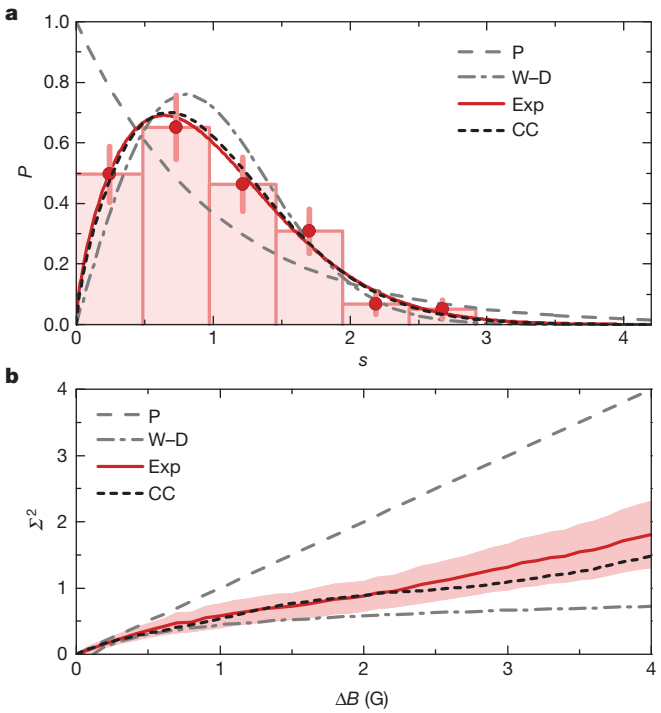


Figure 4 | NNS distribution and number variance. a, ^{168}Er NNS distribution above 30 G with a bin size of 160 mG. The plot shows the experimental data (circles) with the corresponding Brody distribution (which interpolates between Poisson and Wigner-Dyson distributions) (solid line), the Brody distribution for the coupled-channel calculation with $L_{max} = 20$ (short-dashed line), and the parameter-free distributions P_P (long-dashed line) and P_{WD} (dash–dot line). The Brody distribution is given in Methods. For the error bars in the experimental data, we assume a Poisson counting error. **b**, Number variances for the experimental data (solid line) with a 2σ confidence band (shaded area), the coupled-channel data (short-dashed line), P_P (long-dashed line) and P_{WD} (dash–dot line).

deviate from the Poissonian behaviour, showing that Σ^2 tends to the W–D case (Fig. 4b), and confirm the presence of correlations in our system.

Our observations reproduce the salient features predicted by GOE statistics for chaotic systems, the level repulsion and the spectral rigidity. This implies a degree of complexity in Er + Er cold collisions unprecedented in any previous ultracold scattering system. Our results bring the powerful analytical tools of quantum chaos to bear[7]. In particular, these approaches connect the large-scale structure of the spectra to simple features such as the shortest closed classical orbits in the potential energy surface, where these connections are made by the Gutzwiller trace formula[30]. Identifying the most important closed orbits will then shed light on the potential energy surface itself, providing a route to describing ultracold collisions that is complementary to the elaborate close-coupling calculations, which will be difficult to connect in detail with the data.

Erbium systems are the first in which statistical analyses and chaotic behaviour are important to ultracold collisions, but they will not be the last. Specifically, much experimental effort is being exerted on producing ultracold molecular samples, which also have highly anisotropic potential energy surfaces. Learning to understand complex spectra, by acknowledging their essentially chaotic nature, represents a turning point in how the field will consider ultracold collisions in future and provides new inroads into ultracold chemistry.

METHODS SUMMARY

For bosonic sample preparation, we follow the approach of ref. 12. We obtain about 3×10^5 optically trapped atoms at a density of 3×10^{13} cm^{-3}. The trap-loss spectroscopy is performed in a trap with frequencies of $(\nu_x, \nu_y, \nu_z) = (65, 26, 270)$ Hz. The temperature of the cloud is measured by time-of-flight imaging at 0.4 G, which yields $T_{168} = 326(4)$ nK for ^{168}Er and $T_{166} = 415(4)$ nK for ^{166}Er. We ramp the magnetic field to a probe value between 0 and 70 G within 10 ms, and hold the atomic cloud in the optical dipole trap for 400 ms. We observe an increase in the temperature to 560 nK at a magnetic field of about 50 G due to the ramping over many Fano–Feshbach resonances. For fermionic sample preparation, we follow the approach of ref. 13. We obtain about 1.2×10^5 fermionic atoms at a density of 2×10^{14} cm^{-3} and at a temperature of $0.4T_F$, where $T_F = 1.0(1)$ μK is the Fermi temperature. The trap frequencies are $(\nu_x, \nu_y, \nu_z) = (427, 66, 457)$ Hz.

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 6 December 2013; accepted 6 February 2014.
Published online 12 March 2014.

1. Inouye, S. *et al.* Observation of Feshbach resonances in a Bose–Einstein condensate. *Nature* **392,** 151–154 (1998).
2. Chin, C., Grimm, R., Julienne, P. S. & Tiesinga, E. Feshbach resonances in ultracold gases. *Rev. Mod. Phys.* **82,** 1225–1286 (2010).
3. Mayle, M., Ruzic, B. P. & Bohn, J. L. Statistical aspects of ultracold resonant scattering. *Phys. Rev. A* **85,** 062712 (2012).
4. Mayle, M., Quéméner, G., Ruzic, B. P. & Bohn, J. L. Scattering of ultracold molecules in the highly resonant regime. *Phys. Rev. A* **87,** 012709 (2013).
5. Guhr, T., Müller-Groeling, A. & Weidenmüller, H. A. Random-matrix theories in quantum physics: common concepts. *Phys. Rep.* **299,** 189–425 (1998).
6. Bohigas, O., Giannoni, M. J. & Schmit, C. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Phys. Rev. Lett.* **52,** 1–4 (1984).
7. Weidenmüller, H. A. & Mitchell, G. E. Random matrices and chaos in nuclear physics: nuclear structure. *Rev. Mod. Phys.* **81,** 539–589 (2009).
8. Mitchell, G. E., Richter, A. & Weidenmüller, H. A. Random matrices and chaos in nuclear physics: nuclear reactions. *Rev. Mod. Phys.* **82,** 2845–2901 (2010).
9. Brody, T. A. A statistical measure for the repulsion of energy levels. *Lett. N. Cim. Ser. 2* **7,** 482–484 (1973).
10. Lu, M., Burdick, N. Q., Youn, S. H. & Lev, B. L. Strongly dipolar Bose-Einstein condensate of dysprosium. *Phys. Rev. Lett.* **107,** 190401 (2011).
11. Lu, M., Burdick, N. Q. & Lev, B. L. Quantum degenerate dipolar Fermi gas. *Phys. Rev. Lett.* **108,** 215301 (2012).
12. Aikawa, K. *et al.* Bose-Einstein condensation of erbium. *Phys. Rev. Lett.* **108,** 210401 (2012).
13. Aikawa, K. *et al.* Reaching Fermi degeneracy via universal dipolar scattering. *Phys. Rev. Lett.* **112,** 010404 (2014).
14. Hancox, C. I., Doret, S. C., Hummon, M. T., Luo, L. & Doyle, J. M. Magnetic trapping of rare-earth atoms at millikelvin temperatures. *Nature* **431,** 281–284 (2004).
15. Connolly, C. B., Au, Y. S., Doret, S. C., Ketterle, W. & Doyle, J. M. Large spin relaxation rates in trapped submerged-shell atoms. *Phys. Rev. A* **81,** 010702 (2010).
16. Kokoouline, V., Santra, R. & Greene, C. H. Multichannel cold collisions between metastable Sr atoms. *Phys. Rev. Lett.* **90,** 253201 (2003).
17. Krems, R. V., Groenenboom, G. C. & Dalgarno, A. Electronic interaction anisotropy between atoms in arbitrary angular momentum states. *J. Phys. Chem. A* **108,** 8941–8948 (2004).
18. Petrov, A., Tiesinga, E. & Kotochigova, S. Anisotropy-induced Feshbach resonances in a quantum dipolar gas of highly magnetic atoms. *Phys. Rev. Lett.* **109,** 103002 (2012).
19. Berninger, M. *et al.* Feshbach resonances, weakly bound molecular states, and coupled-channel potentials for cesium at high magnetic fields. *Phys. Rev. A* **87,** 032517 (2013).
20. Takekoshi, T. *et al.* Towards the production of ultracold ground-state RbCs molecules: Feshbach resonances, weakly bound states, and the coupled-channel model. *Phys. Rev. A* **85,** 032506 (2012).
21. Kotochigova, S., Levine, H. & Tupitsyn, I. Correlated relativistic calculation of the giant resonance in the Gd^{3+} absorption spectrum. *Int. J. Quantum Chem.* **65,** 575–584 (1997).
22. Kramida, A. *et al.* NIST Atomic Spectra Database Version 5.0. http://www.nist.gov/pml/data/asd.cfm (2013).
23. Lawler, J. E., Wyart, J. & Den Hartog, E. A. Atomic transition probabilities of Er I. *J. Phys. At. Mol. Opt. Phys.* **43,** 235001 (2010).
24. Gao, B. Zero-energy bound or quasibound states and their implications for diatomic systems with an asymptotic van der Waals interaction. *Phys. Rev. A* **62,** 050702(R) (2000).
25. Wigner, E. P. On a class of analytic functions from the quantum theory of collisions. *Ann. Math.* **53,** 36–67 (1951).
26. Dyson, F. J. & Mehta, M. L. Statistical theory of the energy levels of complex systems. IV. *J. Math. Phys.* **4,** 701–712 (1963).
27. Alt, H. *et al.* Gaussian orthogonal ensemble statistics in a microwave stadium billiard with chaotic dynamics: Porter-Thomas distribution and algebraic decay of time correlations. *Phys. Rev. Lett.* **74,** 62–65 (1995).
28. Taylor, J. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements* 272 (Univ. Science Books, 1997).
29. Brody, T. A. *et al.* Random-matrix physics: spectrum and strength fluctuations. *Rev. Mod. Phys.* **53,** 385–479 (1981).

30. Gutzwiller, M. *Chaos in Classical and Quantum Mechanics* (Springer, 1990).

Acknowledgements The Innsbruck group thanks R. Grimm for discussions and S. Baier, C. Ravensbergen and M. Brownnutt for reading the manuscript. S.K. and A.P. thank E. Tiesinga for discussions. J.L.B. is supported by an ARO MURI. The Innsbruck team is supported by the Austrian Science Fund (FWF) through a START grant under project no. Y479-N20 and by the European Research Council under project no. 259435. K.A. is supported within the Lise-Meitner program of the FWF. Research at Temple University is supported by AFOSR and NSF PHY-1308573.

Author Contributions A.F., M.M., K.A. and F.F. did the experimental work and statistical analysis of the data. C.M., A.P. and S.K. did the theoretical work on coupled-channel calculations and RQDT. J.L.B. did the theoretical work on RMT. The manuscript was written with substantial contributions from all authors.

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

LETTER

doi:10.1038/nature13123

Reconciliation of the carbon budget in the ocean's twilight zone

Sarah L. C. Giering[1,2,3], Richard Sanders[1], Richard S. Lampitt[1], Thomas R. Anderson[1], Christian Tamburini[4], Mehdi Boutrif[4], Mikhail V. Zubkov[1], Chris M. Marsay[2,5], Stephanie A. Henson[1], Kevin Saw[1], Kathryn Cook[6] & Daniel J. Mayor[3]

Photosynthesis in the surface ocean produces approximately 100 gigatonnes of organic carbon per year, of which 5 to 15 per cent is exported to the deep ocean[1,2]. The rate at which the sinking carbon is converted into carbon dioxide by heterotrophic organisms at depth is important in controlling oceanic carbon storage[3]. It remains uncertain, however, to what extent surface ocean carbon supply meets the demand of water-column biota; the discrepancy between known carbon sources and sinks is as much as two orders of magnitude[4–8]. Here we present field measurements, respiration rate estimates and a steady-state model that allow us to balance carbon sources and sinks to within observational uncertainties at the Porcupine Abyssal Plain site in the eastern North Atlantic Ocean. We find that prokaryotes are responsible for 70 to 92 per cent of the estimated remineralization in the twilight zone (depths of 50 to 1,000 metres) despite the fact that much of the organic carbon is exported in the form of large, fast-sinking particles accessible to larger zooplankton. We suggest that this occurs because zooplankton fragment and ingest half of the fast-sinking particles, of which more than 30 per cent may be released as suspended and slowly sinking matter, stimulating the deep-ocean microbial loop. The synergy between microbes and zooplankton in the twilight zone is important to our understanding of the processes controlling the oceanic carbon sink.

The global carbon cycle is affected by biological processes in the oceans, which export carbon from surface waters in the form of organic matter and store it at depth, in a process called the 'biological carbon pump'. Most of the exported organic carbon is processed by the water-column biota, which ultimately convert it into CO_2 by means of respiration (remineralization). Variations in the resulting decrease in organic flux with depth[9] can, according to models, lead to changes in atmospheric CO_2 of up to 200 p.p.m. (ref. 3), indicating a strong coupling between biological activity in the ocean interior and oceanic storage of CO_2.

A key constraint in the analysis of carbon fluxes in the twilight zone is that, in the steady state, the attenuation of particulate organic carbon (POC) flux with depth should be balanced by community metabolism. Published estimates of POC flux attenuation with depth are, however, up to two orders of magnitude lower than corresponding estimates of heterotrophic metabolism[4–7]. This discrepancy indicates either that estimates of POC flux, community metabolism, or both, are unreliable, or that additional, unaccounted for, sources of organic carbon to the twilight zone exist[8].

We compiled a comprehensive carbon budget of the twilight zone on the basis of an extensive programme of field measurements at the Porcupine Abyssal Plain (PAP) site (Extended Data Fig. 1a) in July and August 2009. This site is located in the transition region between the subtropical and subpolar gyres of the North Atlantic[10]. The mixed-layer depth remained constant at approximately 50 m throughout the study period. This depth was subsequently used as the upper boundary of the twilight zone, given the need to normalize export measurements to dynamic upper boundaries for the twilight zone[11].

Organic carbon sources to the twilight zone include sinking particles, downward mixing of dissolved organic carbon (DOC), lateral advection of organic matter from the continental shelf, active transport via the daily vertical migration of zooplankton that feed in the mixed layer at night and rest at depth during the day, and chemolithoautotrophy (prokaryotic growth using dissolved inorganic carbon and chemical energy sources).

The downward flux of sinking particles was measured using simultaneous 48-h deployments of free-drifting, neutrally buoyant sediment traps[12] at depths of 50, 150, 300, 450 and 600 m (Extended Data Table 1). Satellite chlorophyll imagery and horizontal velocities (obtained using a 150-kHz vessel-mounted acoustic Doppler current profiler) confirmed that all of the traps were advected along the edge of an anticyclonic eddy for 50 km before surfacing within 3.5 km of each other. The measured POC flux at 50 m (84 ± 8 mg C m^{-2} d^{-1}) was close to estimates independently derived using ^{234}Th budgets and studies of collected marine snow particles[13] (99 ± 41 and 146 ± 26 mg C m^{-2} d^{-1}, respectively). Attenuation of POC flux (F) with depth (z) was fitted to the Martin curve[9] ($F_z = F_{MLD}(z/MLD)^b$), where F_{MLD} is the flux at the bottom of the mixed layer, MLD is the depth of the bottom of the mixed layer and b is the rate of attenuation. The observed attenuation rate ($b = -0.70$; $P < 0.01$, $R^2 = 0.95$, $n = 5$) was consistent with observations in the Pacific Ocean[9,14] ($b = -0.50$ to -1.38; Fig. 1a). Downward POC flux was extrapolated to 1,000 m using $b = -0.70$. The total loss of POC within the twilight zone was 74 ± 9 mg C m^{-2} d^{-1}.

DOC input to the twilight zone was estimated to be 15 mg DOC m^{-2} d^{-1} (range, 0.4–30 mg DOC m^{-2} d^{-1}) on the basis both of the ratio between DOC concentrations and apparent oxygen utilization[15], and of DOC gradients coupled to turbulent diffusivity measured from previous work at the study site[16] (Methods and Extended Data Fig. 2). DOC was estimated to supply 17% of total export, in agreement with previous estimates of 9–20% across the North Atlantic basin[17]. Organic matter input through lateral advection was assumed to be negligible from analyses of back-trajectories (derived from satellite-derived near-surface velocities over 3 months) of the water masses arriving at the PAP site during the study period, which suggested that the water had not passed over the continental slope (Extended Data Fig. 1b). The final source of DOC, namely excretion at depth by active flux, was estimated using net samples of zooplankton biomass and allometric equations[6,18], giving a supply of 3 mg C m^{-2} d^{-1}. Defecation and mortality at depth present further sources of organic carbon to the twilight zone, but these were excluded from the budget owing to large uncertainties associated with their estimation. Finally, chemolithoautotrophy has been suggested to be an important source of organic matter in the deep ocean[19], but without strong evidence that this poorly understood process could provide a major contribution at our study site, we chose to exclude it from our carbon budget.

The remineralization of organic carbon by zooplankton and prokaryotes was estimated from zooplankton biomass and prokaryotic activity.

[1]National Oceanography Centre, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK. [2]Ocean and Earth Sciences, University of Southampton, European Way, Southampton SO14 3ZH, UK. [3]Institute of Biological and Environmental Sciences, Oceanlab, University of Aberdeen, Newburgh AB41 6AA, UK. [4]Aix-Marseille Université, Université de Toulon, CNRS/INSU, IRD, MIO, UM 110, 13288 Marseille Cedex 09, France. [5]Department of Earth and Ocean Sciences, University of South Carolina, Columbia, South Carolina 29208, USA. [6]Marine Laboratory, Marine Scotland Science, Scottish Government, PO Box 101, 375 Victoria Road, Aberdeen AB11 9DB, UK.

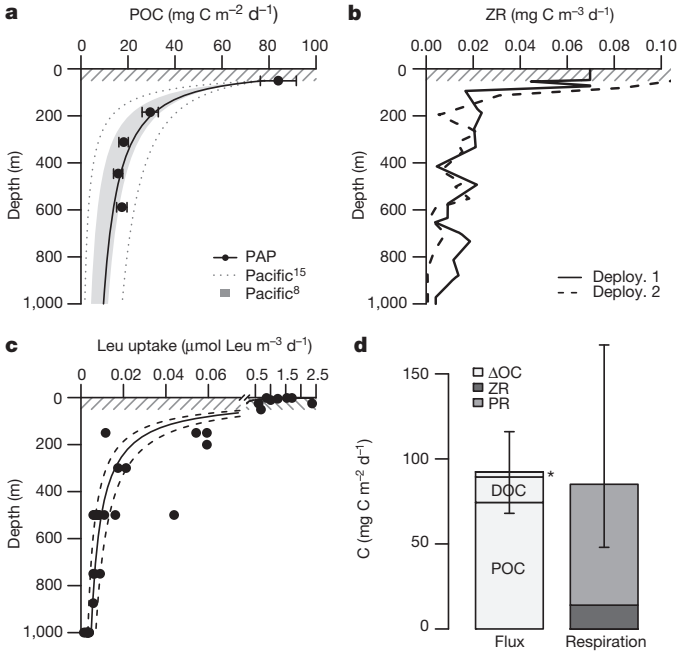


Figure 1 | Sinks and sources of organic carbon to the twilight zone. **a**, POC flux (black dots) below the mixed layer (shaded area) at the PAP site during 3–6 August 2009, fitted to the Martin equation ($F_z = F_{50}(z/50)^b$; solid line; $F_{50} = 78$ mg C m^{-2} d^{-1}, $b = -0.70$; $P < 0.01$, $R^2 = 0.95$, $n = 5$). The observed attenuation is consistent with rates observed in the Pacific[9,14] (grey area, dotted lines). Error bars show analytical error (s.d.). **b**, **c**, Depth profiles of respiration by non-migratory zooplankton (ZR) based on biomass samples ($n = 58$; **b**) and leucine incorporation (μmol Leu m^{-3} d^{-1}) by prokaryotes (power-law fit (solid) and interquartile range (dashed); $P < 0.001$, $R^2 = 0.86$, $n = 37$; **c**). **d**, The sum of net organic carbon supply (ΔOC; light grey) of POC, DOC and active flux (asterisk) matches respiration by non-migratory zooplankton (ZR; dark grey) and prokaryotes (PR; mid grey). Error bars represent upper and lower estimates (see text and Table 1).

It is crucial to note that in a steady-state system, such as we assume this to be, organic carbon is lost from the system only by export or by remineralization. We focus entirely on community respiration as a measure of remineralization, which is a fundamental advance over previous methods to derive budgets (Methods).

Zooplankton respiration was estimated by applying allometric relationships[6] to biomass measurements derived from net samples collected vertically every 80 m, twice during both day and night, using the ARIES net system fitted with 200-μm cod-ends (the narrow, exchangeable ends of the nets, which retain the samples) (Extended Data Table 1 and Extended Data Fig. 3). These allometric relationships are well constrained[6], but they are based on epipelagic zooplankton and our calculated respiration rates for the lower mesopelagic are therefore probably overestimates of the true rates[20]. Zooplankton resident in the twilight zone, mostly detritivorous copepods (*Oithona* and *Oncaea*) and carnivorous chaetognaths, had combined respiration rates of 15.2 and 12.7 mg C m^{-2} d^{-1} (50–1,000 m), respectively, during the two deployment periods (Fig. 1b). Migrating zooplankton (determined as the difference between day and night biomasses) were excluded from these estimates because we assume that they ingest sufficient carbon during grazing at the surface to satisfy their diagnosed respiration rates at depth (Methods). The organic carbon they respire within the twilight zone is thus imported by daily vertical migration.

Prokaryotic heterotrophic production was determined using bioassay isotope-dilution techniques with ^{3}H-leucine tracer[21]. Leucine incorporation rates were 41.7 ± 21.2 nmol Leu m^{-3} d^{-1} at 150 m and 6.6 ± 4.1 nmol Leu m^{-3} d^{-1} at 500–750 m (Fig. 1c), similar to previous estimates in the eastern North Atlantic[19] (37.7 and 7.5 nmol Leu m^{-3} d^{-1}, respectively). Integrated leucine incorporation based on a power-law fit was 14.5 μmol Leu m^{-2} d^{-1} (interquartile range, 13.2–16.1 μmol Leu m^{-2} d^{-1}; $P < 0.001$, $R^2 = 0.86$, $n = 37$). This fit was chosen on the assumption that bacterial activity follows the supply of organic carbon[22], although we lack data from between 50 and 150 m to confirm this fit. The uncertainty in this interpolation possibly leads to a misestimate of integrated leucine incorporation. Integrated leucine incorporation was converted into respiration using leucine-to-carbon conversion factors (0.44 ± 0.27 kg C mol^{-1} Leu) and growth efficiencies (interquartile range, 0.04–0.12) specific to the twilight zone derived from thorough literature surveys (Methods and Extended Data Fig. 4). The uncertainty in this calculation was estimated by bootstrap analysis with 100,000 simulations. The final estimate for integrated (50–1,000 m) prokaryotic respiration was 71 mg C m^{-2} d^{-1} (interquartile range, 35–152 mg C m^{-2} d^{-1}).

The sum of the inputs from POC and DOC matches community respiration (68–116 versus 48–167 mg C m^{-2} d^{-1}; Fig. 1d), with prokaryotes dominating community respiration (70–92%; Table 1).

Our study successfully reconciles the various components of the carbon budget in the twilight zone of the ocean. This was possible because we considered a dynamic upper boundary for the twilight zone (the base of the mixed layer), excluded vertical migrators from the estimate of zooplankton respiration in the twilight zone, and compared respiration rather than carbon demand to net organic carbon supply. Depth-resolved estimates of supply and consumption (Extended Data Fig. 5) show an excess of supply in the upper twilight zone (50–150 m) and a deficit in the lower twilight zone (150–1,000 m). We suggest that this may be caused by a subtle vertical change in ecosystem structure with depth[23,24] or an unaccounted-for vertical transfer of organic carbon between the upper and lower twilight zones.

The suggestion that prokaryotes dominate community respiration seems counterintuitive given that organic carbon supply to the twilight zone is dominated by sinking particles that are accessible to larger (>200 μm) zooplankton. We therefore propose that one of the main roles of zooplankton in the twilight zone is to mechanically degrade particulate material[25] into slow-sinking particulate matter and dissolved organic material that is subsequently remineralized by microbes (prokaryotes and their consumers).

To explore whether this conceptual picture is consistent with our present understanding of twilight-zone ecology, and to provide a full quantitative picture of the twilight-zone carbon cycle, we used a simple steady-state model of that cycle[26]. The model traces the turnover and remineralization of sinking POC along three pathways: colonization and solubilization of detritus by attached microbes, production of free-living microbes following loss of solubilization products during particle degradation, and consumption by detritivorous zooplankton (Methods and Extended Data Fig. 6a). The model was modified to include vertical mixing of DOC and active transport as carbon inputs to the twilight zone and to represent POC in both sinking and suspended forms, the latter produced via zooplankton 'sloppy feeding'[27] (leakage during ingestion). Inputs of carbon to the twilight zone were the measured values given in Table 1.

Modelled respiration rates matched field data well, with 84% of the CO_2 being produced by microbes (prokaryotes and prokaryote consumers) and only 16% by zooplankton (detritivores and carnivores) (Fig. 2). The model further suggests that microzooplankton respiration,

Table 1 | Carbon budget for the twilight zone (50–1,000 m)

Input		Respiration		Community respiration (%)
Sinking POC	74 (65–83)	Zooplankton	14 (13–15)	16 (8–30)
Vertical mixing (DOC)	15 (0–30)			
Active transport (DOC)	3	Prokaryotes	71 (35–152)	84 (70–92)
Lateral advection	0			
Total	92 (68–116)		85 (48–167)	—

Input fluxes and respiration rates (mg C m^{-2} d^{-1}) are based on measurements at the PAP site. Numbers in brackets refer to lower and upper estimates (see text). Community respiration was estimated by combining highest and lowest estimates.

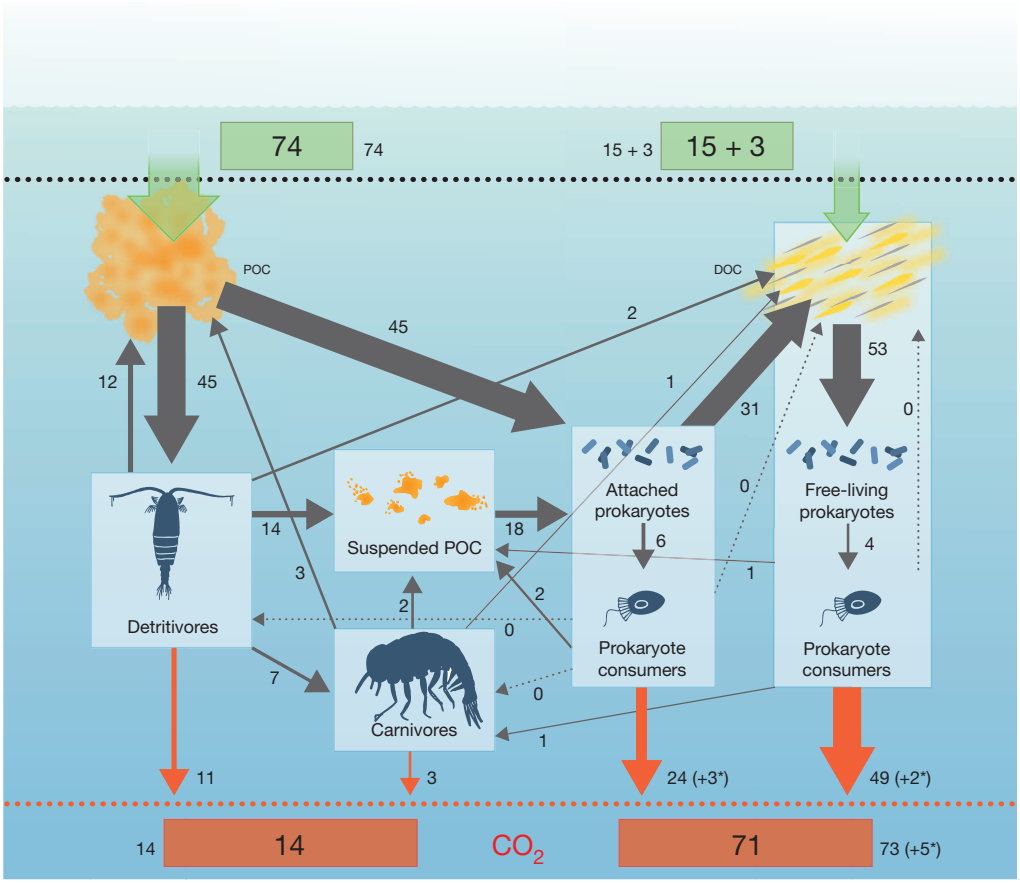


Figure 2 | Predicted carbon cycle in the twilight zone. Organic carbon is supplied to the twilight zone as POC and DOC (vertical mixing plus active transport) (green arrows). POC is processed by detritivores (50%) or attached prokaryotes (50%) and recycled in the twilight zone until eventually remineralized (red arrows), whereby prokaryotes dominate respiration (79%). Observed rates of carbon supply to the twilight zone and remineralization via respiration ('CO_2') are displayed in green and red boxes, respectively. Internal net flows (mg C m^{-2} d^{-1}) derived from a numerical model are represented as arrows (line width reflects flow magnitude; dotted lines are flows <0.5 mg C m^{-2} d^{-1}). Fluxes indicated with an asterisk are for microzooplankton (prokaryote consumers), which are not included in the measured estimates.

which had not been measured during the study, has a small role in the overall budget, contributing only 5 mg C m^{-2} d^{-1}. Attached prokaryotes processed half of the POC flux. The remaining half was processed by detritivorous zooplankton, which released 30% of it as suspended POC, thereby confirming our hypothesis. The relative roles of zooplankton and prokaryotes in processing and respiring sinking POC are robust to changing model parameter values (Methods and Extended Data Fig. 7). Moreover, it is consistent with the general perception that detritivores are sloppy feeders that ingest <40% of processed particles, causing most fast-sinking POC to break up into slow- or non-sinking POC and DOC[25]. This pool of suspended organic matter stimulates the microbial loop[28] in the twilight zone and ultimately fuels the respiration of prokaryotes[6,26].

Our results highlight a synergy between zooplankton and microbes in the twilight zone, where both are important in processing the organic carbon flux and, subsequently, in controlling the strength of the oceanic carbon sink. Large uncertainties remain, however, particularly with regard to estimating prokaryotic activity. A better understanding of prokaryotic metabolism throughout the twilight zone, combined with process studies focusing on the upper twilight zone, is necessary to understand the biological carbon pump fully.

METHODS SUMMARY

We conducted an extensive programme of field measurements at the PAP site (49.0 °N, 16.5 °W) from 8 July to 13 August 2009 aboard RRS *Discovery*. Sinking material was collected for 48 h using free-drifting, neutrally buoyant PELAGRA sediment traps[12]. Samples were screened to remove swimmers, split into aliquots, filtered onto pre-combusted glass fibre filters (grade GF/F), fumed with sulphurous acid and analysed for POC. DOC input was estimated from data collected near the PAP site during May–June 2005 (https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/f3b3d4e9-5ede-2824-e044-000b5de50f38/) and October–November 2005 (https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/f3b3d4e9-5edf-2824-e044-000b5de50f38/). The slope of the correlation between measured DOC and apparent oxygen utilization was compared with the theoretical slope ($C_{org}/-O_2 = 117/170$), giving the relative contribution of DOC to heterotrophic respiration[15]. A lower estimate was calculated using turbulent diffusivity measurements at the PAP site[16], coupled with the aforementioned DOC profiles. Samples for zooplankton biomass profiles (0–1,000 m at 80-m intervals) were preserved in formaldehyde, size-fractionated, identified and enumerated. One to fifty individuals from each group at each depth and size fraction were analysed for dry weight. Zooplankton respiration (in micrograms of carbon per individual per hour) was estimated as a function of body mass (in milligrams dry weight per individual) and temperature[6] (degrees Celsius). DOC excretion at depth was assumed to be equivalent to 31% of respiration by migrating zooplankton[18]. Leucine incorporation rates were estimated on samples ($n = 37$) recovered from depth using a conductivity–temperature–depth rosette sampler. Both time-course experiments and concentration-series bioassays were carried out. In these respective types of experiment, ^{3}H-leucine was added at final concentrations of 10–20 nM and 0.025–0.5 nM and incubated in the dark at *in situ* temperatures for 4–8 h and 0.5–2 h. Samples were filtered onto 0.2-μm polycarbonate filters and washed with deionized water, and their radioactivity was then measured.

The cruise metadata report is available from the British Oceanographic Data Centre (http://www.bodc.ac.uk/data/documents/cruise/9451/).

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 February 2013; accepted 31 January 2014.
Published online 19 March 2014.

1. Laws, E. A., Falkowski, P. G., Smith, W. O. J., Ducklow, H. & McCarthy, J. J. Temperature effects on export production in the open ocean. *Glob. Biogeochem. Cycles* **14,** 1231–1246 (2000).
2. Henson, S. A. *et al.* A reduced estimate of the strength of the ocean's biological carbon pump. *Geophys. Res. Lett.* **38,** L04606 (2011).
3. Kwon, E. Y., Primeau, F. & Sarmiento, J. L. The impact of remineralization depth on the air–sea carbon balance. *Nature Geosci.* **2,** 630–635 (2009).
4. Boyd, P. W. *et al.* Transformations of biogenic particulates from the pelagic to the deep ocean realm. *Deep-Sea Res. II* **46,** 2761–2792 (1999).
5. Reinthaler, T., Van Aken, H. M. & Veth, C. Prokaryotic respiration and production in the meso- and bathypelagic realm of the eastern and western North Atlantic basin. *Limnol. Oceanogr.* **51,** 1262–1273 (2006).
6. Steinberg, D. K. *et al.* Bacterial vs zooplankton control of sinking particle flux in the ocean's twilight zone. *Limnol. Oceanogr.* **53,** 1327–1338 (2008).
7. Baltar, F., Aristegui, J., Gasol, J. M., Sintes, E. & Herndl, G. J. Evidence of prokaryotic metabolism on suspended particulate organic matter in the dark waters of the subtropical North Atlantic. *Limnol. Oceanogr.* **54,** 182–193 (2009).
8. Burd, A. B. *et al.* Assessing the apparent imbalance between geochemical and biochemical indicators of meso- and bathypelagic biological activity: what the @$é! is wrong with present calculations of carbon budgets? *Deep-Sea Res. II* **57,** 1557–1571 (2010).
9. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. VERTEX: carbon cycling in the northeast Pacific. *Deep-Sea Res. I* **34,** 267–285 (1987).
10. Henson, S. A., Dunne, J. P. & Sarmiento, J. L. Decadal variability in North Atlantic phytoplankton blooms. *J. Geophys. Res.* **114,** C04013 (2009).
11. Buesseler, K. O. *et al.* Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. *Limnol. Oceanogr.* **54,** 1210–1232 (2009).
12. Lampitt, R. S. *et al.* Particle export from the euphotic zone: estimates using a novel drifting sediment trap, ^{234}Th and new production. *Deep-Sea Res. I* **55,** 1484–1502 (2008).
13. Riley, J. S. *et al.* The relative contribution of fast and slow sinking particles to ocean carbon export. *Glob. Biogeochem. Cycles* **26,** GB1026 (2012).
14. Buesseler, K. O. *et al.* Revisiting carbon flux through the ocean's twilight zone. *Science* **316,** 567–570 (2007).
15. Doval, M. D. & Hansell, D. A. Organic carbon and apparent oxygen utilization in the western South Pacific and the central Indian oceans. *Mar. Chem.* **68,** 249–264 (2000).
16. Martin, A. P. *et al.* The supply of nutrients due to vertical turbulent mixing: a study at the Porcupine Abyssal Plain study site in the northeast Atlantic. *Deep-Sea Res. II* **57,** 1293–1302 (2010).
17. Carlson, C. A. *et al.* Dissolved organic carbon export and subsequent remineralization in the mesopelagic and bathypelagic realms of the North Atlantic basin. *Deep-Sea Res. II* **57,** 1433–1445 (2010).
18. Steinberg, D. K. *et al.* Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep-Sea Res. I* **47,** 137–158 (2000).
19. Reinthaler, T., Van Aken, H. M. & Herndl, G. J. Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep-Sea Res. II* **57,** 1572–1580 (2010).
20. Ikeda, T., Sano, F., Yamaguchi, A. & Matsuishi, T. Metabolism of mesopelagic and bathypelagic copepods in the western North Pacific Ocean. *Mar. Ecol. Prog. Ser.* **322,** 199–211 (2006).
21. Kirchman, D., K'nees, E. & Hodson, R. Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. *Appl. Environ. Microbiol.* **49,** 599–607 (1985).
22. Ducklow, H. W., Kirchman, D. L., Quinby, H. L., Carlson, C. A. & Dam, H. G. Stocks and dynamics of bacterioplankton carbon during the spring bloom in the eastern North Atlantic Ocean. *Deep-Sea Res. II* **40,** 245–263 (1993).
23. DeLong, E. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311,** 496–503 (2006).
24. Iversen, M. H., Nowald, N., Ploug, H., Jackson, G. A. & Fisher, G. High resolution profiles of vertical particulate organic matter export off Cape Blanc, Mauritania: degradation processes and ballasting effects. *Deep-Sea Res. I* **57,** 771–784 (2010).
25. Lampitt, R., Noji, T. & Von Bodungen, B. What happens to zooplankton faecal pellets? Implications for material flux. *Mar. Biol.* **104,** 15–23 (1990).
26. Anderson, T. R. & Tang, K. W. Carbon cycling and POC turnover in the mesopelagic zone of the ocean: insights from a simple model. *Deep-Sea Res. II* **57,** 1581–1592 (2010).
27. Jumars, P., Penry, D. & Baross, J. Closing the microbial loop: dissolved carbon pathway to heterotrophic bacteria from incomplete ingestion, digestion and absorption in animals. *Deep-Sea Res. A* **36,** 483–495 (1989).
28. Azam, F. *et al.* The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10,** 257–263 (1983).

Acknowledgements We thank the captain and crew of the RRS *Discovery* and scientists during D341, especially J. Hunter for executing the ARIES deployments and S. Ward for the PELAGRA deployments. We thank the OSCAR Project Office and BODC for providing data. Finally, we thank T. Cornulier for statistical help. This work was funded by Oceans 2025 and EU FP7-ENV-2010 Collaborative Project 264933 BASIN Basin-Scale Analysis, Synthesis and Integration. C.T. and M.B. were funded by the ANR-POTES program (no. ANR-05-BLAN-0161-01, awarded to C. T.) supported by the Agence Nationale de la Recherche (ANR, France). D.J.M. was funded by NERC (NE/G014744/1).

Author Contributions R.S. and S.L.C.G. designed and conducted the study; R.S.L. was involved in the conceptual development; S.L.C.G., C.T., M.B., M.V.Z., C.M.M. and S.A.H. all contributed data; K.C. was involved in sample analyses; and K.S. coordinated the PELAGRA deployments. T.R.A. and D.J.M. developed the model and implemented it for field data interpretation; S.L.C.G. analysed the data and wrote the manuscript together with R.S., D.J.M. and T.R.A. All authors discussed 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 T.R.A. (tra@noc.ac.uk, for model R code) or S.L.C.G. (s.giering@abdn.ac.uk, for all other aspects of this work).

LETTER

doi:10.1038/nature13122

Geomagnetic fluctuations reveal stable stratification at the top of the Earth's core

Bruce Buffett[1]

Modern observations of the geomagnetic field reveal fluctuations[1–3] with a dominant period of about 60 years. These fluctuations are probably a result of waves in the liquid core[4], although the precise nature of the waves is uncertain. Common suggestions include a type of magnetic wave, known as a torsional oscillation[5], but recent studies[6] favour periods that are too short to account for a 60-year fluctuation. Another possibility involves MAC waves[7], which arise from the interplay between magnetic, Archimedes and Coriolis forces. Waves with a suitable period can emerge when the top of the core is stably stratified. Here I show that MAC waves provide a good description of time-dependent zonal flow at the top of the core[8], as inferred from geomagnetic secular variation[9]. The same wave motion can also account for unexplained fluctuations in the dipole field[10]. Both of these independent predictions require a 140-kilometre-thick stratified layer with a buoyancy frequency comparable to the Earth's rotation rate. Such a stratified layer could have a thermal origin[11,12], implying a core heat flow of about 13 terawatts. Alternatively, the layer could result from chemical stratification[13,14]. In either case, the existence of a stratified layer at the top of the core obscures the nature of flow deeper in the core, where the magnetic field is continually regenerated.

Evidence for a 60-year fluctuation in the geomagnetic field was first reported[1] more than 40 years ago. Subsequent studies[2,3] using longer records have confirmed the existence of a dominant 60-year period, and find evidence for a weaker 30-year fluctuation. Initial explanations appealed to torsional oscillations[5], which are nearly geostrophic motions in the form of fluid cylinders aligned with the rotation axis. The restoring force for wave motion is due to the component of the magnetic field that threads across the surface of the cylinders. A nominal 60-year period implies a relatively weak magnetic field of roughly 0.3 mT (refs 15 and 16). Stronger magnetic fields in numerical geodynamo models[17] imply shorter periods and recent observations[6] support the existence of torsional oscillations with periods of 4 to 6 years.

Longer-period waves are possible when the top of the fluid core is stably stratified[7]. The required ingredients for wave motion include fluid buoyancy, rapid rotation and a radial magnetic field that passes through the stratified layer. Figure 1 shows a schematic illustration of the associated fluid velocity. A radial component of flow V_r alters the distribution of fluid density in the stratified layer and induces a perturbation in pressure. Owing to the effects of rapid rotation, the pressure perturbation causes an azimuthal flow V_ϕ much like cyclonic flow around a pressure low in the atmosphere. Azimuthal flow is opposed by the presence of the radial magnetic field, which drives a weaker meridional flow V_θ towards the pole. The restoring force for the waves comes partly from the buoyancy force, which opposes the radial motion, and partly from magnetic tension, which opposes the horizontal flow. The direction of flow reverses after half a cycle; an eastward azimuthal flow drives a meridional flow towards the Equator.

Braginsky[7] obtained analytical solutions for these waves when the radial magnetic field is approximated by an axial dipole. Different modes of oscillations are distinguished by the spatial pattern of radial flow. The gravest modes are associated with upwelling at the poles and downwelling at the Equator over half a cycle. For each pattern of radial flow there are overtones associated with shorter wavelengths in the radial direction. In general the overtones are more heavily damped than the fundamental modes. In addition, waves with a larger number of upwellings and downwellings have shorter periods.

A description of waves with a more general form of radial magnetic field requires a numerical solution. Indirect inferences from geodetic observations[18,19], as well as numerical geodynamo models[17], suggest that a large part of the radial magnetic field near the core–mantle boundary resides in wavelengths that are too small to detect at the Earth's surface. A simple, but plausible, description of the radial magnetic field at the core–mantle boundary assumes a random distribution. Because magnetic tension does not depend on the direction of the field, I can approximate the root-mean-square (r.m.s.) value of the radial magnetic field as a constant over the surface of the core. Changes in the radial magnetic field across the stratified layer are small when the layer is sufficiently thin.

Two parameters specify the waves when the r.m.s. value of the radial magnetic field is fixed by geodetic constraints at 0.6 mT (see Methods). One parameter is the thickness H of the stratified layer and the other is the buoyancy frequency N defined as

$$N = \sqrt{-\frac{g}{\rho}\frac{\partial \rho}{\partial r}}$$

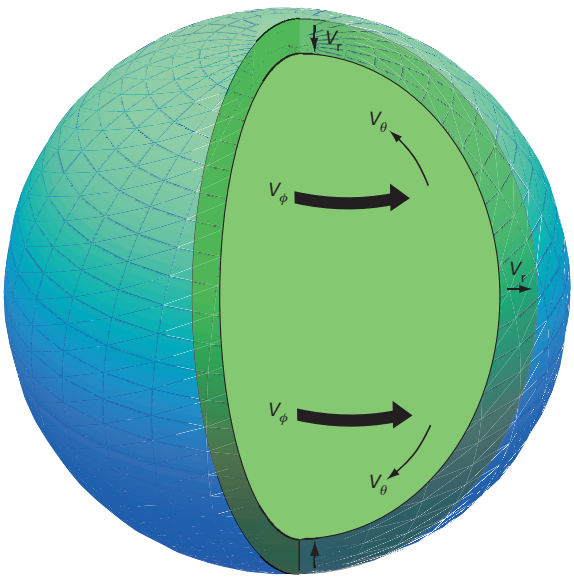


Figure 1 | Schematic illustration of the wave motion. Radial motion V_r causes a pressure perturbation, which drives an azimuthal flow V_ϕ in the stratified layer. The presence of a radial magnetic field opposes V_ϕ and induces a meridional flow V_θ. The fluid velocities reverse direction over a full cycle of the wave.

[1]Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA.

where $\rho \cong 10^4\ \mathrm{kg\,m^{-3}}$ is the fluid density, $\partial\rho/\partial r$ is the radial derivative and $g \cong 10\ \mathrm{m\,s^{-2}}$ is the acceleration due to gravity. Variations in density due to changes in hydrostatic pressure do not contribute to the buoyancy force, so the radial derivative in density is defined relative to variations in a well-mixed, isentropic fluid. For simplicity H is assumed to be constant and N varies linearly with radius across the layer. The value of N is largest at the core–mantle boundary and decreases to zero at the base of the stratified layer. As a result, the waves are completely defined by H and the maximum value of N (denoted as $N_{\max}$).

Figure 2 shows a representative wave with a symmetric pattern of upwelling and downwelling in the Northern and Southern hemispheres. This particular wave is computed using $H = 140$ km and ${N_{\max}}^2 = 1.02\Omega^2$, where $\Omega = 0.73 \times 10^{-4}\ \mathrm{s^{-1}}$ is the rotation rate of the planet. The corresponding period of the wave is 58 years, which is within the uncertainty of estimates for geomagnetic fluctuations.

The absolute amplitude of the wave in Fig. 2 is not uniquely set by the computations, but the relative size of the velocity components reveals the structure of the wave. The largest flow is associated with V_ϕ, whereas V_θ is about an order of magnitude smaller. Estimates of V_ϕ and V_θ have been inferred at the top of the core from variations in the geomagnetic field[8,9], so in principle we can look for evidence of wave motion in the surface core flow. The zonal part of V_ϕ (that is, independent of longitude) has a value of the order of several kilometres per year, and exhibits large variations in both space and time (see Fig. 3). On the other hand, the zonal part of V_θ is often prohibited by dynamical constraints in the inversion for flow[8,20]. Consequently, I begin my search for waves using only the zonal part of V_ϕ.

Estimates of V_ϕ over the time interval between 1930 and 1990 are well represented by a linear combination of waves. The chosen time interval is long enough to capture the dominant 60-year period, but restricts the record to times when the surface core flow is most reliable. Values for H and $N_{\max}$ are iteratively adjusted to minimize the misfit between the core-flow estimates and a linear combination of six MAC waves (see Extended Data Table 1). For each pair of model parameters (H, $N_{\max}$), the amplitude and phase of the waves are estimated by the method of least squares. The optimal fit to the core flow is shown in Fig. 3 using $H = 140$ km and ${N_{\max}}^2 = 1.02\Omega^2$. Good agreement means that the surface core flow is compatible with waves, but I cannot conclude that the wave motion uniquely accounts for the core flow.

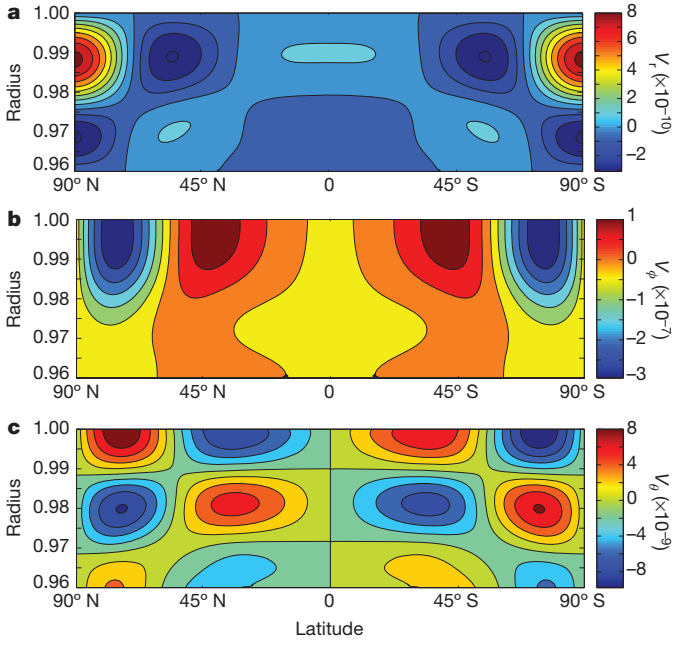


Figure 2 | A representative wave inside the stratified layer. The velocity components V_r (**a**), V_ϕ (**b**) and V_θ (**c**) are shown as a function of radius r and latitude, where $r = 1$ denotes the core–mantle boundary. This particular wave has a period of 58 years for the best-fitting values of layer thickness H and buoyancy frequency $N_{\max}$.

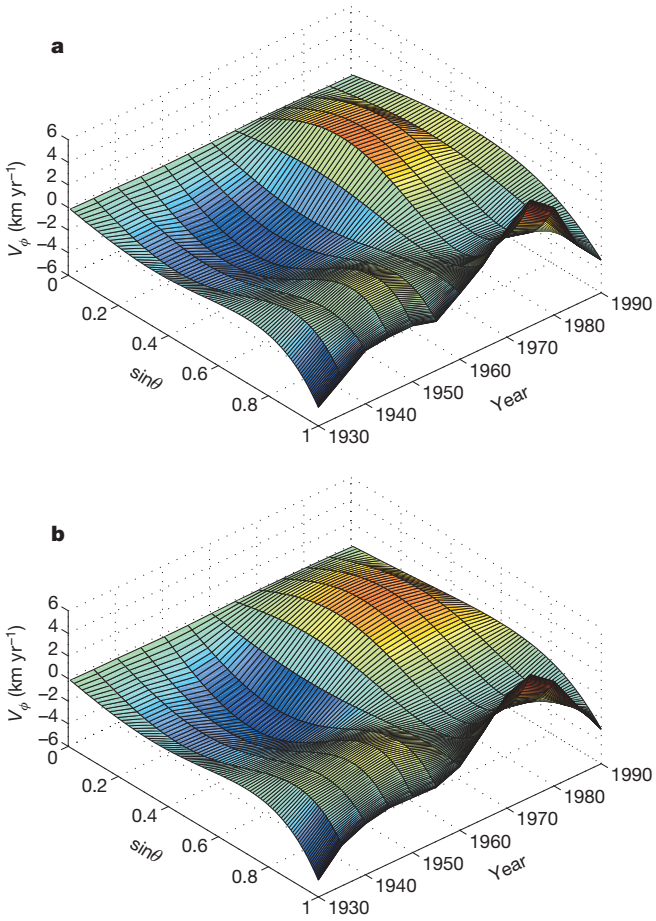


Figure 3 | Estimate of surface core flow and prediction due to MAC waves. **a**, Zonal flow V_ϕ from the uvm-s model of Jackson[8] is shown as a function of sinθ and time. Only the long-wavelength part of the flow in spherical harmonic degrees 1, 3 and 5 is retained for fitting the MAC waves. The components of the spherical harmonic expansion that are least affected are the small-scale flow and magnetic field at the core–mantle boundary[28]. **b**, A linear combination of six MAC waves can reliably reproduce the surface core flow. The model parameters for the waves (H, $N_{\max}$) are iteratively adjusted to find the optimal fit to the surface core flow.

An independent indication of wave motion is found in fluctuations of the dipole field[10]. Once a linear combination of waves is chosen to fit V_ϕ, the resulting V_θ component of flow alters the strength of the dipole field by sweeping the radial magnetic field into the polar regions. To quantify the changes in the dipole, I assume that the effects of magnetic diffusion are small for decadal fluctuations[20]. In this case time variations in the radial magnetic field B_r are described by

$$\dot{B}_r = -\nabla_H \cdot (\boldsymbol{V} B_r)$$

where ∇_H denotes the horizontal part of the divergence and $\boldsymbol{V}$ is the velocity due to waves. Only zonal components of B_r at the longest wavelengths contribute significantly to the dipole fluctuation because the waves are zonal and predominantly large-scale. When B_r is evaluated at the mid-range epoch of 1960 using the magnetic field model gufm1 (ref. 21), I obtain predictions for the fluctuation in the dipole field. Restoring the effects of magnetic diffusion causes only a small change in the amplitude and phase of the dipole fluctuation (see Methods). The resulting predictions are compared with time variations in the gufm1 model[21], after removing a linear trend in the strength of the dipole (Fig. 4).

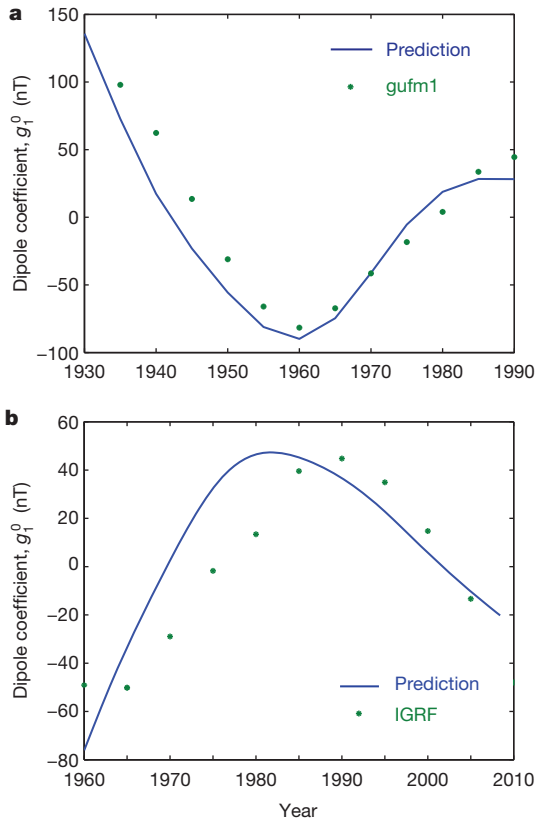


Figure 4 | Fluctuations in the dipole field due to MAC waves. Predictions for the interval 1930–1990 (**a**) and 1960–2008 (**b**) are based on the flow models of Jackson[8] and Wardinski and Lesur[22], respectively. The observed fluctuations are represented in terms of the Gauss coefficient $g_1^0(t)$ from the gufm1 model[21] and the IGRF model[29], after removing linear trends.

The amplitude and frequency of the dipole fluctuation agrees remarkably well with the prediction based on MAC waves. Even the decay of the dipole fluctuation is broadly consistent with the damping associated with the waves, although an excitation source must be present to sustain the waves over longer time intervals.

Evidence for MAC waves can also be found in more recent models of surface core flow. One example is the extended C^{3}FM model[22], which gives estimates for the flow between 1957 and 2008 from a combination of satellite and observatory observations. The set of waves that best fit the model of Jackson[8] also gives a good fit to the C^{3}FM model, although it is not possible to capture fluctuations with periods shorter than the shortest-period wave (see Methods). Slightly better fits to the C^{3}FM model are possible with a modest increase in stratification to $N_{max}{}^2 = 1.08\Omega^2$, but I retain the original set of waves to predict dipole fluctuations in more recent times. These predictions match the observations reasonably well with no adjustable parameters (see Fig. 4b).

Satisfactory predictions for the dipole fluctuations in Fig. 4 require an electrical conductivity of 10^6 S m^{-1}, which is consistent with recent theoretical predictions[11,12]. Lower values of electrical conductivity cause excessive damping of the waves. Adopting larger values for H can offset the excess damping, but thicker layers underestimate the amplitude of the dipole fluctuation. In addition, the model appears to require a fairly uniform r.m.s. radial magnetic field over the surface of the core. Calculations with a dipolar radial field alter the structure of the waves enough to prevent a good fit to both V_ϕ and the dipole fluctuation.

Stratification at the top of the core can arise in several ways. Thermal stratification is expected when the core heat flow is less than a hypothetical heat flow conducted through a well-mixed region at the top of the core[23,24] (sometimes called the adiabatic heat flow). A simple model for thermal stratification[25] suggests that a core heat flow of 13 TW would produce a stratified layer 140 km thick when the adiabatic heat flow is 15 TW (ref. 11). These two heat flows correspond to temperature gradients that differ by 0.12 K km^{-1}, yielding a density gradient of 0.012 kg km^{-1} at the core–mantle boundary. The resulting buoyancy frequency is 1.5Ω, which is close to the estimate recovered by fitting V_ϕ to MAC waves. Alternatively, the stratification might arise from radial variations in composition. Convection in the underlying region would probably provide a continuous source of excitation for the waves because buoyant parcels strike the base of the stratified layer. The resulting wave motion would obscure deeper motions in the core at periods of roughly a century and less.

METHODS SUMMARY

Numerical solutions for the waves are adapted from a numerical geodynamo model[26]. Perturbations in the velocity and magnetic fields are expanded in vector spherical harmonics using a truncation at degree $l = 40$. Both of these fields are defined on a radial grid with 200 equally spaced levels. Radial derivatives in the governing equations are represented by second-order finite differences. When the time dependence is periodic, the governing equations and boundary conditions (see Methods) yield an eigenvalue problem for the frequency and structure of the waves. Iterative solutions are obtained using an incomplete Arnoldi method[27]. The resulting waves are used to compute the associated fluctuations in the dipole field. The induction equation for the perturbation in the radial magnetic field at the core–mantle boundary is discretized in latitude and solved using the finite difference method. The dipole component is extracted using the orthogonality of spherical harmonics.

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

Received 1 November 2013; accepted 30 January; corrected online 26 March 2014 (see full-text HTML version for details).

1. Currie, R. G. Geomagnetic line spectra—2 to 70 years. *Astrophys. Space Sci.* **21,** 425–438 (1973).
2. Roberts, P. H., Yu, Z. J. & Russell, C. T. On the 60-year signal from the core. *Geophys. Astrophys. Fluid Dyn.* **101,** 11–35 (2007).
3. Jackson, L. P. & Mound, J. E. Geomagnetic variation on decadal time scales: what can we learn from empirical mode decomposition? *Geophys. Res. Lett.* **37,** L14307 (2010).
4. Finlay, C. C., Dumberry, M., Chulliat, A. & Pais, M. A. Short timescale core dynamics: theory and observation. *Space Sci. Rev.* **155,** 177–218 (2010).
5. Braginsky, S. I. Torsional magnetohydrodynamic vibrations in the Earth's core and variations in day length. *Geomagn. Aeron.* **10,** 1–10 (1970).
6. Gillet, N., Jault, D., Canet, E. & Fournier, A. Fast torsional waves and strong magnetic field with the Earth's core. *Nature* **465,** 74–77 (2010).
7. Braginsky, S. I. MAC-oscillations of the hidden ocean of the core. *J. Geomag. Geoelectr.* **45,** 1517–1538 (1993).
8. Jackson, A. Time dependency of tangentially geostrophic core surface motion. *Phys. Earth Planet. Inter.* **103,** 293–311 (1997).
9. Bloxham, J. & Jackson, A. Time-dependent mapping of the magnetic field at the core-mantle boundary. *J. Geophys. Res.* **97,** 19537–19563 (1992).
10. Yokoyama, Y. & Yukutake, T. Sixty-year variation in a time series of the geomagnetic gauss coefficients between 1910 and 1983. *J. Geomag. Geoelectr.* **43,** 563–584 (1991).
11. Pozzo, M. Davies, C. Gubbins, D. & Alfe, D. Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* **485,** 355–358 (2012).
12. Pozzo, M., Davies, C., Gubbins, D. & Alfe, D. Transport properties for liquid silicon-oxygen-iron mixtures at Earth's core conditions. *Phys. Rev. B* **87,** 014110 (2013).
13. Buffett, B. A. & Seagle, C. T. Stratification at the top of the core due to chemical interaction with the mantle. *J. Geophys. Res.* **115,** B04407 (2010).
14. Gubbins, D. & Davies, C. J. The stratified layer at the core-mantle boundary caused by barodiffusion of oxygen, sulphur and silicon. *Phys. Earth Planet. Inter.* **215,** 21–28 (2013).
15. Zatman, S. & Bloxham, J. Torsional oscillations and the magnetic field within the Earth's core. *Nature* **388,** 760–763 (1997).
16. Buffett, B. A., Mound, J. E. & Jackson, A. Inversion of torsional oscillations for the structure and dynamics of Earth's core. *Geophys. J. Int.* **177,** 878–890 (2009).
17. Christensen, U. R. Geodynamo models: tools for understanding properties of Earth's magnetic field. *Phys. Earth Planet. Inter.* **187,** 157–169 (2011).
18. Buffett, B. A., Matthews, P. M. & Herring, T. A. Modeling of nutation and precession: effects of electromagnetic coupling. *J. Geophys. Res.* **107,** B42070 (2002).
19. Koot, L. *et al.* Constraints on the coupling at the core-mantle and inner-core boundaries inferred from nutation observations. *Geophys. J. Int.* **182,** 1279–1294 (2010).
20. Holme, R. & Olsen, N. Core surface flow modeling from high-resolution secular variation. *Geophys. J. Int.* **166,** 518–528 (2006).
21. Jackson, A., Jonkers, A. R. T. & Walker, M. R. Four centuries of geomagnetic secular variation from historical records. *Phil. Trans. R. Soc. Lond. A* **358,** 957–990 (2000).

22. Wardinski, I. & Lesur, V. An extended version of the C^3FM geomagnetic field model: application of a continuous frozen-flux constraint. *Geophys. J. Int.* **189,** 1409–1429 (2012).
23. Gubbins, D., Thomson, C. J. & Whaler, K. A. Stable regions in the Earth's liquid core. *Geophys. J. R. Astron. Soc.* **68,** 241–251 (1982).
24. Labrosse, S., Poirier, J. P. & LeMouel, J. L. On the cooling of the Earth's core. *Phys. Earth Planet. Inter.* **99,** 1–17 (1997).
25. Lister, J. R. & Buffett, B. A. Stratification of the outer core at the core-mantle boundary. *Phys. Earth Planet. Inter.* **105,** 5–19 (1998).
26. Kuang, W. & Bloxham, J. Numerical modeling of magnetohydrodynamic conv6ction in a rapidly rotating spherical shell: Weak and strong field dynamo action. *J. Comput. Phys.* **153,** 51–81 (1999).
27. Sorensen, D. C. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Matrix Anal. Appl.* **13,** 357–385 (1992).
28. Eymin, C. & Hulot, G. On core surface flows inferred from satellite magnetic data. *Phys. Earth Planet. Inter.* **152,** 200–220 (2005).
29. Finlay, C. *et al.* International geomagnetic reference field: the eleventh generation. *Geophys. J. Int.* **183,** 1216–1230 (2010).

Acknowledgements A. Jackson and I. Wardinski provided models of surface core flow and magnetic field. Comments and suggestions from R. Holme substantially improved the final text. This work was supported in part by the US National Science Foundation (EAR-1045277).

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

LETTER

doi:10.1038/nature13164

Methane fluxes show consistent temperature dependence across microbial to ecosystem scales

Gabriel Yvon-Durocher[1], Andrew P. Allen[2], David Bastviken[3], Ralf Conrad[4], Cristian Gudasz[5,6]†, Annick St-Pierre[7], Nguyen Thanh-Duc[8] & Paul A. del Giorgio[7]

Methane (CH_4) is an important greenhouse gas because it has 25 times the global warming potential of carbon dioxide (CO_2) by mass over a century[1]. Recent calculations suggest that atmospheric CH_4 emissions have been responsible for approximately 20% of Earth's warming since pre-industrial times[2]. Understanding how CH_4 emissions from ecosystems will respond to expected increases in global temperature is therefore fundamental to predicting whether the carbon cycle will mitigate or accelerate climate change. Methanogenesis is the terminal step in the remineralization of organic matter and is carried out by strictly anaerobic Archaea[3]. Like most other forms of metabolism, methanogenesis is temperature-dependent[4,5]. However, it is not yet known how this physiological response combines with other biotic processes (for example, methanotrophy[6], substrate supply[3,7], microbial community composition[8]) and abiotic processes (for example, water-table depth[9,10]) to determine the temperature dependence of ecosystem-level CH_4 emissions. It is also not known whether CH_4 emissions at the ecosystem level have a fundamentally different temperature dependence than other key fluxes in the carbon cycle, such as photosynthesis and respiration. Here we use meta-analyses to show that seasonal variations in CH_4 emissions from a wide range of ecosystems exhibit an average temperature dependence similar to that of CH_4 production derived from pure cultures of methanogens and anaerobic microbial communities. This average temperature dependence (0.96 electron volts (eV)), which corresponds to a 57-fold increase between 0 and 30 °C, is considerably higher than previously observed for respiration (approximately 0.65 eV)[11] and photosynthesis (approximately 0.3 eV)[12]. As a result, we show that both the emission of CH_4 and the ratio of CH_4 to CO_2 emissions increase markedly with seasonal increases in temperature. Our findings suggest that global warming may have a large impact on the relative contributions of CO_2 and CH_4 to total greenhouse gas emissions from aquatic ecosystems, terrestrial wetlands and rice paddies.

Biogenic CH_4 fluxes are a major component of global CH_4 emissions, yet they are poorly constrained[2,13,14]. There are large uncertainties not only in the current magnitude of these fluxes but also in the factors that regulate them[2,13]. In particular, there is substantial uncertainty in the parameterization of the temperature dependence of natural CH_4 emissions in process-based biogeochemistry models[15–18], which greatly hinders our ability to predict the response of this key component of the carbon cycle to global warming. For example, temperature dependencies for ecosystem-level CH_4 emissions have reported apparent activation energies that vary from 0.2 to 2.5 eV (refs 6, 19–21) (1 eV = 96 kJ mol^{-1}).

In a bid to reduce this uncertainty, which is fundamental to improving projections of feedbacks between the carbon cycle and future climate change[15–18], we quantified variation in the temperature dependence of CH_4 fluxes in meta-analyses of three different types of experiments (pure cultures of methanogens, laboratory incubations of anaerobic sediments, and seasonal field surveys of CH_4 emissions) that correspond to three distinct levels of biological organization (population, community and ecosystem). In particular, we assess whether ecosystem-level CH_4 emissions exhibit temperature dependencies similar to those of the underlying methanogenic process. To do this, we first established the magnitude and variability of the temperature dependencies of key rate processes for populations of methanogens in culture (methanogenesis, growth) and laboratory incubations of anaerobic microbial communities from natural sediment samples (CH_4 production). We then assessed whether these temperature dependencies differ from those observed in an ecosystem-level analysis of the seasonal temperature dependence of CH_4 emissions from aquatic, wetland and rice-paddy ecosystems.

To characterize the temperature dependencies of physiological rate processes for methanogens, we fit the Boltzmann–Arrhenius function (which describes the exponential relationship between metabolic rate and temperature, assuming a single enzyme-catalysed reaction is rate-limiting[22]), separately, to the data compiled from the population and community-level experiments using linear mixed-effects models (see Methods).

The population-level analysis reveals that the average temperature dependencies for the rates of methanogenesis and growth are similar. Specifically, the improvement in model fit going from a null model, which assumes a common average activation ($\overline{E_M}$ in equation (1), see Methods) energy for both rate processes to an alternative model, which assumes a distinct average activation energy for each rate process, is not statistically significant (likelihood ratio test: $\chi^2 = 0.39$, d.f. = 1, $P = 0.53$). Thus, the average temperature dependencies for both rate processes (methanogenesis and growth) can be characterized using the same average apparent activation energy ($\overline{E_M} = 1.10$ eV, 95% confidence interval of 0.93–1.27 eV; Fig. 1a).

The community-level analysis of CH_4 production rates from anaerobic sediment incubations produces a similar value for the average activation energy ($\overline{E_M} = 0.93$ eV, 95% confidence interval of 0.82 – 1.03 eV), suggesting that the temperature dependence of CH_4 production at the community level largely reflects the kinetics of the physiological processes generating this flux. More detailed analyses of these community-level data indicate that the average activation energies for the sediments from three broadly defined ecosystem types (that is, aquatic ecosystems, wetlands and rice paddies) are statistically indistinguishable (likelihood ratio test comparing a null model with a single activation energy to an alternative model with separate activation energies for each ecosystem type: $\chi^2 = 1.62$, d.f. = 2, $P = 0.44$). However, for both the population- and community-level analyses, temperature dependencies do vary among experimental units, *i*, as indicated by the statistical significance of random effects on the activation energies (characterized by $\varepsilon_{E_M,i}$ in equation (1), see Methods and Table 1), owing, for example, to inherent physiological

[1]Environment and Sustainability Institute, University of Exeter, Penryn, Cornwall, TR10 9EZ. UK. [2]Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia. [3]Department of Thematic Studies – Water and Environmental Studies, Linköping University, SE-581 83 Linköping, Sweden. [4]Max-Planck-Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany. [5]Department of Ecology and Environmental Sciences, Umeå University, Linnaeus väg 6, SE-901 87 Umeå, Sweden. [6]Department of Ecology and Genetics, Limnology, Uppsala University, Norbyvägen 18D, SE-752 36, Uppsala Sweden. [7]Département des sciences biologiques, Université du Québec à Montréal, Montréal, Province of Québec, H2X 3X8, Canada. [8]Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire 03824, USA. †Present address: Department of Ecology and Evolutionary Biology, Princeton University, Princeton, 106A Guyot Hall, New Jersey 08544, USA.

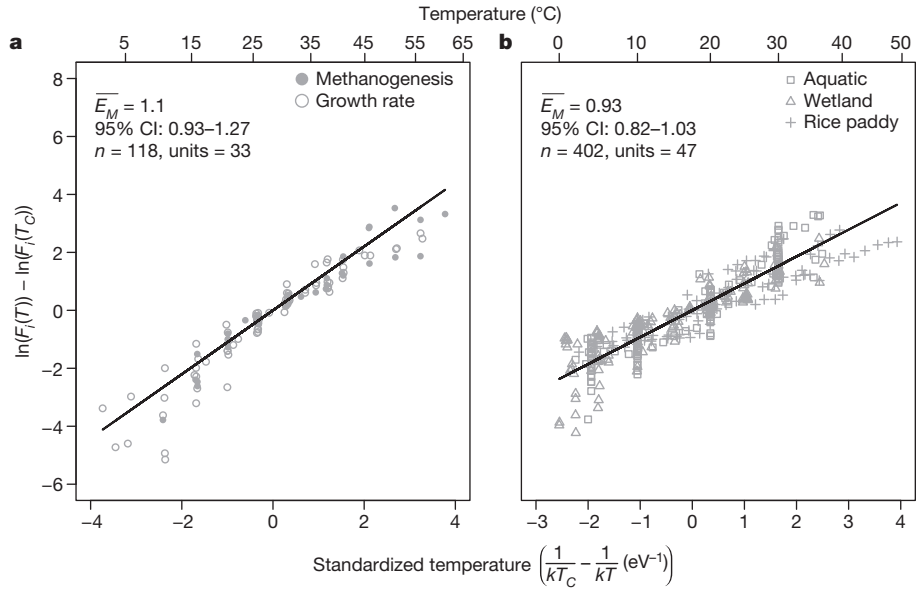


Figure 1 | Temperature dependence of CH_4 production and related processes at population and community levels. Temperature dependencies for methanogen populations in culture (**a**) and anaerobic microbial communities from natural sediment samples (**b**) are separately characterized using mixed-effects models by fitting Boltzmann-Arrhenius functions with experimental-unit-level random effects on the apparent activation energy and rate at fixed temperature (see equation (1) in the Methods). Analyses reveal that average temperature dependencies are statistically indistinguishable for rates of growth and methanogenesis (**a**, open circles and filled circles, respectively) by populations of methanogens ($P = 0.53$), and rates of CH_4 production by laboratory-incubated anaerobic microbial communities from aquatic, wetland and rice-paddy sediments (**b**, squares, triangles and crosses, respectively) ($P = 0.44$). The fitted solid lines correspond to the average apparent activation energies estimated from the mixed-effects models ($\overline{E_M} = 1.10$ and 0.93 eV for the population (**a**) and community (**b**) analyses, respectively). Data have been standardized by subtracting from each measurement the estimated experiment-specific intercept; that is, the estimated rate at fixed temperature, $\ln F(T_C)$, where T_C is the average of the temperature measurements for the population-level data (27.2 °C in **a**) and the community-level data (17.5 °C in **b**). These standardizations are used for visualization of the data only.

differences between strains of methanogens in culture and variation between sediment samples in microbial community composition.

To assess how the average temperature dependence of CH_4 production for methanogenic populations and communities compares to that of CH_4 emissions from entire ecosystems, we fit the Boltzmann–Arrhenius equation to a database of 1,553 measurements of CH_4 emission and temperature, measured seasonally for 127 field sites that span the globe and encompass wetlands, rice paddies and aquatic ecosystems. Analysis revealed that the estimated average activation energies are statistically indistinguishable for these three ecosystem types (likelihood ratio test comparing a null model with a single activation energy to an alternative model with separate activation energies for each ecosystem type: $\chi^2 = 4.97$, d.f. $= 2$, $P = 0.10$), and can therefore be characterized by a common parameter ($\overline{E_M} = 0.96$ eV; 95% confidence interval of 0.86–1.07; Fig. 2). This average temperature dependence is statistically indistinguishable to those observed for CH_4 production from populations in culture and incubations of microbial communities from sediments—that is, all three estimates for $\overline{E_M}$ have 95% confidence intervals that overlap (Table 1)—which is remarkable given the multitude of processes that may confound the temperature dependence of CH_4 emissions over a seasonal cycle at the ecosystem level. Thus, our analysis is consistent with the hypothesis that the seasonal temperature dependence of CH_4 emissions at the ecosystem level largely reflects the kinetics of the methanogenic processes generating this flux (see Supplementary Information section 2 for further discussion of the potential mechanisms governing the scaling of the temperature dependence of CH_4 fluxes from populations to ecosystems).

Importantly, the average apparent activation energy we report here for the seasonal temperature dependence of ecosystem-level CH_4 emissions ($\overline{E_M} = 0.96$ eV) is considerably higher than that reported previously for CO_2 fluxes attributable to respiration (~0.65 eV)[11] and photosynthesis (~0.3 eV)[12], which could have important implications for the effect of global warming on the balance of CH_4 and CO_2 emissions from ecosystems[23]. Given these differential temperature dependencies, we expected the ratio of CH_4 to CO_2 emissions to increase, on average, with seasonal increases in temperature for a collection of sites (see Supplementary Information section 1).

To test this prediction, we analysed ecosystem-level data for the subset of sites in our compilation where simultaneous measurements of CH_4 and CO_2 fluxes have been made, enabling us to calculate the efflux ratio of these greenhouse gases in response to seasonal variation in temperature. This data set comprises 177 estimates from 38 field sites. In exactly the same way as for the analyses of CH_4 emissions, we fit a linear mixed-effects model using the Boltzmann–Arrhenius function

Table 1 | Estimates of the parameters used to characterize the temperature dependencies of CH_4 production, emission and related variables

Rate or ratio	Average activation energy (eV) (95% CI)	Standard deviation of random effect on activation energy	Standard deviation of random effect on rate or ratio at fixed temperature
Population-level rate of methanogenesis or growth	1.10 (0.93–1.27)	0.42	2.28
Community-level rate of CH_4 production	0.93 (0.82–1.03)	0.32	2.45
Ecosystem-level CH_4 emission	0.96 (0.86–1.07)	0.42	2.17
Ecosystem-level CH_4:CO_2 emission ratio	0.71 (0.46–0.97)	0.59	1.03

Parameters were determined by fitting Boltzmann–Arrhenius functions to rates (equation (1)) or ratios of rates (equation (2)) using mixed-effects models. The standard deviations of the random effects used to characterize differences among sites or experimental units in temperature dependencies ($\varepsilon_{E_{M,j}}$ and $\varepsilon_{E_{M,C,j}}$ in equations (1) and (2) of the Methods, respectively) are significantly greater than zero in all analyses ($P < 0.05$), as are the standard deviations of the random effects used to characterize differences in rates (or rate ratios) among sites or experimental units at fixed temperature ($\varepsilon_{F,j}$ and $\varepsilon_{Q,j}$, respectively) ($P < 0.001$).

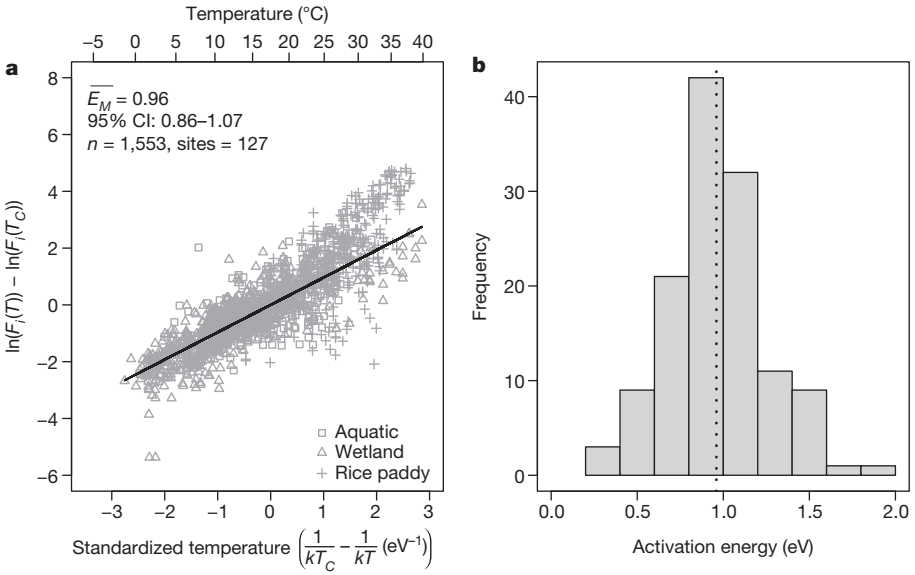


Figure 2 | Temperature dependence of CH_4 emissions at the ecosystem level. Temperature dependencies in **a** are characterized using a mixed-effects model by fitting a Boltzmann–Arrhenius function to the emission data with site-level random effects on the apparent activation energy and rate at fixed temperature (see equation (1) in the Methods). Analysis reveals that CH_4 emissions exhibit average temperature dependencies that are statistically indistinguishable for aquatic (squares), wetland (triangles) and rice-paddy (crosses) ecosystems ($P = 0.10$). The fitted solid line corresponds to the average apparent activation energy estimated from the mixed-effects model ($\overline{E_M} = 0.96$ eV). The distribution of apparent activation energies among sites (**b**) yields an average ($\overline{E_M} = 0.96$ eV, represented by a dashed line) that is statistically indistinguishable ($P > 0.05$) from the averages derived from the population- and community-level analyses (Fig. 1 and Table 1). In **a**, data have been standardized by subtracting from each measurement the estimated site-specific intercept; that is, the estimated rate of CH_4 emission at fixed temperature, $\ln F(T_C)$, where T_C is the average of the temperature measurements in the field emissions data set (15.6 °C). This standardization is used for visualization of the data only.

to the natural logarithm of the CH_4:CO_2 flux data. As predicted, this ratio increases with increasing temperature for the majority of the field sites (35 of 38), yielding average temperature dependence across sites (characterized by $\overline{E_{M:C}}$ in equation (2), see Methods) of 0.71 eV (95% confidence interval of 0.46–0.97 eV; Fig. 3). This finding suggests that, on average, the relative contribution of CH_4 to total greenhouse gas emissions increases with seasonal increases in temperature due to the differences in the biochemical kinetics of methanogenesis, respiration and photosynthesis.

Our analyses demonstrate that, on average, CH_4 emissions for a wide range of ecosystems show a temperature dependence similar to that of methanogenesis by methanogen populations in culture and CH_4 production from incubations of microbial communities from sediments. Moreover, this average temperature dependence is much higher than that of both respiration[11] and photosynthesis[12], resulting in a greater relative contribution of CH_4 to total carbon emissions at higher temperatures. Although this consistent temperature dependence of CH_4 fluxes from microbes to ecosystems is remarkable, our results also emphasize that temperature is not the only variable that controls CH_4 emissions from ecosystems. The substantial site-to-site variation we report for the temperature dependence of ecosystem-level CH_4 emissions (see Fig. 2b and Table 1) highlights the importance of other variables in driving deviations from the underlying physiological response. Moreover, the average apparent activation energy we report here ($\overline{E_M} = 0.96$ eV), based on the seasonal temperature dependence of CH_4 emissions within sites, does not apply across sites along geographic temperature gradients. Indeed, the response of CH_4 emissions to geographic variation in temperature (Extended Data Figs 1 and 2) is consistently weaker than the

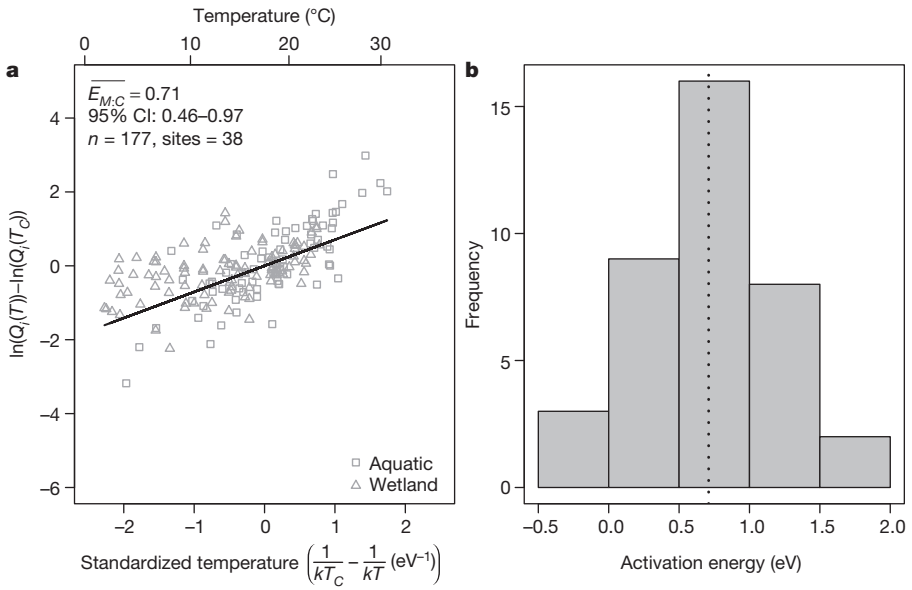


Figure 3 | Temperature dependence of the CH_4:CO_2 emission ratio. Temperature dependencies are characterized in **a** using a mixed-effects model by fitting a Boltzmann–Arrhenius function with site-level random effects on the temperature dependence and ratio at fixed temperature (see equation (2) in the Methods). The fitted solid line in **a** corresponds to the average temperature dependence estimated from the mixed-effects model ($\overline{E_{M:C}} = 0.71$ eV). The distribution of site-level temperature dependencies (presented as a histogram of slope estimates in **b**) yields an average ($\overline{E_{M:C}} = 0.71$ eV, represented by a dashed line in **b**) that is positive and significant ($P > 0.05$). Data in **a** have been standardized by subtracting from each measurement the estimated site-specific intercept; that is, the estimated ratio of CH_4:CO_2 emissions at fixed temperature, $\ln F(T_C)$, where T_C is the average temperature for the ratio data (15.3 °C). This standardization is used for visualization of the data only.

response to seasonal variation within sites (see Fig. 2). Furthermore, for natural wetlands, which span the broadest gradient in average temperature, we observe a significant negative correlation between the site-level intercept—that is, CH_4 emissions at fixed temperature, $\ln F(T_C)$ (see equation (1) in Methods)—and average site temperature (Extended Data Fig. 2e). These results suggest that temperature-dependent variation in other biotic (for example, methanotrophy[6], substrate supply[3,7], microbial community structure and/or composition[8], physiological acclimation and/or adaptation[24]) and abiotic variables (for example, water-table depth[9,10]) may have an important role in constraining the response of CH_4 emissions to warming in the long-term, as has been suggested for rates of ecosystem respiration[11].

Overall, our findings provide an empirical basis for refining representations of the temperature dependence of CH_4 fluxes in coupled climate–carbon cycle models[15–18,25]. In particular, they provide evidence that the temperature dependencies of methanogenesis, CH_4 production and CH_4 emissions are substantially higher than two other key rate processes in the carbon cycle, heterotrophic respiration and photosynthesis. The observation of a general increase in the CH_4:CO_2 ratio with increasing temperature, driven by the relatively high temperature dependence of CH_4 production, may have important implications for the magnitude of future positive feedbacks between global warming and the carbon cycle given the greater potency of CH_4 compared to CO_2 as a greenhouse gas[2,13].

METHODS SUMMARY

We used linear mixed-effects modelling[26] to quantify the temperature dependencies of metabolic rate processes for pure cultures of methanogens, anaerobic microbial communities and seasonal field surveys of CH_4 emissions by fitting Boltzmann–Arrhenius functions of the form:

$$\ln F_i(T) = \left(\overline{E_M} + \varepsilon_{E_M,i}\right)\left(\frac{1}{kT_C} - \frac{1}{kT}\right) + \left(\overline{\ln F(T_C)} + \varepsilon_{F,i}\right) \quad (1)$$

where $\ln F_i(T)$ is the natural logarithm of the rate of CH_4 flux at absolute temperature, T(K), for some arbitrary experimental unit, i. In this expression, each experimental unit, i, corresponds to a distinct strain of methanogen (population-level analysis), sediment sample (community-level analysis) or site (ecosystem-level analysis). The parameter $\overline{E_M}$ (in eV) corresponds to an average among experimental units for the apparent activation energy, which characterizes the temperature dependence of $F_i(T)$, and k is the Boltzmann constant (8.62×10^{-5} eV K^{-1}). We centred the temperature data separately in each analysis using the mean temperature for the data set, T_C, so that $\ln F(T_C)$ corresponds to an average among experimental units for the rate at T_C. For each analysis, we expect estimates of the apparent activation energy, E_M, and the rate of CH_4 production at a fixed temperature, $\ln F(T_C)$, to vary between experimental units. We account for this variation by treating the slopes and intercepts as random variables with averages of $\overline{E_M}$ and $\overline{\ln F(T_C)}$, respectively, and deviations from these averages of $\varepsilon_{E_M,i}$ and $\varepsilon_{F,i}$ for each experimental unit, i.

We used linear mixed-effects modeling[26] to quantify the temperature dependence of the CH_4:CO_2 emission ratio using a Boltzmann–Arrhenius-type approximation (see Supplementary Information section 1) of the form:

$$\ln Q_i(T) = \left(\overline{E_{M:C}} + \varepsilon_{E_{M:C},i}\right)\left(\frac{1}{kT_C} - \frac{1}{kT}\right) + \left(\overline{\ln Q(T_C)} + \varepsilon_{Q,i}\right) \quad (2)$$

where $Q_i(T)$ is the ratio of CH_4 to CO_2 emissions for site i at temperature T, $\overline{E_{M:C}}$ and $\overline{\ln Q(T_C)}$ are the respective averages across sites for the temperature dependence of this ratio and the magnitude of the ratio at a fixed temperature T_C; $\varepsilon_{E_{M:C},i}$ and $\varepsilon_{Q,i}$ are random-effects terms used to represent site-level deviations from these averages.

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 October 2013; accepted 21 February 2014.
Published online 19 March 2014.

1. Forster, P. *et al.* in *Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) (Cambridge Univ. Press, 2007).
2. Kirschke, S. *et al.* Three decades of global methane sources and sinks. *Nature Geosci.* **6,** 813–823 (2013).
3. Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Rev. Microbiol.* **6,** 579–591 (2008).
4. Westermann, P., Ahring, B. K. & Mah, R. A. Temperature compensation in *Methanosarcina barkeri* by modulation of hydrogen and acetate affinity. *Appl. Environ. Microbiol.* **55,** 1262–1266 (1989).
5. Zinder, S. H., Anguish, T. & Cardwell, S. C. Effects of temperature on methanogenesis in a thermophilic (58°C) anaerobic digester. *Appl. Environ. Microbiol.* **47,** 808–813 (1984).
6. Segers, R. Methane production and methane consumption: a review of processes underlying wetland methane fluxes. *Biogeochemistry* **41,** 23–51 (1998).
7. Whiting, G. J. & Chanton, J. P. Primary production control of methane emission from wetlands. *Nature* **364,** 794–795 (1993).
8. Fey, A. & Conrad, R. Effect of temperature on carbon and electron flow and on the archaeal community in methanogenic rice field soil. *Appl. Environ. Microbiol.* **66,** 4790–4797 (2000).
9. Moore, T. R., Roulet, N. T. & Waddington, J. M. Uncertainty in predicting the effect of climatic change on the carbon cycling of Canadian peatlands. *Clim. Change* **40,** 229–245 (1998).
10. Moore, T. R. & Roulet, N. T. Methane flux: water table relations in northern wetlands. *Geophys. Res. Lett.* **20,** 587–590 (1993).
11. Yvon-Durocher, G. *et al.* Reconciling the temperature dependence of respiration across time scales and ecosystem types. *Nature* **487,** 472–476 (2012).
12. Allen, A. P., Gillooly, J. F. & Brown, J. H. Linking the global carbon cycle to individual metabolism. *Funct. Ecol.* **19,** 202–213 (2005).
13. Bridgham, S. D., Cadillo-Quiroz, H., Keller, J. K. & Zhuang, Q. Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. *Glob. Change Biol.* **19,** 1325–1346 (2013).
14. Bastviken, D., Tranvik, L. J., Downing, J. A., Crill, P. M. & Enrich-Prast, A. Freshwater methane emissions offset the continental carbon sink. *Science* **331,** 50 (2011).
15. Gedney, N., Cox, P. M. & Huntingford, C. Climate feedback from wetland methane emissions. *Geophys. Res. Lett.* **31,** L20503 (2004).
16. Riley, W. J. *et al.* Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in CESM. *Biogeosciences* **8,** 1925–1953 (2011).
17. Spahni, R. *et al.* Constraining global methane emissions and uptake by ecosystems. *Biogeosciences* **8,** 1643–1665 (2011).
18. Wania, R., Ross, I. & Prentice, I. C. Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1.3.1. *Geosci. Mod. Dev.* **3,** 565–584 (2010).
19. Christensen, T. R. *et al.* Factors controlling large scale variations in methane emissions from wetlands. *Geophys. Res. Lett.* **30,** 1414 (2003).
20. Crill, P. M. *et al.* Methane flux from Minnesota peatlands. *Glob. Biogeochem. Cycles* **2,** 371–384 (1988).
21. Walter, B. P. & Heimann, M. A process-based, climate-sensitive model to derive methane emissions from natural wetlands: application to five wetland sites, sensitivity to model parameters, and climate. *Glob. Biogeochem. Cycles* **14,** 745–765 (2000).
22. Eyring, H. The activated complex and the absolute rate of chemical reactions. *Chem. Rev.* **17,** 65–77 (1935).
23. Yvon-Durocher, G., Montoya, J. M., Woodward, G., Jones, J. I. & Trimmer, M. Warming increases the proportion of primary production emitted as methane from freshwater mesocosms. *Glob. Change Biol.* **17,** 1225–1234 (2011).
24. Bradford, M. A. *et al.* Thermal adaptation of soil microbial respiration to elevated temperature. *Ecol. Lett.* **11,** 1316–1327 (2008).
25. Melton, J. R. *et al.* Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences* **10,** 753–788 (2013).
26. Zuur, A., Ieno, E., Walker, N. & Saveliev, A. *Mixed Effects Models and Extensions in Ecology with R* (Springer Verlag, 2009).

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

Acknowledgements We thank M. Trimmer for early discussions that inspired much of this work, as well as P. Cox and T. Lenton for comments on earlier drafts of the manuscript.

Author Contributions G.Y.-D., D.B. and C.G. had initial discussions. G.Y.-D. conceived the study, analysed the data and wrote the first draft of the manuscript. D.B., P.A.d.G., C.G., N.T.-D., R.C. and A.S. contributed original data. A.P.A. wrote the theory for the CH_4:CO_2 temperature dependence. All authors contributed to revisions of 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 G. Y.-D. (g.yvon-durocher@exeter.ac.uk).

METHODS

Statistical analyses. We used linear mixed-effects modelling[26] to quantify the temperature dependencies of metabolic rate processes for pure cultures of methanogens, anaerobic microbial communities, and seasonal field surveys of CH_4 emissions, by fitting Boltzmann–Arrhenius functions of the form:

$$\ln F_i(T) = \left(\overline{E_M} + \varepsilon_{E_M,i}\right)\left(\frac{1}{kT_C} - \frac{1}{kT}\right) + \left(\overline{\ln F(T_C)} + \varepsilon_{F,i}\right) \quad (1)$$

where $\ln F_i(T)$ is the natural logarithm of the rate of CH_4 flux at absolute temperature, T (K), for some arbitrary experimental unit, i. In this expression, each experimental unit, i, corresponds to a distinct strain of methanogen (population-level analysis), sediment sample (community-level analysis) or site (ecosystem-level analysis). The parameter $\overline{E_M}$ (in eV) corresponds to an average among experimental units for the apparent activation energy, which characterizes the temperature dependence of $F_i(T)$, and k is the Boltzmann constant (8.62×10^{-5} eV K^{-1}). We centred the temperature data separately in each analysis using the mean temperature for the data set, T_C, so that $\ln F(T_C)$ corresponds to an average among experimental units for the rate at T_C. For each analysis, we expect estimates of the apparent activation energy, E_M, and the rate of CH_4 production at a fixed temperature, $\ln F(T_C)$, to vary between experimental units, which we account for by treating the slopes and intercepts as random variables with averages of $\overline{E_M}$ and $\overline{\ln F(T_C)}$, respectively, and deviations from these averages of $\varepsilon_{E_M,i}$ and $\varepsilon_{F,i}$ for each experimental unit, i.

We used linear mixed-effects modeling[26] to quantify the temperature dependence of the CH_4:CO_2 emission ratio using a Boltzmann–Arrhenius-type approximation (see Supplementary Information section 1) of the form:

$$\ln Q_i(T) = \left(\overline{E_{M:C}} + \varepsilon_{E_{M:C},i}\right)\left(\frac{1}{kT_C} - \frac{1}{kT}\right) + \left(\overline{\ln Q(T_C)} + \varepsilon_{Q,i}\right) \quad (2)$$

where $Q_i(T)$ is the ratio of CH_4 to CO_2 emissions for site i at temperature T, $\overline{E_{M:C}}$ and $\overline{\ln Q\ (T_C)}$ are the respective averages across sites for the temperature dependence of this ratio and the magnitude of the ratio at a fixed temperature T_C; $\varepsilon_{E_{M:C},i}$ and $\varepsilon_{Q,i}$ are random-effects terms used to represent site-level deviations from these averages.

Mixed-effects models are the most appropriate tool for meta-analyses of the kinds of data compiled for this study because they allow for nested covariance structures, where site- or experimental-unit-level relationships are nested within overall relationships, and can accommodate unbalanced designs, where the number of measurements vary among experimental units[26,27]. This approach enabled us to determine the overall average temperature dependence of flux ($\overline{E_M}$, represented by a 'fixed' effect) for a collection of experiments or sites, i, while accounting for the fact that multiple flux–temperature relationships (corresponding to 'random' effects on the slope ($\varepsilon_{E_M,i}$) and intercept ($\varepsilon_{F,i}$) for each i) are nested within this overall relationship. It also allowed us to explicitly quantify the variation in temperature dependencies among sites (or experimental units).

The appropriateness of this technique is perhaps most easily explained for the analysis of field data in this study. Given that there are many other factors, besides temperature, that influence CH_4 emissions, we expect the estimated apparent activation energy to exhibit variation among sites. In particular, we expect the apparent activation energy to differ from that of the physiological flux whenever other variables that affect the rate process exhibit covariation with temperature (see ref. 11 for a theoretical derivation of this result). In this analysis, we are interested in two types of quantities: first, the fixed effects representing the average temperature response of CH_4 emissions across all 127 sites, $\overline{E_M}$, and the average emission rate at fixed temperature, $\ln F(T_C)$; and second, the random effects ($\varepsilon_{E_M,i}$,$\varepsilon_{F,i}$) representing the magnitudes of the deviations from these fixed effects for each of the 127 sites. The deviation from the fixed effects for any given site will be driven by other variables, besides temperature, that affect CH_4 emission and seasonally covary with temperature in the field; for example, methanotrophy[6], water-table depth[9,10], substrate supply[7], community composition[8]. Similar reasoning can be used to justify applying the mixed-effects modelling approach to the population- and community-level data. Although this statistical approach does not allow us to identify the particular variables that contribute to differences in flux–temperature relationships among sites or experimental units, it does allow us to quantify the overall magnitude of their effects as standard deviations of the random-effects terms. Given that the magnitudes of these deviations are assumed to represent a sample from a multivariate normal distribution, the resulting model can be generalized to a larger population of sites.

Parameters were estimated by fitting the Boltzmann–Arrhenius equation to the data using mixed-effects models via the 'lmer' function in the 'lme4' package of R statistical software (v 3.0.2)[28]. For these analyses, we adopted a top-down approach, starting with the most complex model, to determine the significance of the fixed and random effects in a two-stage analysis[26]. In stage one, we determined whether it was necessary to include random effects corresponding to variation in both the slope and intercept among units, i, by using a likelihood ratio test[26,29] to assess the improvement in model fit going from a null model, which includes all potential fixed effects and only one random effect (corresponding to intercept variation), to an alternative model, which includes all potential fixed effects and both random effects (that is, on the slope and intercept). Analysis revealed that the random-effects structure that best described each data set included random variation in both the slope and the intercept (see Table 1 for details of the statistics). In the second stage, we applied the random-effects structure determined in stage one to assess the significance of the fixed effects (averages across sites for the apparent activation energy and intercept), and other potential covariates, using likelihood ratio tests. The results of these likelihood ratio tests are given below and in the main text.

Data compilations. For population-level experiments, measurements of rates of methanogenesis and temperature from cultures of methanogenic Archaea were compiled based on an exhaustive search of the literature. Experiments were only included in the analysis if they met the following criteria: rates were measured at three or more distinct temperatures at or below the optimum; rates were measured during the logarithmic (that is, exponential) phase of population growth, either in batch or continuous culture; rates could be normalized by population biomass. These criteria were met for 12 separate strains. Only data up to the optimum temperature were analysed. Two experiments of enrichment cultures for acetic acid conversion to CH_4 were also included because these culturing techniques result in numerical dominance by methanogenic populations[30]. In our analysis, these enrichment cultures had estimates of the apparent activation energy within the range exhibited by the pure cultures.

To explore further the temperature dependence of metabolic rate processes in methanogens, we compiled additional data on methanogen growth rates, which have been reported more frequently in the literature than rates of methanogenesis. Given the rapid growth rates of these populations (that is, doubling times of less than 1 day), and that the energy transformations required to fuel this growth result from methanogenesis, growth rates of methanogens are likely to be closely linked to rates of methanogenesis and hence ecosystem-level CH_4 emissions. Growth rate data meeting the same three criteria listed above for methanogenesis were compiled for 21 strains of methanogens, and encompass both the acetoclastic and hydrogentrophic groups.

Mixed-effects model analysis indicates that the average activation energies are not significantly different for rates of methanogenesis and growth (likelihood ratio test comparing a null model with a single activation energy to an alternative model with separate activation energies for methanogenesis and growth rate; $\chi^2 = 0.39$, d.f. = 1, $P = 0.53$). These findings suggest that methanogen growth efficiency is approximately independent of temperature for populations that are growing exponentially and are not therefore substrate limited. We also found no significant difference between the acetoclastic or hydrogenotrophic methanogens in the temperature dependencies of either growth rate or methanogenesis (likelihood ratio test comparing a null model with a single activation energy to an alternative model with separate activation energies for acetoclastic or hydrogenotrophic methanogens; $\chi^2 = 1.30$, d.f. = 1, $P = 0.25$). We therefore combined the data on rates of growth and methanogenesis for both groups to characterize an overall average temperature dependence for these metabolic rate processes in methanogen populations (Fig. 1a).

For community-level experiments, rates of CH_4 production were compiled from the literature from experiments in which sediment samples from aquatic, wetland and rice-paddy ecosystems, where incubated anaerobically under laboratory conditions at multiple temperatures. Our data compilation consists of 402 estimates of CH_4 production from 47 separate anaerobic communities sampled from 16 aquatic ecosystems, 26 wetlands and 5 rice-paddy ecosystems. Unlike the culture data, these data characterize the overall effect of temperature on the production of CH_4 by the entire sediment microbial community[31,32].

CH_4 is produced as the terminal product of organic matter decomposition, and is known to be dependent on a series of syntrophic associations with anaerobic bacteria, protozoa and fungi, which supply the methanogens with the precursors for methanogenesis[3], H_2, CO_2 and acetate. It is accomplished by the more or less sequential reduction of NO_3^-, Mn(IV), Fe(III), and SO_4^{2-}, before CH_4 production becomes the sole process of organic matter degradation[31,32]. The length of this reduction period, during which organic matter is oxidised to CO_2, and which precedes the production of CH_4, is determined by the relative amounts of organic matter and inorganic electron acceptors in the sediment[31]. These factors are likely to vary between experiments carried out on sediments from different sites. Thus, between-experiment variation in the length of the incubation period over which the estimate of CH_4 production is determined may influence the measured rate, and may also interact with temperature (for example, if the sequential reduction of electron acceptors proceeds faster at higher temperatures[32]) to influence the estimated apparent activation energy.

The experimental techniques applied in these studies were heterogeneous with respect to the duration over which the flux of CH_4 was determined—ranging from 2 to 30 days—and with respect to whether or not there was a pre-incubation period. We found no evidence that length of the incubation duration affected estimates of the apparent activation energy when this variable was included as a fixed factor in our linear mixed-effects model (likelihood ratio test comparing an alternative model including an interaction between temperature and the length of the incubation duration to a null model with no interaction term: $\chi^2 = 0.95$, d.f. = 1, $P = 0.33$). These findings suggest that the estimated apparent activation energy is largely unaffected by potential effects of temperature on the speed of alternative electron acceptor reduction. Among the 47 studies, 15 entailed pre-incubation periods ranging from 1 to 60 days in the dark at 4 °C. We found no evidence that the presence or absence of an incubation period affected the estimated temperature dependence when this variable was included as a binary fixed factor in our mixed-effects model (likelihood ratio test comparing a null model with a single activation energy to an alternative model with separate activation energies for experiments with and without pre-incubation periods; $\chi^2 = 1.92$, d.f. = 1, $P = 0.38$). In summary, estimates of the apparent activation energy appear to be largely unaffected by differences in experimental protocols among studies.

Our database of ecosystem-level CH_4 emission data comprises both published data compiled from the literature and unpublished data analysed here for the first time (see below for data-collection methods for the unpublished data set). In total, this database comprises 1,553 estimates of CH_4 emission and temperature taken from 127 sites across the globe, including 57 aquatic ecosystems, 51 wetlands and 19 rice paddies (Extended Data Fig. 1). CH_4 emissions were estimated using one of three methods: the floating-chamber technique; the eddy-covariance method; or occasionally (<5% of the estimates) the modelled-diffusion method, which yields estimates based on the concentration of CH_4 just below the air–water interface. The estimated apparent activation energies of seasonal CH_4 emissions did not vary significantly among the three methods when this variable was included as a fixed factor in the mixed-effects model (likelihood ratio test comparing a null model with a single activation energy to an alternative model with separate activation energies for each measurement method; $\chi^2 = 0.52$, d.f. = 2, $P = 0.77$), which justifies combining data obtained using all three methods for the analysis.

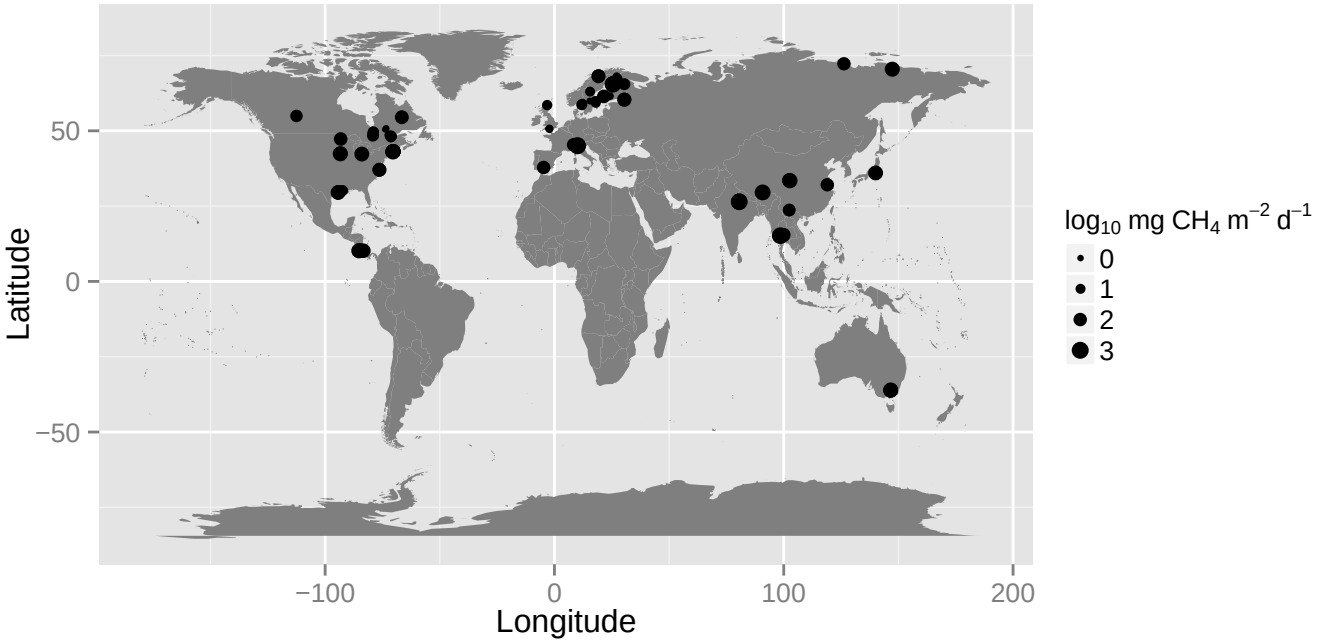
We assessed potential differences in the temperature dependence of CH_4 emissions among three 'ecosystem types'—aquatic, wetland and rice-paddy ecosystems—owing to differences in the physical and chemical characteristics of these ecosystems. 'Aquatic' ecosystems were defined as those that have standing water throughout the year, and included lakes and rivers. Wetlands were defined as natural systems with either permanently or seasonally water-saturated soils, and include peatlands, bogs and swamps. Finally, rice paddies were defined as man-made water-saturated sites established for rice cultivation. We separated these systems for the analysis, because the differences in their physical and chemical characteristics have the potential to modify the seasonal temperature dependence of CH_4 emission in a systematic way. For example, methanotrophy in aerobic zones can re-oxidize up to 90% of the CH_4 produced in anaerobic sediments[6]. Among these three ecosystem types, aquatic systems (lakes and rivers) are expected to have the largest aerobic zones beneath the air–water interface, and may therefore be expected to exhibit temperature dependencies that deviate furthest from the intrinsic temperature dependence of methanogenesis. However, as noted in the main text, although the temperature dependencies clearly varied between sites (Fig. 2b), we found no evidence of systematic differences in temperature dependence between the three ecosystem types (likelihood ratio test comparing a null model with a single activation energy to an alternative model with separate activation energies for each ecosystem type; $\chi^2 = 4.97$, d.f. = 2, $P = 0.10$). This finding is broadly consistent with recent studies[33–35], which show that methanotrophy is primarily limited by substrate availability under *in situ* CH_4 concentrations in aquatic ecosystems, and exhibits no temperature dependence when CH_4 is limiting. Therefore, our results, and those of refs 33–35, suggest that methanotrophy generally does not interact with CH_4 production and temperature under *in situ* CH_4 concentrations. Rather, methanotrophy seems to track the seasonal dynamics of CH_4 production, resulting in little modulation of the observed temperature dependence.

Another factor that may systematically affect the observed temperature dependence of CH_4 emissions, at least for the fluxes estimated through the floating-chamber technique, is the relative contributions of fluxes attributable to diffusion versus ebullition. This is because ebullition has the potential to bypass the methanotrophic microbial populations in the oxic zones of the sediment and water column. We believe that the estimates of CH_4 emissions in our database are largely attributable to diffusive fluxes, due to the extensive sampling effort (in both space and time) required to accurately quantify ebullition fluxes. Although none of our flux estimates could be partitioned into diffusive and ebullition fluxes, we attempted to assess indirectly whether ebullition influenced the estimated temperature dependence by including the duration over which the flux chamber was deployed as a covariate in our mixed-effects model. In carrying out this analysis, we reasoned that ebullition fluxes were more likely to be assayed for chambers deployed for longer periods of time. However, we found no evidence of a significant interaction between 'flux duration' and 'temperature' in the mixed-effects model (likelihood ratio test comparing an alternative model including an interaction between temperature and flux duration to a null model with no interaction term: $\chi^2 = 0.85$, d.f. = 1, $P = 0.36$) when analysing fluxes estimated using the floating-chamber technique ($n = 1{,}007$, 118 sites).

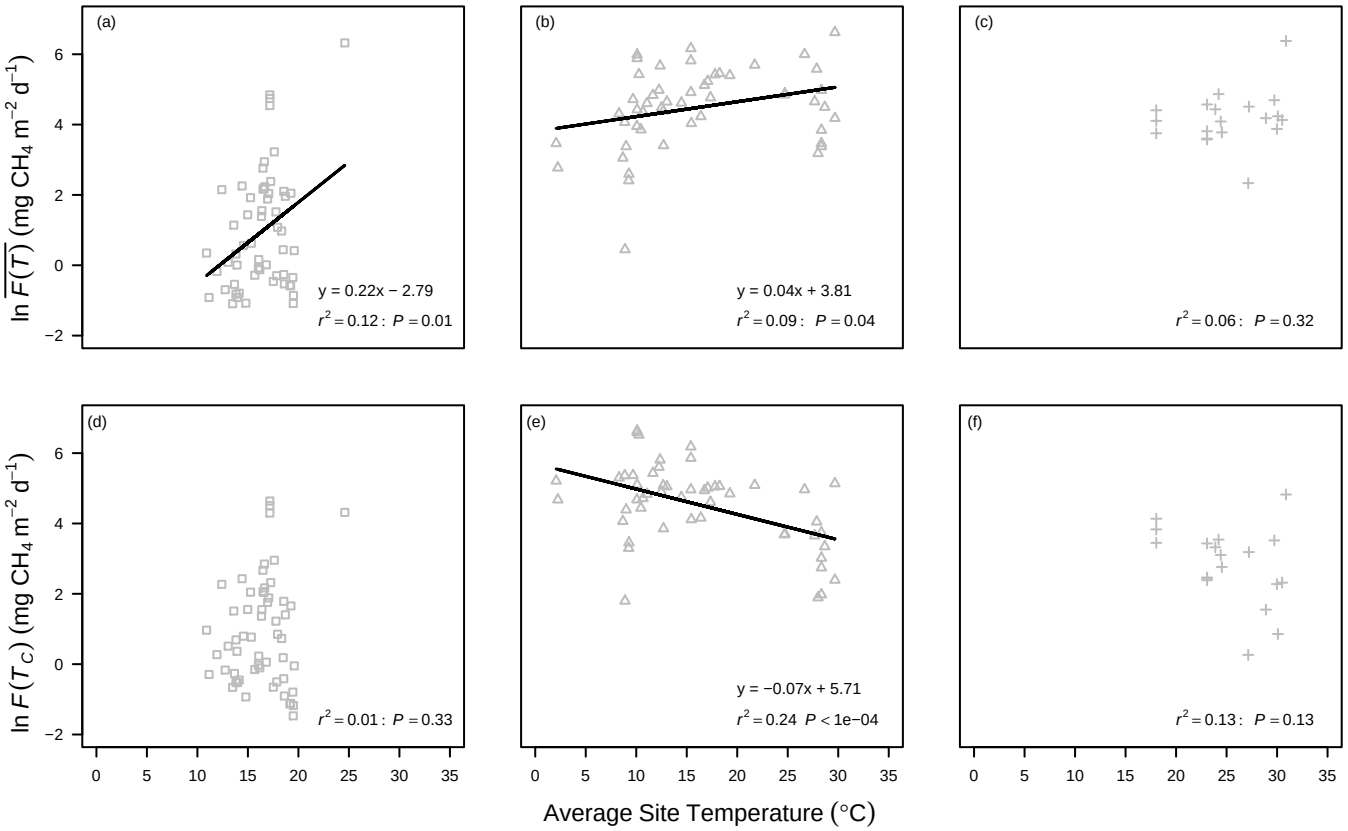
The unpublished emission data included in our analysis come from lakes and rivers located in four distinct boreal regions of Québec, eastern Canada: Abitibi south (48.5° N, 79° W), Abitibi Borth (51° N, 79° W), Chibougamau (49.5° N, 74° W) and Chicoutimi (48° N, 71° W), and in boreal Sweden (58–59° N, 12–18° E). Geographic and climatic details of these regions can be found in ref. 36 and at http://www.smhi.se, respectively. The selected lakes and rivers were visited from three to 12 times during the open water season (May to November 2012). Surface-water partial pressures of CO_2 (pCO_2; Canadian sites only) and of CH_4 (pCH_4) were measured using the headspace method. Water was sampled at 0.1–0.5 m from the surface from the deepest measured point of lakes, and near the shores of streams, rivers and wetlands. Samples were taken using a 60-ml polypropylene syringe by extracting 30 ml of water, followed by 30 ml of ambient air, to create a headspace in a 1:1 ratio of ambient air to water. The syringe was then vigorously shaken for 1 min in order to equilibrate the gases in water and air. Ambient pCO_2 was measured *in situ* using an EGM-4 infrared gas analyser (PP-systems). pCH_4 samples (in duplicate) were placed in 30-ml glass vials equipped with crimped rubber stoppers, filled with saturated saline solution, and then stored inverted until subsequent analysis in laboratory with Shimadzu GC-8A Gas chromatograph with FID (flame ionization detector). In Canada, the CO_2 and CH_4 fluxes were measured on the basis of changes in pCO_2 and pCH_4 with time in floating chambers. The chamber was connected in closed loop to an EGM-4 infrared gas analyser (PP-systems), with pCO_2 in the chamber measured every minute for 10 min. The accumulation rate was estimated from these data using linear regression. Samples for pCH_4 were taken from a sampling port every 10 min, and analysed as described above. Flux calculations were carried out following ref. 37. The measurements in Sweden focused on CH_4 only allowing 24 h floating flux chamber deployments. The details of field approaches and flux calculations for these measurements are found in ref. 38.

For ecosystem-level studies including CH_4:CO_2 ratio, the seasonal temperature dependence of the CH_4:CO_2 efflux ratio was estimated for the subset of field sites in our database (38 of 127) with simultaneous measurements of CO_2 and CH_4 emissions.

27. Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24,** 127–135 (2009).
28. R core development team R: a language and environment for statistical computing. http://www.R-project.org (R Foundation for Statistical Computing, 2014).
29. Pinheiro, J. & Bates, D. M. *Mixed-effects models in S and S-PLUS* (Springer Verlag, 2000).
30. Van den Berg, L., Patel, G. B., Clark, D. S. & Lentz, C. P. Factors affecting the rate of methane formation from acetic-acid by enriched methanogeneic cultures. *Can. J. Microbiol.* **22,** 1312–1319 (1976).
31. Yao, H. & Conrad, R. Effect of temperature on reduction of iron and production of carbon dioxide and methane in anoxic wetland rice soils. *Biol. Fertil. Soils* **32,** 135–141 (2000).
32. van Bodegom, P. M. & Stams, A. Effects of alternative electron acceptors and temperature on methanogenesis in rice paddy soils. *Chemosphere* **39,** 167–182 (1999).
33. Lofton, D., Whalen, S. & Hershey, A. Effect of temperature on methane dynamics and evaluation of methane oxidation kinetics in shallow Arctic Alaskan lakes. *Hydrobiologia* **721,** 209–222 (2014).
34. Duc, N. T., Crill, P. & Bastviken, D. Implications of temperature and sediment characteristics on methane formation and oxidation in lake sediments. *Biogeochemistry* **100,** 185–196 (2010).
35. Megonigal, J. P. & Schlesinger, W. H. Methane-limited methanotrophy in tidal freshwater swamps. *Glob. Biogeochem. Cycles* **16,** 1088–1095 (2002).
36. Lapierre, J.-F. & del Giorgio, P. A. Geographical and environmental drivers of regional differences in the lake pCO_2 versus DOC relationship across northern landscapes. *J. Geophys. Res.-Biogeosci.* **117,** (2012).
37. Vachon, D., Prairie, Y. T. & Cole, J. J. The relationship between near-surface turbulence and gas transfer velocity in freshwater systems and its implications for floating chamber measurements of gas exchange. *Limnol. Oceanogr.* **55,** 1723–1732 (2010).
38. Bastviken, D. *et al.* Methane emissions from Pantanal, South America, during the low water season: toward more comprehensive sampling. *Environ. Sci. Technol.* **44,** 5450–5455 (2010).



Extended Data Figure 1 | Global distribution of the field sites included in our analysis of ecosystem-level CH_4 emissions. The size of each point relates to the logarithm (base 10) of the mean rate of CH_4 emission (mg CH_4 per m^2 per day) over the duration of the experiment.



Extended Data Figure 2 | Correlations of average site temperatures with average CH_4 emissions and CH_4 emissions at fixed temperature for globally distributed ecosystems. Average CH_4 emissions are shown in **a**–**c**, and CH_4 emissions at fixed temperature are shown in **d**–**f**. Average site temperature is positively correlated with average CH_4 emissions, $\ln \overline{F(T)}$, for aquatic ecosystems (**a**) and natural wetlands (**b**), although temperature explains only 12% and 9% of the variance, respectively, for CH_4 emissions from these two ecosystem types. In contrast, it is not significantly correlated with average CH_4 emissions for rice paddies (**c**). Average site temperature is also not correlated with CH_4 emissions at fixed temperature, $\ln F(T_C)$ (where T_C is the average temperature across the field emissions data set (15.6 °C)), for aquatic ecosystems (**d**) and rice paddies (**f**), but is significantly negatively correlated for natural wetlands (**e**). The latter finding suggests that temperature-dependent biotic (for example, methanotrophy, substrate supply, microbial community structure, physiological acclimation and/or adaptation) and abiotic factors (for example, water-table depth) may play an important role in constraining variation in total CH_4 emissions among wetlands along geographic temperature gradients.

LETTER

doi:10.1038/nature12976

Geographical limits to species-range shifts are suggested by climate velocity

Michael T. Burrows[1], David S. Schoeman[2], Anthony J. Richardson[3,4], Jorge García Molinos[1], Ary Hoffmann[5], Lauren B. Buckley[6], Pippa J. Moore[7,8], Christopher J. Brown[9], John F. Bruno[6], Carlos M. Duarte[10,11,12], Benjamin S. Halpern[13,14], Ove Hoegh-Guldberg[9], Carrie V. Kappel[13], Wolfgang Kiessling[15,16], Mary I. O'Connor[17], John M. Pandolfi[18], Camille Parmesan[19,20], William J. Sydeman[21], Simon Ferrier[22], Kristen J. Williams[22] & Elvira S. Poloczanska[3]

The reorganization of patterns of species diversity driven by anthropogenic climate change, and the consequences for humans[1], are not yet fully understood or appreciated[2,3]. Nevertheless, changes in climate conditions are useful for predicting shifts in species distributions at global[4] and local scales[5]. Here we use the velocity of climate change[6,7] to derive spatial trajectories for climatic niches from 1960 to 2009 (ref. 7) and from 2006 to 2100, and use the properties of these trajectories to infer changes in species distributions. Coastlines act as barriers and locally cooler areas act as attractors for trajectories, creating source and sink areas for local climatic conditions. Climate source areas indicate where locally novel conditions are not connected to areas where similar climates previously occurred, and are thereby inaccessible to climate migrants tracking isotherms: 16% of global surface area for 1960 to 2009, and 34% of ocean for the 'business as usual' climate scenario (representative concentration pathway (RCP) 8.5)[8] representing continued use of fossil fuels without mitigation. Climate sink areas are where climate conditions locally disappear, potentially blocking the movement of climate migrants. Sink areas comprise 1.0% of ocean area and 3.6% of land and are prevalent on coasts and high ground. Using this approach to infer shifts in species distributions gives global and regional maps of the expected direction and rate of shifts of climate migrants, and suggests areas of potential loss of species richness.

By reorganizing natural systems[2], climate change is poised to be one of the greatest threats to biodiversity of this century[3], compromising the integrity, goods and services of living systems[1]. Increased understanding of how species distributions and persistence are likely to be affected can inform effective conservation under climate change, as part of a range of considerations[9]. Predictions from complex models may incorporate ecological complexity but come with a high degree of uncertainty[10]. A simpler approach is to consider the local speed and direction of shifting climate contours as an expectation of how species' distributions would have to shift to track the location of their thermal niches[6,10]. This is the velocity of climate change[6,7]: the temporal trend divided by the spatial gradient in a climate variable such as temperature[6] or precipitation[10,11]. Landscapes and seascapes have different patterns of climate velocity[7] and consequential residence times of climate, giving different implications for species' persistence and priorities for conservation[12]. Patterns of shifts in distributions of many taxa in the ocean have already been shown to follow the velocity of climate[5].

Here we use velocity-derived trajectories to indicate global regions susceptible to effects of geographical limits to climate-driven distribution shifts. Climate trajectories are paths that points on an isotherm will travel over specific periods (Fig. 1, see Methods for details), integrating spatially variable speeds and directions of climate velocity along the paths to show effects that static velocity maps cannot. Geographical limits to trajectories, either barriers such as coasts and mountains or lack of connections to cooler or warmer environments, suggest limits to climatic niche shifts and, by inference, species persistence at a local level or replacement from warmer environments (Figs 2 and 3). Velocity fields were derived for 1960–2009 for land and ocean surface temperatures[7] on a 1° grid, allowing inference at a global scale, but sacrificing small-scale detail such as thermal minima on mountain tops or sharp gradients associated with ocean fronts[13].

We categorized types of trajectory behaviour using trajectory length and the percentages of trajectories starting in, ending in and passing through cells (Fig. 1 and Extended Data Fig. 1). Short trajectories indicated non- or slow-moving thermal niches. Cells were classed as relative climate sinks if a high proportion of trajectories terminated there. Absolute climate sinks were also distinguished: coastal climate sinks where trajectories were blocked by coasts, and internal climate sinks where velocities in neighbouring cells converged. Cells were classed as climate sources if no trajectories ended there. Thereafter, cells with a high proportion of trajectories passing through were classed as corridors. Divergence cells were identified as those where fewer trajectories ended than started in that cell, and convergence cells if the opposite were true.

Uncertainty evaluated by resampling of annual average temperature maps gave a 'likely' (>66%) consistency of designation of trajectory classes between bootstrap samples and the overall classification for 59% of ocean and 72% of land cells (Fig. 2 and Extended Data Figs 2–4). Consistency was <66% where spatial gradients and temperature trends are most uncertain, such as where inter- and multidecadal climate variability dominates as for the El Niño-Southern Oscillation in the tropical Pacific. Classification of 'very likely' (>90%) was achieved for 40% of land and 26% of ocean cells, mainly sources, coastal sinks and low velocity areas (Extended Data Fig. 4).

[1]Department of Ecology, Scottish Association for Marine Science, Scottish Marine Institute, Oban, Argyll, PA37 1QA, Scotland, UK. [2]School of Science and Engineering, University of the Sunshine Coast, Maroochydore, Queensland QLD 4558, Australia. [3]Climate Adaptation Flagship, CSIRO Marine and Atmospheric Research, Ecosciences Precinct, GPO Box 2583, Brisbane, Queensland 4001, Australia. [4]Centre for Applications in Natural Resource Mathematics (CARM), School of Mathematics and Physics, The University of Queensland, St Lucia, Queensland 4072, Australia. [5]Department of Genetics, University of Melbourne, 30 Flemington Road, Parkville, Victoria 3010, Australia. [6]Department of Biology, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3280, USA. [7]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 3DA, UK. [8]Centre for Marine Ecosystems Research, Edith Cowan University, Perth 6027, Australia. [9]The Global Change Institute, The University of Queensland, Brisbane, Queensland 4072, Australia. [10]The UWA Oceans Institute, University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia. [11]Department of Global Change Research, IMEDEA (UIB-CSIC), Instituto Mediterráneo de Estudios Avanzados, Esporles 07190, Spain. [12]Department of Marine Biology, Faculty of Marine Sciences, King Abdulaziz University, PO Box 80207, Jeddah 21589, Saudi Arabia. [13]Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106, USA. [14]Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK. [15]GeoZentrum Nordbayern, Paläoumwelt, Universität Erlangen-Nürnberg, Loewenichstrasse 28, 91054 Erlangen, Germany. [16]Museum für Naturkunde, Invalidenstr asse 43, 10115 Berlin, Germany. [17]Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver V6T 1Z4, Canada. [18]School of Biological Sciences, Australian Research Council Centre of Excellence for Coral Reef Studies, The University of Queensland, Brisbane, Queensland 4072, Australia. [19]Integrative Biology, University of Texas, Austin, Texas 78712, USA. [20]Marine Institute, Drake Circus, University of Plymouth, Devon PL4 8AA, UK. [21]Farallon Institute for Advanced Ecosystem Research, 101 H Street, Suite Q, Petaluma, California 94952, USA. [22]Climate Adaptation Flagship, CSIRO Ecosystem Sciences, GPO Box 1700, Canberra, Australian Capital Territory 2601, Australia.

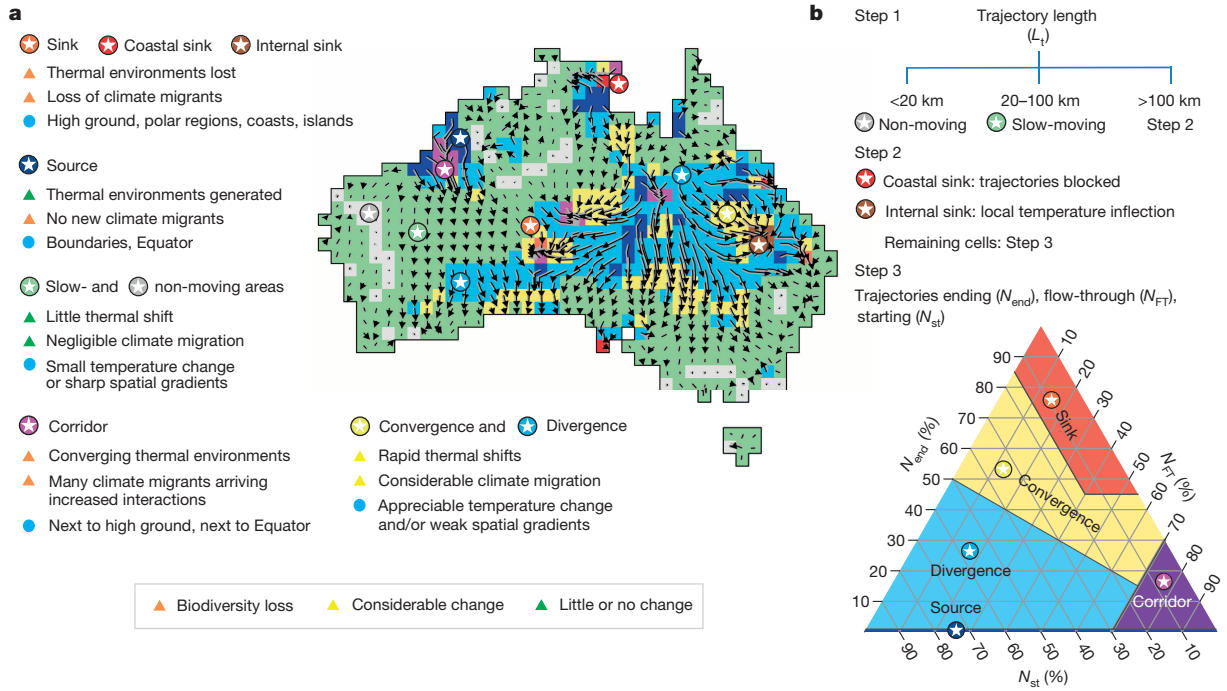


Figure 1 | Climate-change features emerging from the properties of climate velocity trajectories using the Australian landmass as an example. **a**, Trajectories (indicated by arrows) show predicted 50-year (1960–2009) shifts based on isotherm velocity. Features described by trajectory properties are shown; typical locations and physical characteristics are indicated on the map (stars denoting typical locations of sinks, sources, corridors, convergence and divergence, and slow- and non-moving areas). The respective implications of these physical characteristics for climate migrants are listed in the boxes (triangles). **b**, Hierarchical sequential classification of climate change features based on length of trajectories (step 1), geographical features (step 2) and the relative abundance of trajectories ending in, starting from and flowing through each cell (step 3).

The proportions of land and sea areas classed as climate sources and sinks are similar at a global scale (sources: land 17.4%, ocean 15.9%; sinks: land 3.6%, ocean 1.0%), but the latitudinal pattern differs (Fig. 2 and Extended Data Figs 5–9 for regional maps). For ocean surface temperatures, climate sources are concentrated within 10° latitude of the Equator (Figs 2b, d), a pattern not evident on land (Fig. 2a, c).

Topographic complexity on land generates more local warm and cold source and sink areas compared to the ocean. Importantly, 12.0% of land trajectories and 5.4% of ocean trajectories terminate in sinks, representing 'lost' local conditions: 6.1% (5.0%) land (ocean) ending in coastal sinks, 5.9% (0.4%) in internal sinks. These losses are analogous to 'disappearing climates'[14] but here result from local connections

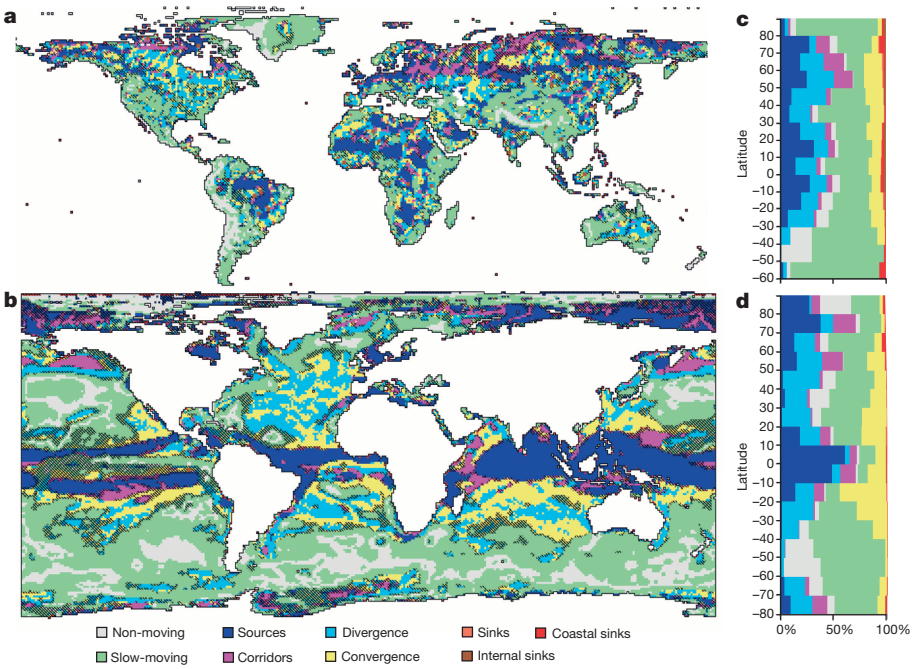


Figure 2 | Global patterns of climate trajectory classes. **a–d**, On land (**a**) and in the ocean (**b**), and proportional areas by latitude (**c**, **d**). Uncertainty in classification of areas is shown by the cross hatching on (**a**, **c**): <66% of 500 bootstrap class maps consistent with underlying class map. Non-hatched areas have classification that is 'likely' (>66% consistent). Uncertainty between convergence and divergence areas and slow and non-moving areas is not shown (see Extended Data Fig. 2).

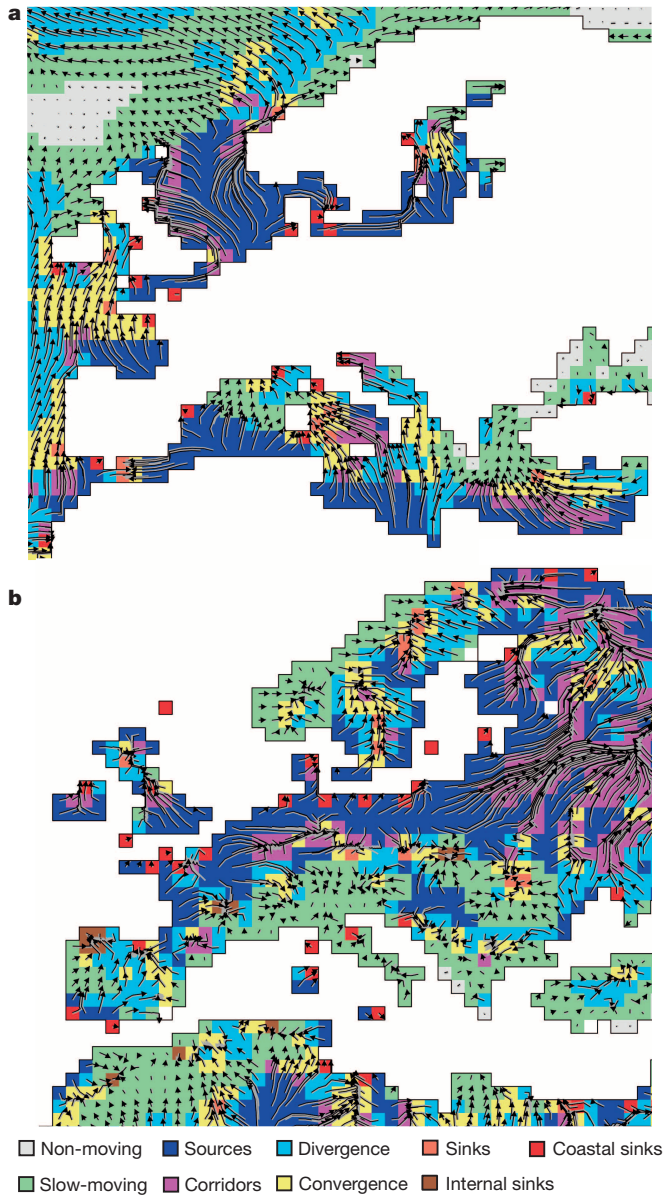


Figure 3 | Regional patterns of climate velocity trajectory classes for land and sea surface temperatures. a, **b**, We show 50-year trajectories (shown by arrows) of climates (1960–2009) for Europe for 1° grid cells, overlaid on climate trajectory classes for ocean (**a**) and land (**b**).

to thermal minima rather than global loss of combinations of climatic conditions.

In the ocean, coastal sources form where poleward trajectories depart from coastlines, as in North Africa, and coastal sinks where equatorward coastlines block trajectories, as in southern France (Fig. 3a). The opposite occurs on land to a degree: sources on equatorward coastlines and sinks along poleward coastlines. However, cooler, higher regions of continents, such as mountain ranges in Europe, 'attract' trajectories towards the interior, disrupting the otherwise-poleward flow, resulting in internal sinks at greater elevations (Fig. 3b). Elevated land is also more likely to comprise non- or slow-moving areas. Corridors are evident in areas of convergent trajectories and high climate velocities, as in the northern North Sea and in southern Germany (Fig. 3a, b).

Future climate trajectories for sea surface temperature (SST) using an example global climate model (ACCESS1.0) for 2006–2100 produce similar patterns of shifting climates to those seen globally for 1960–2009. Sources were indicated at thermal maxima around the equator (Fig. 4a, c). Greater warming in the RCP 8.5 'business as usual' scenario[8] (2.95 °C increase in annual average SST for 2080–2099 over 1960–2009) results in longer trajectories than the RCP 4.5 scenario run (1.75 °C increase) and a doubling in size of areas identified as sources (RCP 8.5, 34.1% of scenario cells as sources versus 19.8% for RCP 4.5). Local-scale patterns were also similar (Fig. 4b, d), as thermal gradients that determine trajectory direction are reproduced in the climate models. No-analogue climate futures[15] will emerge in source areas with novel climates[14], particularly around the equator, but not in sources that result from coastal barriers disconnected from similar climates elsewhere.

Each climate trajectory feature has different implications for the migration patterns of climate-sensitive species through climate connectivity (Extended Data Table 1). Species richness in climate source areas may decline, because climate migrants leaving may not be replaced: sources lack connection routes for new migrants. Converging temperature isotherms in relative climate sinks may concentrate climate migrants, increasing local diversity, whereas absolute climate sinks represent climatic dead-ends where species cannot spread along thermal gradients into cooler areas, creating potential for local extinction. Large numbers of trajectories traversing a limited pathway suggest important corridors for climate migrants.

Patterns of distribution shifts indicated by trajectory behaviour raise questions for ecology and conservation of climate-sensitive species[12], particularly when considered alongside the magnitude and latitude of

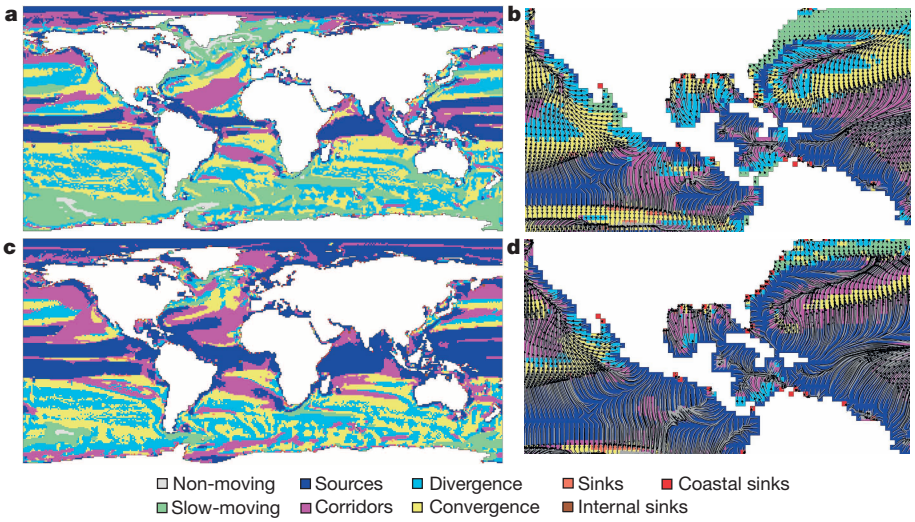


Figure 4 | Global and regional patterns of 50-year climate trajectory classes based on trends from ensembles of global climate models for 2006–2100. a–d, Classes for two CMIP5 scenarios: the 4.5 W m^{-2} (RCP 4.5) scenario (**a**, **b**) and the 8.5 W m^{-2} (RCP 8.5) scenario (**c**, **d**), derived from temperature trends and spatial patterns in temperature in data from the ACCESS1.0 CMIP5 global climate model. Arrows show the expected shift in location of points along isotherms over a 50-year period.

climate change[16]. Long 'climate residence times' in areas of low velocity[10] have been associated with high levels of historical endemism[17] and have led to such areas being proposed as genetic and evolutionary refugia[18]. The non-replacement of climate migrants in climate source areas may result in net loss of species richness, and facilitate the establishment of new species into abandoned niches[19], such as in the eastern Mediterranean Sea via the Suez Canal[20]. A larger number of inbound climate migrants in convergent areas, corridors and sinks implies that local communities should face greater reshuffling of species and novel ecological interactions[21], and compromised genetic diversity through gene swamping[22] but with increased adaptive gene flow[23]. Climate migrants face local extinction in climate sink areas, unless the species involved can adapt to changed conditions.

Climate velocity is emerging as a good predictor for range shifts in the ocean[4,5], whereas the relationship is still to be investigated on land, although terrestrial species distribution shifts have been related to latitudinal shifts in isotherms[24]. The similarity in trajectory maps of future projections with past reconstructions suggests that the ecological implications and therefore the management actions for conserving biodiversity, as informed by current climate trajectories, could remain effective into the future. Species losses from both source and sink areas may accelerate with climate change, with greater warming resulting in areas of novel and lost climates, suggesting that impacts may be greatest in these areas. The approach developed here offers a rapid global method to quantify and map patterns of shifting thermal niches, and by implication those species tracking thermal conditions, and highlights those areas of the globe that may be at risk from the effects of geographical barriers to climate migrants.

METHODS SUMMARY

Velocity fields were derived for 1960–2009 land and ocean surface temperatures using Hadley Centre HadISST v1.1 and Climate Research Unit CRU TS3.1 (ref. 7) data sets, and for global climate model (GCM) projections for 2006–2100 using example GCM model data from CMIP5 (Coupled Model Intercomparison Project Phase 5) experiments for RCP 4.5 and RCP 8.5 scenarios. Velocity was calculated by dividing the 50-year temperature trend by the spatial gradient in 50-year means, taking the direction from the spatial gradient[6]. Trajectories of climate were obtained by calculating displacements using local velocity. If trajectories hit a coastal barrier and a cooler or warmer cell was found locally then the trajectory continued towards that cell, else the cell was designated as a coastal sink.

The collective behaviour of trajectories was obtained from 0.1°-spaced trajectories projected over 50 years at 0.1-year intervals. Cells were classed (Fig. 1b) as: first, slow or non-moving where 50-year trajectories were less than 100 km; second, coastal and internal sinks, where trajectories halted on coastlines or converged towards a central point; third, five types based on proportions of trajectories starting from, ending in, and flowing through cells. These five types of cells are: sources, cells where no trajectories ended; sinks where many trajectories ended; corridors, cells with a high proportion of trajectories passing through; and divergence and convergence cells as those where fewer/more trajectories ended than started in that cell.

Uncertainty of classification into areas based on trajectory behaviour was estimated using bootstrap resampling ($n = 500$) of the temperature data sets. Each bootstrap sample comprised a random selection with replacement of 50 annual mean temperatures from the original series, from which mapped temporal and spatial temperature trends, and thereby velocity, were calculated. 50-year trajectories based on these maps gave a bootstrap classification of trajectory areas. Consistency of types by grid cell was expressed as the percentage of cells in bootstrap maps that were the same type as the original classification.

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 13 September; accepted 30 December 2013.
Published online 9 February; corrected online 26 March 2014 (see full-text HTML version for details).

1. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486,** 59–67 (2012).
2. Barnosky, A. D. *et al.* Approaching a state shift in Earth's biosphere. *Nature* **486,** 52–58 (2012).
3. Pereira, H. M. *et al.* Scenarios for global biodiversity in the 21st century. *Science* **330,** 1496–1501 (2010).
4. Poloczanska, E. S. *et al.* Global imprint of climate change on marine life. *Nature Climate Change* **3,** 919–925 (2013).
5. Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L. & Levin, S. A. Marine taxa track local climate velocities. *Science* **341,** 1239–1242 (2013).
6. Loarie, S. R. *et al.* The velocity of climate change. *Nature* **462,** 1052–1055 (2009).
7. Burrows, M. T. *et al.* The pace of climate change in marine and terrestrial ecosystems. *Science* **334,** 652–655 (2011).
8. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93,** 485–498 (2012).
9. Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C. & Mace, G. M. Beyond predictions: biodiversity conservation in a changing climate. *Science* **332,** 53–58 (2011).
10. Ackerly, D. D. *et al.* The geography of climate change: implications for conservation biogeography. *Divers. Distrib.* **16,** 476–487 (2010).
11. Ordonez, A. & Williams, J. W. Projected climate reshuffling based on multivariate climate-availability, climate-analog, and climate-velocity analyses: implications for community disaggregation. *Clim. Change* **119,** 659–675 (2013).
12. Gillson, L., Dawson, T. P., Jack, S. & McGeoch, M. A. Accommodating climate change contingencies in conservation strategy. *Trends Ecol. Evol.* **28,** 135–142 (2013).
13. Dobrowski, S. Z. *et al.* The climate velocity of the contiguous United States during the 20th century. *Glob. Chang. Biol.* **19,** 241–251 (2013).
14. Williams, J. W., Jackson, S. T. & Kutzbach, J. E. Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl Acad. Sci. USA* **104,** 5738–5742 (2007).
15. Williams, J. W. & Jackson, S. T. Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ.* **5,** 475–482 (2007).
16. Deutsch, C. A. *et al.* Impacts of climate warming on terrestrial ectotherms across latitude. *Proc. Natl Acad. Sci. USA* **105,** 6668–6672 (2008).
17. Sandel, B. *et al.* The influence of late Quaternary climate-change velocity on species endemism. *Science* **334,** 660–664 (2011).
18. Budd, A. F. & Pandolfi, J. M. Evolutionary novelty is concentrated at the edge of coral species distributions. *Science* **328,** 1558–1561 (2010).
19. Facon, B. *et al.* A general eco-evolutionary framework for understanding bioinvasions. *Trends Ecol. Evol.* **21,** 130–135 (2006).
20. Hiddink, J. G., Ben Rais Lasram, F., Cantrill, J. & Davies, A. J. Keeping pace with climate change: what can we learn from the spread of Lessepsian migrants? *Glob. Chang. Biol.* **18,** 2161–2172 (2012).
21. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421,** 37–42 (2003).
22. Norberg, J., Urban, M. C., Vellend, M., Klausmeier, C. A. & Loeuille, N. Eco-evolutionary responses of biodiversity to climate change. *Nature Climate Change* **2,** 747–751 (2012).
23. Sgrò, C. M., Lowe, A. J. & Hoffmann, A. A. Building evolutionary resilience for conserving biodiversity under climate change. *Evol. Appl.* **4,** 326–337 (2011).
24. Chen, I. C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science* **333,** 1024–1026 (2011).

Acknowledgements This work was conducted as a part of the Towards Understanding Marine Biological Impacts of Climate Change Working Group supported by the National Center for Ecological Analysis and Synthesis, a center funded by the NSF (grant no. EF-0553768), the University of California, Santa Barbara and the State of California. M.T.B., P.J.M. and J.G.M. were supported by the UK Natural Environment Research Council grant NE/J024082/1. D.S. was supported by the Australian Research Council's Collaborative Research Network. J.P. thanks the Australian Research Council Centre of Excellence for Coral Reef Studies for additional support, and A.J.R. was supported by the Australian Research Council Discovery Grant DP0879365 and Future Fellowship Grant FT0991722.

Author Contributions M.T.B., D.S.S., A.J.R. and E.S.P. conceived the research. M.T.B., J.G.M. and D.S.S. analysed the data. M.T.B., D.S.S., A.J.R., E.S.P., J.G.M. and M.I.O. wrote the first draft. M.T.B., D.S.S., A.J.R., J.G.M., A.H., L.B.B., P.J.M., C.J.B., J.F.B., C.M.D., B.S.H., O.H.G., C.V.K.,W.K., M.I.O., J.M.P., C.P., W.J.S., S.F., K.J.W. and E.S.P. contributed equally to discussion of ideas and analyses, and all authors commented on the manuscript.

Author Information Data used in analyses are available from the University of East Anglia Climate Research Unit and the UK Meteorological Office Hadley Centre, with online access at the British Atmospheric Data Centre. Maps are available as Google Earth files on http://www.figshare.com. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.T.B. (mtb@sams.ac.uk).

LETTER

doi:10.1038/nature13010

A suspension-feeding anomalocarid from the Early Cambrian

Jakob Vinther[1], Martin Stein[2], Nicholas R. Longrich[3] & David A. T. Harper[4]

Large, actively swimming suspension feeders evolved several times in Earth's history, arising independently from groups as diverse as sharks, rays and stem teleost fishes[1], and in mysticete whales[2]. However, animals occupying this niche have not been identified from the early Palaeozoic era. Anomalocarids, a group of stem arthropods that were the largest nektonic animals of the Cambrian and Ordovician periods, are generally thought to have been apex predators[3–5]. Here we describe new material from *Tamisiocaris borealis*[6], an anomalocarid from the Early Cambrian (Series 2) Sirius Passet Fauna of North Greenland, and propose that its frontal appendage is specialized for suspension feeding. The appendage bears long, slender and equally spaced ventral spines furnished with dense rows of long and fine auxiliary spines. This suggests that *T. borealis* was a microphagous suspension feeder, using its appendages for sweep-net capture of food items down to 0.5 mm, within the size range of mesozooplankton such as copepods. Our observations demonstrate that large, nektonic suspension feeders first evolved during the Cambrian explosion, as part of an adaptive radiation of anomalocarids. The presence of nektonic suspension feeders in the Early Cambrian, together with evidence for a diverse pelagic community containing phytoplankton[7,8] and mesozooplankton[7,9,10], indicate the existence of a complex pelagic ecosystem[11] supported by high primary productivity and nutrient flux[12,13]. Cambrian pelagic ecosystems seem to have been more modern than previously believed.

T. borealis, from the Early Cambrian Sirius Passet fauna of North Greenland, has been described previously as a possible anomalocarid on the basis of a disarticulated frontal appendage[6]. New fossils not only substantiate the anomalocarid affinities of *Tamisiocaris* but also suggest that it was adapted to prey microphagously on mesozooplankton (please note the revised terminology of the group, see Supplementary Information).

T. borealis is at present known from five isolated frontal appendages and two appendages associated with a head shield. Frontal appendages (Fig. 1) measure ≥120 mm in length, comparable in size to the later *Anomalocaris canadensis*[14], but the total size of the body is not known. As in other anomalocarids, the appendage consists of discrete, sclerotized articles. All specimens are preserved with the ventral spines parallel to the bedding plane, and the articles show no evidence of distortion due to compaction. It is therefore assumed that the articles were transversely compressed, with an oval cross-section in life. The appendage consists of at least 18 articles, versus 14 in *A. canadensis*, for example. Articles are separated by triangular arthrodial membranes (Extended Data Fig. 2b, c). These extend almost to the dorsal margin of the appendage; ventrally, the membrane is 33–50% the length of the articles, suggesting a well-developed flexural ability.

The appendage curves downward distally, with the strongest curvature around the second and third article. The first article is straight, and longer than the next three combined. It bears a single pair of ventral

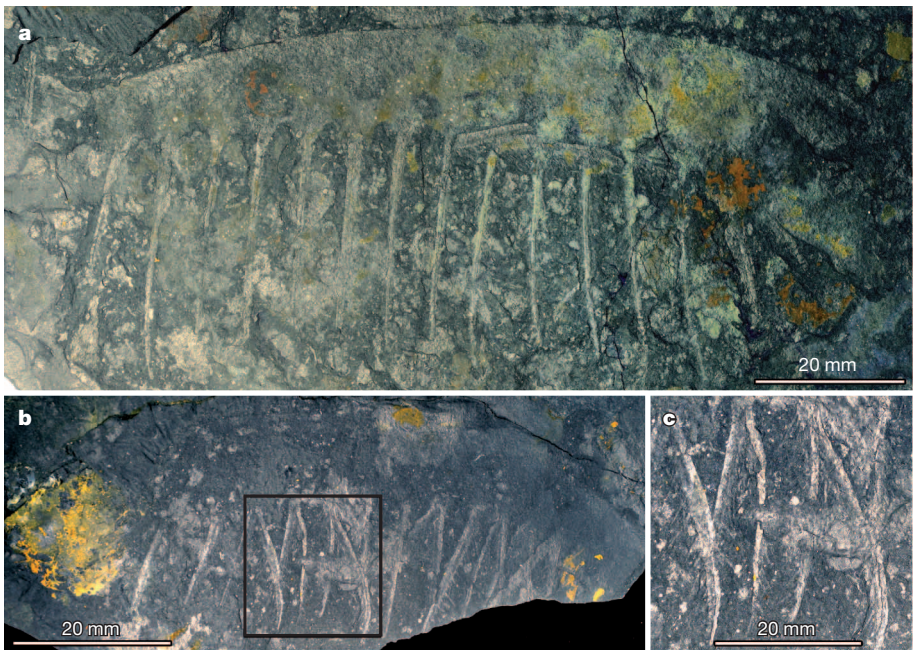


Figure 1 | *T. borealis* Daley and Peel, 2010, frontal appendages from Sirius Passet, Lower Cambrian, North Greenland. **a**, Isolated and relatively complete appendage, MGUH 30500 (Geological Museum at the University of Copenhagen). **b**, Isolated appendage, preserving auxiliary spines in great detail, MGUH 30501. **c**, Detail of spine (boxed area in **b**). All specimens were photographed submerged in water with high-angle illumination.

[1]Schools of Earth Sciences and Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK. [2]Natural History Museum of Denmark, Copenhagen University, Universitetsparken 15, 2100, Denmark. [3]Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK. [4]Palaeoecosystems Group, Department of Earth Sciences, Durham University, Durham DH1 3LE, UK.

spines near its distal margin, which are stout and angled backwards (Fig. 1a), as in *A. briggsi*[5]. The next 17 articles each bear pairs of long and delicate ventral spines inserted at the mid-length of the article. These are evenly spaced along the appendage about 5–6 mm apart. The spines diverge ventrally such that each pair forms an inverted V-shape. Unlike *A. canadensis*, in which longer and shorter spines alternate and taper distally, the ventral spines are all of similar length, measuring 26–27.5 mm along the full length of the appendage (Fig. 1a, b and Extended Data Figs 1–3). A similar condition is seen in *A. briggsi*. The ventral spines curve posteriorly, again as in *A. briggsi*, but unlike in any other anomalocarids. Individual spines appear flattened, with a median rod and thinner lamellar margins (Extended Data Fig. 1c). In addition, ventral spines are frequently kinked, and sometimes broken, suggesting that they were weakly sclerotized and flexible.

As in many other anomalocarids[5,15], the anterior and posterior margins of the ventral spines bear auxiliary spines (Fig. 1c and Extended Data Figs 1c, 2d and 3), but they are unusually long in *Tamisiocaris*—measuring 4.2–5.0 mm in length—and extremely slender. Auxiliary spines form a comb-like array, being spaced 0.3–0.85 mm apart, with a median spacing of 0.49 mm. The length and spacing are such that adjacent spine combs between spines would overlap or interdigitate. A three-dimensional digital reconstruction was produced (Fig. 2a) with the inferred relative proportions in place to infer its dexterity (Fig. 2b).

One specimen consists of two associated appendages in subparallel orientation (Extended Data Fig. 4). Proximally, they join a large, elliptical head shield. The head shield is larger than in *A. canadensis*, but is not enlarged to the same degree as seen in *Peytoia nathorsti* and *Hurdia victoria*. Eyes are not preserved.

The affinities of *Tamisiocaris* were examined in a cladistic analysis to explore its position within the anomalocarids. The analysis recovers a clade consisting of *T. borealis* and *A. briggsi* (Fig. 3). This clade, which we name the Cetiocaridae (cetus: whale, shark or other large marine animal; and caris: sea crab), is diagnosed by long, slender and recurved ventral spines, and the presence of numerous auxiliary spines. *Tamisiocaris* is more specialized, however, in having flexible ventral spines and densely packed auxiliary spines. The cetiocarids are a sister to Hurdiidae, a clade containing *H. victoria*, *P. nathorsti* and related species. Outside these taxa lies a clade of plesiomorphic forms including *A. canadensis*, *A. saron*, *Amplectobelua* spp. and relatives.

The hypothesis that *T. borealis* engaged in suspension feeding can be evaluated by comparisons with extant analogues (Extended Data Fig. 5). Suspension-feeding crustaceans, such as cirripedes (barnacles), atyid shrimp, copepods, cladocerans, mysids and euphausiaceans (krill) share a suite of adaptations for sieving particles out of the water column that are also found in the Cetiocaridae (Extended Data Fig. 5). These include appendages with very elongate, flexible setae and/or setules; regular spacing; and close spacing of setae/setules. These features create a net with a regular mesh size that efficiently traps all particles above a threshold set by the setal spacing. The feeding limbs sieve particles out of the water, concentrate them by contraction, and carry them to the mouth[16]. The suspension-feeding apparatuses of vertebrates have a similar morphology. Suspension-feeding teleosts and some sharks use a mesh formed by long, slender and closely spaced gill rakers. The feeding apparatus of mysticete whales consists of arrays of baleen plates that wear into elongate fringes[17].

The mesh size of the capture apparatus is closely related to prey size: right whales specialize in feeding on small copepods (fringe diameter 0.2 mm), whereas blue whales (fringe diameter 1 mm) feed on larger krill[18]. A survey of diverse suspension feeders, from cladocerans to blue whales, shows a linear relationship between mesh size and minimum prey size (Fig. 4). Although larger prey can be captured, the bulk of the prey is close to the mesh size of the suspension apparatus.

On the basis of the morphologies seen in modern animals, a suspension-feeding anomalocarid would be predicted to have evolved a setal mesh, with large appendages bearing long, flexible setae to increase capture area, with close, regular setal spacing. This is indeed the morphology observed in *Tamisiocaris*. Furthermore, the mesh dimensions can be used to predict the size of the prey caught by *Tamisiocaris*. Spacing of the auxiliary spines in *T. borealis* suggests that it could capture food items suspended in the water column down to 0.49 mm, whereas linear regression from extant suspension feeders (Fig. 4) predicts a slightly larger minimum particle size of 0.70 mm. Known mesozooplankton, from small carbonaceous fossil assemblages from the Cambrian Series 2 (refs 9, 10), contain isolated feeding appendages from crustaceans,

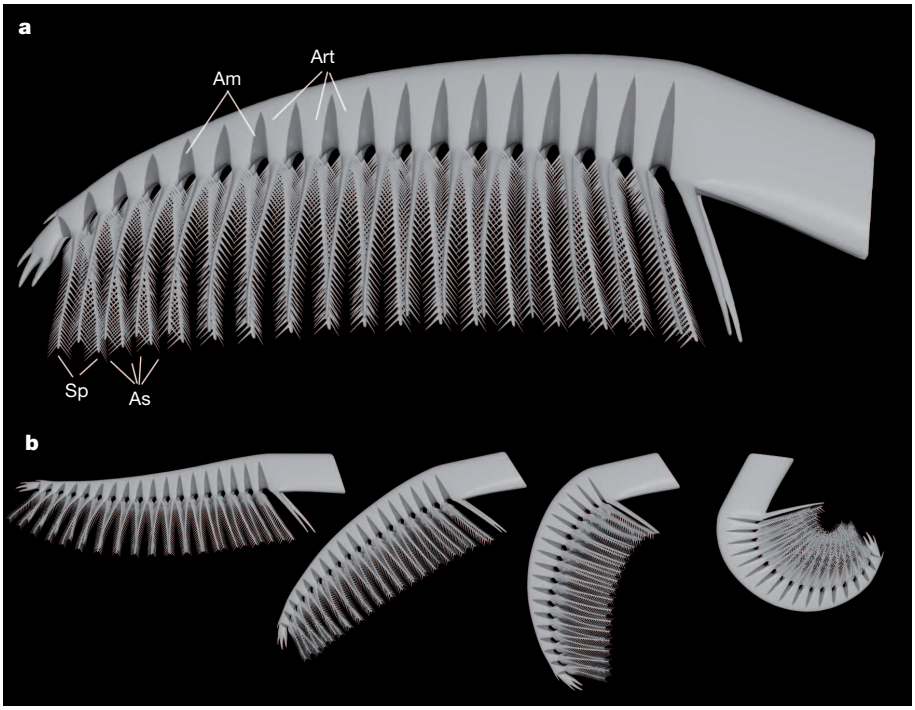


Figure 2 | A digital reconstruction of *Tamisiocaris*. **a**, Single appendage indicating the articulating membranes (Am), articles (Art), spines (Sp) and auxiliary spines (As). **b**, Possible sequence of movement of the frontal appendage of *Tamisiocaris*. See also Supplementary Videos 1 and 2.

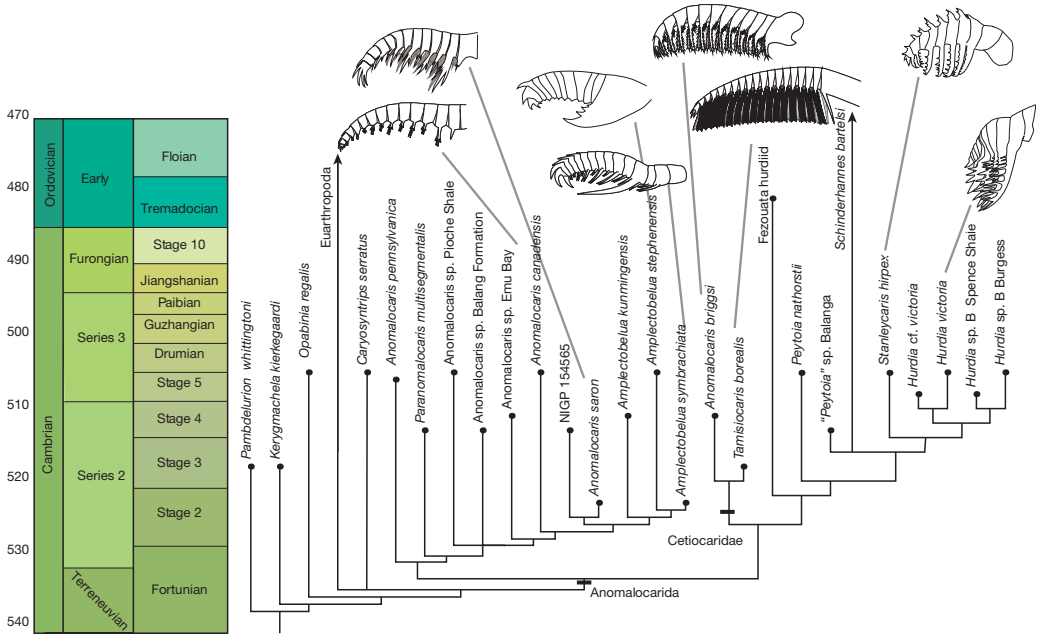


Figure 3 | Phylogeny of anomalocarids. Strict consensus of 91 trees derived from an analysis of 31 taxa and 54 characters using parsimony in PAUP* 4.0 b10. *T. borealis* forms a clade with *A. briggsi*, here named Cetiocaridae. Numerical ages (left) are given as millions of years ago.

including putative copepods. On the basis of comparisons with mandibles of modern counterparts[10], the largest known specimens reached diameters of 1.5–2.7 mm. We suggest that feeding was accomplished by alternate sweeping of the appendages, with entrapped prey being sucked[19] up by the oral cone (Supplementary Videos 1 and 2).

In the context of the phylogenetic analysis presented here (Fig. 3), we see that anomalocarid clades evolved distinct frontal appendage morphologies and feeding strategies. Primitive forms such as *A. canadensis* had raptorial appendages with stout, trident-like spines, well-suited to impaling large, free-swimming or epifaunal prey[3] (Extended Data Fig. 6a, b). *Amplectobelua* had pincer-like appendages[20] (Extended Data Fig. 6c, d) that would have been effective in seizing and tearing apart relatively large, slow-moving animals. In hurdiids, the appendages bear opposing pairs of spines, which may have functioned as jaws or in sediment sifting[15] (Extended Data Fig. 6e, f). Finally, cetiocaridid frontal appendages are specialized as sweep nets (Extended Data Fig. 6g, h). This extraordinary range of appendage morphologies shows that, far from being a failed experiment, anomalocarids staged a major adaptive radiation during the Cambrian explosion, evolving to fill a range of niches as nektonic predators, much like the later radiations of vertebrates and cephalopods, including suspension feeders[21,22].

The existence of suspension feeding in anomalocarids also has implications for the structure of Early Cambrian pelagic food webs (Extended Data Fig. 7). It had been assumed that a diverse planktonic fauna and suspension-feeding animals did not evolve until the Late Cambrian[23] and thus the complexity of the pelagic food web evolved in a delayed, piecemeal fashion. However, the discovery of large suspension feeders in the Early Cambrian suggests a well-developed pelagic biota supported by high primary productivity and abundant mesozooplankton, because large animals can only exploit small prey when they exist at high densities. Whales, whale sharks and basking sharks exploit highly productive areas such as upwelling zones and seasonal plankton blooms at high latitudes[24]. This general observation holds for all microphagous suspension feeders ranging from cladocerans, to anchovies, to red salmon,

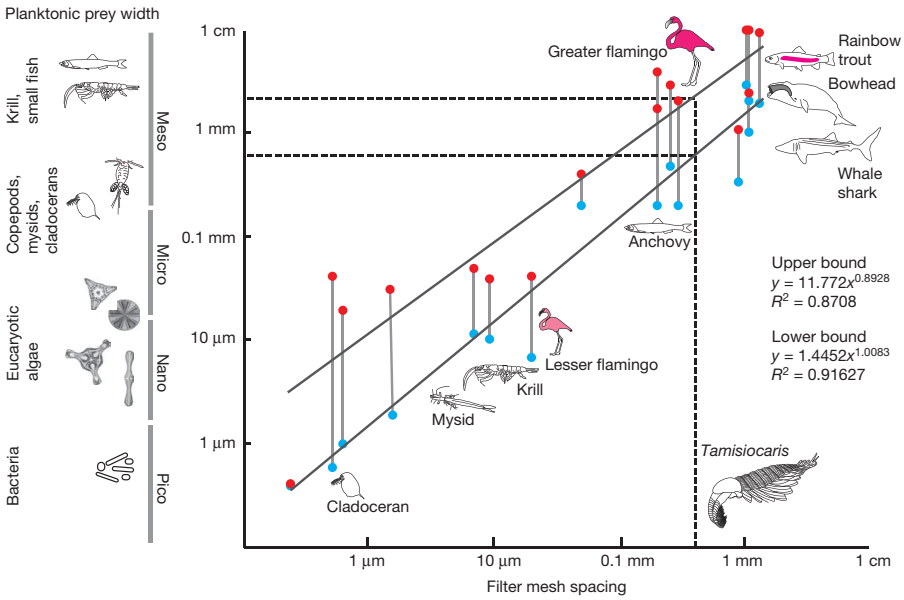


Figure 4 | Diagram depicting the relationship between suspension mesh size and the food items consumed by suspension feeders. Blue and red dots are minimum and maximum food particle size respectively recorded for a given taxon. *Tamisiocaris* is indicated by the dotted line based on the average mesh width of 0.49 mm. The diagram is collated from a range of modern suspension feeders (see Methods).

to blue whales: a high density of food particles is required to sustain an actively swimming suspension feeder.

Other evidence for high primary productivity in the Cambrian, such as vast deposits of phosphorites and increased terrestrial nutrient flux[12,13,25], imply that high productivity may have been a global phenomenon in this period. Furthermore, the Cambrian also witnessed a radiation of spiny acritarchs, which are thought to have lived as microscopic phytoplankton, replacing larger Neoproterozoic benthic forms[7,8]. Complex minute crustacean feeding appendages also occur in Lower and Middle–Upper Cambrian rocks[9,10], demonstrating the presence of diverse mesozooplankton, preying on phytoplankton. Abundant vetulicolians in Sirius Passet[26] (with hundreds of specimens collected on recent expeditions) may also have been suspension feeding upon phytoplankton (Extended Data Fig. 7). One tier up, *Tamisiocaris* would have preyed upon the mesozooplankton, as would the common nektonic arthropod *Isoxys volucris*[27]. Other pelagic predators known from Lagerstätten elsewhere would also have fed on mesozooplankton, including ctenophores, cnidarians, chaetognaths[11] and pelagic arthropods[28] (Extended Data Fig. 7). The Cambrian pelagic food web was therefore highly complex[28,29], containing multiple trophic levels, including pelagic predators[11] and multiple tiers of suspension feeders. This underscores the remarkable speed with which a modern food chain was assembled during the Cambrian explosion.

Finally, the discovery of a suspension-feeding anomalocarid has implications for debates concerning the predictability of evolution, or lack thereof. One view holds that evolution is ultimately unpredictable[30]. The notable convergence between *Tamisiocaris* and extant suspension feeders, however, suggests that although different groups occupy ecological niches at different times, the number of viable niches and viable strategies for exploiting them are limited. Furthermore, the derivation of the suspension-feeding *Tamisiocaris* from a large apex predator parallels the evolution of suspension-feeding pachycormid fish[1,21], sharks and whales[2]. In each case, suspension feeders evolved from nektonic macropredators. This suggests that evolution is canalized not only in terms of outcomes, but in terms of trajectories. The result is that independent evolutionary experiments by animals as different as anomalocarids, fish and whales have converged on broadly similar outcomes.

METHODS SUMMARY

Specimens were collected in the field and photographed in the laboratory, coated or uncoated and submerged in water. A digital reconstruction of the *Tamisiocaris* feeding appendage was made to infer the range of motions. The suspension mesh diameter and prey width were collected from the literature on extant suspension feeders to depict the linear relationship between these (see Methods). A cladistic analysis containing 31 taxa and 51 characters was collated and analysed in PAUP* 4.0 b10 and TNT (see Supplementary Information).

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

Received 28 November 2013; accepted 9 January 2014.

1. Friedman, M. *et al.* 100-Million-year dynasty of giant planktivorous bony fishes in the Mesozoic seas. *Science* **327,** 990–993 (2010).
2. Marx, F. G. & Uhen, M. D. Climate, critters, and cetaceans: Cenozoic drivers of the evolution of modern whales. *Science* **327,** 993–996 (2010).
3. Whittington, H. B. & Briggs, D. E. G. The largest Cambrian animal, *Anomalocaris*, Burgess Shale, British Columbia. *Phil. Trans. R. Soc. B* **309,** 569–609 (1985).
4. Paterson, J. R. *et al.* Acute vision in the giant Cambrian predator *Anomalocaris* and the origin of compound eyes. *Nature* **480,** 237–240 (2011).
5. Daley, A. C., Paterson, J. R., Edgecombe, G. D., García-Bellido, D. C. & Jago, J. B. New anatomical information on *Anomalocaris* from the Cambrian Emu Bay Shale of South Australia and a reassessment of its inferred predatory habits. *Palaeontology* **56,** 971–990 (2013).
6. Daley, A. C. & Peel, J. S. A possible anomalocaridid from the Cambrian Sirius Passet Lagerstatte, North Greenland. *J. Paleont.* **84,** 352–355 (2010).
7. Butterfield, N. J. Plankton ecology and the Proterozoic-Phanerozoic transition. *Paleobiology* **23,** 247–262 (1997).
8. Vidal, G. & Knoll, A. H. Radiations and extinctions of plankton in the late Proterozoic and early Cambrian. *Nature* **297,** 57–60 (1982).
9. Harvey, T. H. P. & Butterfield, N. J. Sophisticated particle-feeding in a large Early Cambrian crustacean. *Nature* **452,** 868–871 (2008).
10. Harvey, T. H. P., Vélez, M. I. & Butterfield, N. J. Exceptionally preserved crustaceans from western Canada reveal a cryptic Cambrian radiation. *Proc. Natl Acad. Sci. USA* **109,** 1589–1594 (2012).
11. Vannier, J., Steiner, M., Renvoisé, E., Hu, S.-X. & Casanova, J.-P. Early Cambrian origin of modern food webs: evidence from predator arrow worms. *Proc. R. Soc. Lond. B* **274,** 627–633 (2007).
12. Brasier, M. Nutrient flux and the evolutionary explosion across the Precambrian-Cambrian boundary interval. *Hist. Biol.* **5,** 85–93 (1991).
13. Peters, S. E. & Gaines, R. R. Formation of the 'Great Unconformity' as a trigger for the Cambrian explosion. *Nature* **484,** 363–366 (2012).
14. Briggs, D. E. G. *Anomalocaris*, the largest known Cambrian arthropod. *Palaeontology* **22,** 631–664 (1979).
15. Daley, A. C. & Budd, G. E. New anomalocaridid appendages from the Burgess Shale, Canada. *Palaeontology* **53,** 721–738 (2010).
16. Jørgensen, C. B. *Biology of Suspension Feeding* (Pergamon, 1966).
17. Pivorunas, A. The feeding mechanisms of baleen whales. *Am. Sci.* **67,** 432–440 (1979).
18. Nemoto, T. in *Marine Food Chains* (ed. Steele, J. H.) 241–252 (Univ. California Press, 1970).
19. Daley, A. C. & Bergström, J. The oral cone of *Anomalocaris* is not a classic "peytoia". *Naturwissenschaften* **99,** 501–504 (2012).
20. Hou, X.-G., Bergström, J. & Ahlberg, P. *Anomalocaris* and other large animals in the lower Cambrian Chengjiang fauna of southwest China. *GFF* **117,** 163–183 (1995).
21. Friedman, M. Parallel evolutionary trajectories underlie the origin of giant suspension-feeding whales and bony fishes. *Proc. R. Soc. B* **279,** 944–951 (2012).
22. Kruta, I., Landman, N., Rouget, I., Cecca, F. & Tafforeau, P. The role of ammonites in the Mesozoic marine food web revealed by jaw preservation. *Science* **331,** 70–72 (2011).
23. Signor, P. W. & Vermeij, G. J. The plankton and the benthos: origins and early history of an evolving relationship. *Paleobiology* **20,** 297–319 (1994).
24. Tynan, C. T. Ecological importance of the Southern Boundary of the Antarctic Circumpolar Current. *Nature* **392,** 708–710 (1998).
25. Cook, P. J. & Shergold, J. H. Phosphorus, phosphorites and skeletal evolution at the Precambrian–Cambrian boundary. *Nature* **308,** 231–236 (1984).
26. Vinther, J., Smith, M. P. & Harper, D. A. T. Vetulicolians from the Lower Cambrian Sirius Passet Lagerstätte, North Greenland, and the polarity of morphological characters in basal deuterostomes. *Palaeontology* **54,** 711–719 (2011).
27. Stein, M., Peel, J. S., Siveter, D. J. & Williams, M. *Isoxys* (Arthropoda) with preserved soft anatomy from the Sirius Passet Lagerstatte, lower Cambrian of North Greenland. *Lethaia* **43,** 258–265 (2010).
28. Vannier, J., García-Bellido, D. C., Hu, S.-X. & Chen, A.-L. Arthropod visual predators in the early pelagic ecosystem: evidence from the Burgess Shale and Chengjiang biotas. *Proc. R. Soc. B* **276,** 2567–2574 (2009).
29. Dunne, J. A., Williams, R. J., Martinez, N. D., Wood, R. A. & Erwin, D. H. Compilation and network analyses of Cambrian food webs *PLoS Biol.* **6,** e102 (2008).
30. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton & Co., 1989).

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

Acknowledgements Our expeditions to North Greenland were financed by Geocenter Denmark, the Agouron Institute and the Carlsberg Foundation. We are grateful for discussions with members of the Bristol Palaeobiology group, E. Sperling, C. Hull, M. Matz and M. Friedman. M.S. was supported by the Carlsberg Foundation. We thank POLOG for logistic support. A. T. Nielsen and M. P. Smith assisted in the field in 2009. S. Powell assisted with figures. S. L. Jakobsen and A. T. Nielsen facilitated work and curation of the collected material at the Statens Naturhistoriske Museum, Copenhagen.

Author Contributions J.V., M.S. and N.R.L. designed, analysed and performed research. D.A.T.H. obtained funding for the fieldwork. J.V., M.S., N.R.L. and D.A.T.H. wrote and discussed 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 J.V. (jakob.vinther@bristol.ac.uk).

LETTER

doi:10.1038/nature12980

A primitive placoderm sheds light on the origin of the jawed vertebrate face

Vincent Dupret[1], Sophie Sanchez[1,2], Daniel Goujet[3], Paul Tafforeau[2] & Per E. Ahlberg[1]

Extant vertebrates form two clades, the jawless Cyclostomata (lampreys and hagfishes) and the jawed Gnathostomata (all other vertebrates), with contrasting facial architectures[1,2]. These arise during development from just a few key differences in the growth patterns of the cranial primordia: notably, the nasal sacs and hypophysis originate from a single placode in cyclostomes but from separate placodes in gnathostomes, and infraoptic ectomesenchyme migrates forward either side of the single placode in cyclostomes but between the placodes in gnathostomes[3–8]. Fossil stem gnathostomes preserve cranial anatomies rich in landmarks that provide proxies for developmental processes and allow the transition from jawless to jawed vertebrates to be broken down into evolutionary steps[7,9–12]. Here we use propagation phase contrast synchrotron microtomography to image the cranial anatomy of the primitive placoderm (jawed stem gnathostome) *Romundina*[13], and show that it combines jawed vertebrate architecture with cranial and cerebral proportions resembling those of cyclostomes and the galeaspid (jawless stem gnathostome) *Shuyu*[11]. This combination seems to be primitive for jawed vertebrates, and suggests a decoupling between ectomesenchymal growth trajectory, ectomesenchymal proliferation, and cerebral shape change during the origin of gnathostomes.

One of the most distinctive features of cyclostome facial architecture is the median nasohypophysial duct, which forms when the posthypophyseal processes grow forward on either side of the nasohypophysial placode to create the hood (lamprey) or oronasal membrane plus tentacles (hagfish)[2,4–7]. In gnathostomes, by contrast, these processes converge in the midline between the hypophysial and nasal placodes to create the trabeculae, while the maxillary process of the mandibular arch grows forward laterally to form the upper jaw[1,4–7,14]. Cyclostomes have very short forebrains and nasal sacs that lie well posterior to the upper lip of the mouth, whereas gnathostomes have longer forebrains and nasal sacs located above the jaw margin[4–7]. The cyclostome hypophysis projects ventrally into the nasohypophysial duct, whereas the gnathostome hypophysis projects posteroventrally towards the palate; the distance between hypophysis and nasal sacs is thus much greater in gnathostomes than cyclostomes.

Comparative analysis of extant cyclostomes and gnathostomes alone does not allow us to polarize this transformation or break it down into steps. However, the lower part of the gnathostome stem group comprises fossil jawless vertebrates with nasohypophysial ducts[15], showing that the cyclostome pattern precedes the gnathostome pattern and indicating that fossils can potentially illuminate the transformation itself (Fig. 1 and Extended Data Figs 1 and 2). The 430-million-year-old galeaspid (jawless stem gnathostome) *Shuyu* seems to show an early transitional condition: nasal sacs and hypophysis are separated by a small trabecular process, but open into a common nasohypophysial duct[11]. This suggests that the posthypophysial processes of *Shuyu* followed a lateral growth path as in cyclostomes, but that nasal and hypophysial placodes were separate, and that a small component of the premandibular ectomesenchyme grew into the midline above the hypophysial placode to form the trabecular process (Supplementary Information).

Placoderms (extinct armoured jawed fishes) form the paraphyletic upper part of the gnathostome stem group[9,10,12]. They may thus include the most primitive examples of jawed vertebrate morphology. *Romundina stellina* Ørvig, 1975 (ref. 13) (Fig. 2 and Extended Data Figs 3 and 4), from the earliest Devonian period (410–419 million years ago) of Canada, is one of a small number of placoderms that differ from all other jawed vertebrates by having nasal capsules positioned between the eyes, well behind the upper lip. These 'posterior-nosed' forms, which also include *Brindabellaspis* and antiarchs, are consistently recovered among the phylogenetically deepest placoderms[9,10,12,15,16] (Fig. 1, Supplementary Information and Extended Data Figs 1 and 2). The position of the capsules suggests that they had short forebrains, but their endocranial anatomy has remained largely unknown[15,17–19]. The only properly documented cranial cavity of a posterior-nosed form is that of *Brindabellaspis*, reconstructed from two acid-prepared braincases[18]. *Brindabellaspis* has some puzzling features, such as nasal sacs located within the orbits[18], which make it difficult to assess whether its endocranial anatomy is representative for posterior-nose placoderms in general.

Here we present the braincase of *Romundina*, reconstructed from a propagation phase contrast synchrotron microtomography (PPC-SRμCT) scan of specimen MNHN CPW1 (Fig. 2a, b and Supplementary Video 1). The specimen is complete apart from an oblique ventral breakage and the absence of the rostronasal capsule. This capsule, comprising the nasal capsules and their dermal bone cover, was separated from the braincase by an encircling optic fissure (Fig. 2a, b); it is known

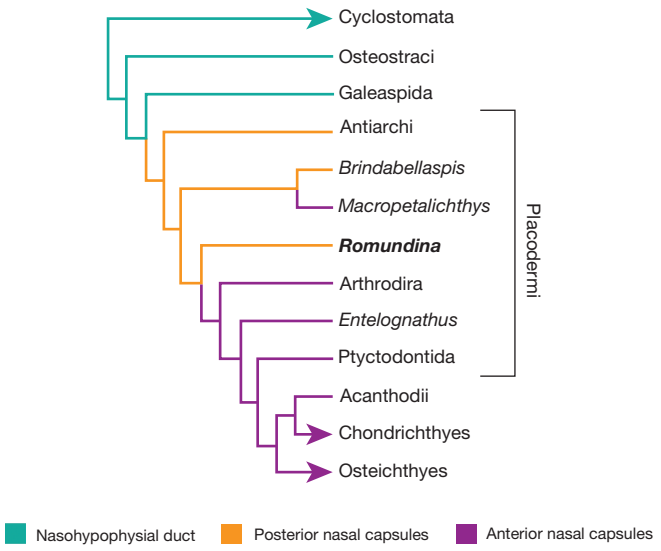


Figure 1 | Phylogenetic position of *Romundina*. Simplified vertebrate phylogeny, based on the character matrix reported previously[12] with the addition of *Romundina*, showing the distribution of different nasal architectures. Arrowheads indicate extant groups. For full analysis see Supplementary Information.

[1]Uppsala University, Department of Organismal Biology, Subdepartment of Evolution and Development, Norbyvägen 18A, SE-752 36, Uppsala, Sweden. [2]European Synchrotron Radiation Facility, 6 rue Jules Horowitz, 38043 Grenoble Cedex, France. [3]Muséum national d'Histoire naturelle, UMR 7207 CR2P CNRS/MNHN/UPMC, 8 rue Buffon, CP 38,75231 Paris Cedex 05, France.

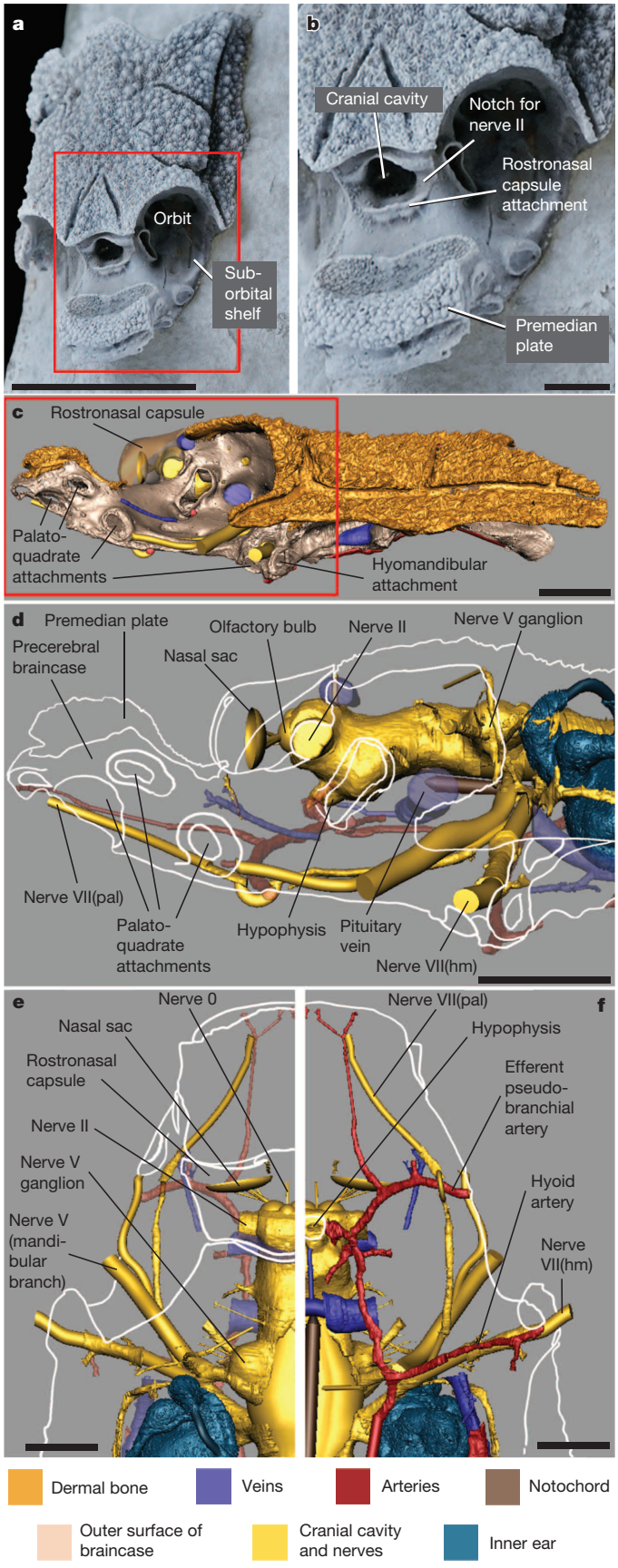


Figure 2 | Endocranial anatomy of *Romundina*. **a**, Photograph of specimen MNHN CPW1 in left anterodorsolateral view. The red box indicates the area enlarged in **b**. **b**, Orbital region and snout in close-up view. **c**, Braincase and skull roof in left lateral view, modelled from PPC-SRμCT scan. Red box indicates region shown in **d**–**f**. **d**–**f**, Cranial cavity and associated structures in lateral (**d**), dorsal (**e**) and ventral (**f**) views. Outline of braincase shown in white. Rostronasal capsule, nasal sacs and olfactory bulbs are reconstructed from previously reported data[13] (see also Supplementary Information). **e** and **f** both extend slightly across the midline so that midline structures such as nerve 0 and the hypophysial fossa are shown complete. Scale bars: **a**, 10 mm; **b**–**d**, 2 mm; **e**, **f**, 1 mm. All images are original.

from other specimens of *Romundina*[13,20] (Extended Data Fig. 4), allowing us to reconstruct its position and internal spaces (Figs 2c–f and 3, Extended Data Fig. 3c–f and Supplementary Video 2). The posterior part of the braincase is broadly similar to those of other early placoderms[15,17,21–23] and will be the subject of a subsequent publication. We focus here on the anterior region, from the canal for the hyomandibular branch of the facial nerve (VII(hm)) to the tip of the snout (Fig. 2d–f and Extended Data Fig. 3d–f), which differs radically from the corresponding region of an extant gnathostome.

The telencephalic portion of the cranial cavity is extremely short (Figs 2c–e, 3). It is bounded anteriorly by the lamina cribrosa of the rostronasal capsule (Extended Data Fig. 4 and ref. 13), which carries on its posterior face a central pair of foramina for the terminal nerve (nerve 0) and two lateral concavities for the olfactory bulbs. Posterior to the optic foramen (nerve II), which straddles the optic fissure, the diencephalic portion of the cranial cavity passes down to the anteroventrally directed hypophysial recess. Large recesses for the nerve V ganglia mark the level of the pons. Nasal–hypophysial distance is much smaller than in an extant gnathostome (Fig. 3). Similar proportions characterize the cranial cavities of the posterior-nosed placoderms *Brindabellaspis*[18] and *Jagorina*[23], the galeaspid *Shuyu*[11], and the brains of extant cyclostomes[15] (Supplementary Information). We conclude that cerebral proportions remained substantially unchanged through the transition from jawless to jawed cranial architecture, and that the characteristic elongate telencephalon and posteroventrally oriented hypophysis of extant gnathostomes evolved later (Fig. 3).

The braincase of *Romundina* includes a large precerebral region and broad suborbital shelves that terminate caudally at the canal for the hyomandibular branch of the facial nerve (nerve VII(hm)) (Fig. 2 and Extended Data Fig. 3). The suborbital shelves carry a series of articulations for the palatoquadrate (known from detached examples[13]) from the anterior end of the precerebral region to immediately in front of the VII(hm) nerve foramen. The spatial relationship of this region to the palatoquadrate and rostronasal capsule, and the fissure separating it from the latter, indicate that the precerebral braincase and suborbital shelves are formed by the trabeculae. In extant gnathostomes, the posterior part of the trabeculae consists of infraoptic ectomesenchyme, whereas the anterior part seems to contain varying components of supra- and infraoptic ectomesenchyme, depending on taxon[24,25]; the nasal capsules are always composed of supraoptic ectomesenchyme. In *Romundina*, the sharp anterior demarcation and bounding fissure of the rostronasal capsule suggest that it formed the anteriormost part of the supraoptic ectomesenchyme and that the whole trabecular region was composed of infraoptic material (Supplementary Information and Figs 3 and 4).

The braincase of *Brindabellaspis* resembles that of *Romundina* in having a long precerebral region[18]. The partly preserved braincase of the antiarch *Minicrania* and attachment scars on the dermal skull bones of other antiarchs indicate a similar configuration[19,26]. The rostronasal capsule was small and fissure bounded in antiarchs (as indicated by the dermal component), whereas in *Brindabellaspis* the capsule was fused to the braincase and the nasal sacs lay in the anterior corners of the orbits[17,18]. We conclude that the anatomical interpretation and inferred ectomesenchymal map of *Romundina* are applicable to posterior-nosed placoderms in general.

Although *Romundina* is an unambiguous jawed vertebrate, the proportions of its brain closely resemble those of jawless vertebrates. Furthermore, its facial morphology suggests that the region formed from infraoptic ectomesenchyme was proportionately much larger than in an extant

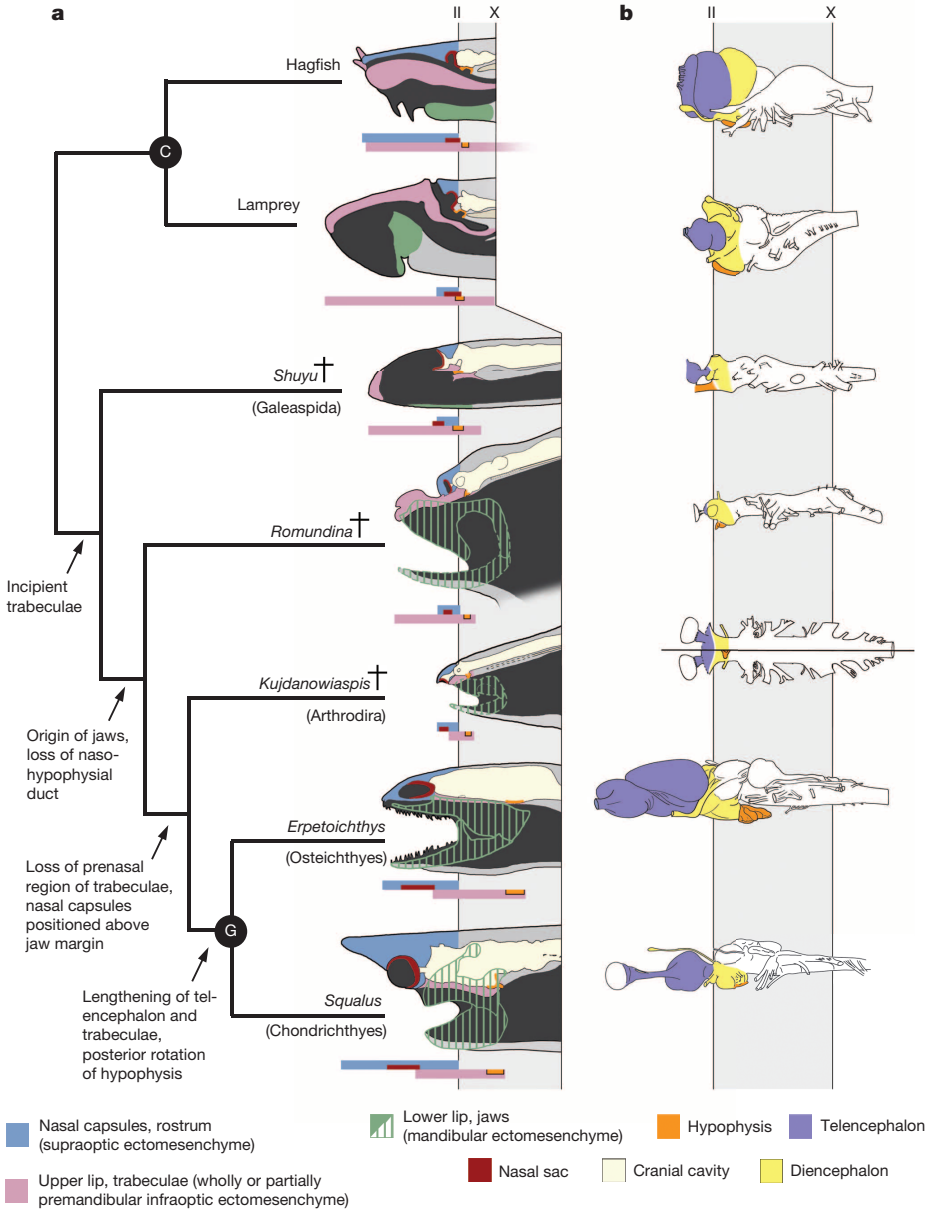


Figure 3 | Transformation of the vertebrate face and brain. **a**, Phylogeny of representative vertebrates, with diagrammatic sagittal sections of their heads. Heads are aligned on the exits of the optic nerve (II) and vagus nerve (X); cyclostomes are shown at a smaller scale because of their short brains and large precerebral regions. Bar diagram beneath each head shows the longitudinal position and extent of cranial components. Posterior boundary of supraoptic ectomesenchyme is positioned arbitrarily at the level of nerve II. Hagfish and lamprey are modified from ref. 7, *Shuyu* is based on data from ref. 11, with permission, all other images are original. C, cyclostome crown group node; G, gnathostome crown group node. Arrows indicate important evolutionary steps. **b**, Brains/cranial cavities of the same taxa, aligned on the optic nerve (II) and vagus nerve (X), all in lateral view except *Kujdanowiaspis*, which is shown in composite dorsal (top)/ventral (bottom) view because no lateral view is available. Diagrams of brains of hagfish, lamprey, *Erpetoichthys* are modified with permission from ref. 29; brain of *Squalus* is modified with permission from ref. 30; cranial cavity of *Shuyu* is modified with permission from ref. 11; cranial cavity of *Kujdanowiaspis* is modified from ref. 23.

gnathostome and projected anteriorly beyond the supraoptically derived rostronasal capsule (Figs 3 and 4). This resembles the condition in extant cyclostomes, where infraoptic material forms the bulk of the upper lip or oronasal membrane (Fig. 3). The phylogenetic position of *Romundina* and the similar morphology of other posterior-nosed placoderms suggest that these characteristics are primitive for jawed stem gnathostomes. Immediately crownward to *Romundina*, our phylogenetic analysis recovers the clade Arthrodira (Figs 1 and 3, Supplementary Information and Extended Data Figs 1 and 2). Primitive arthrodires such as *Kujdanowiaspis*[21,23] and *Dicksonosteus*[22] have terminal rostronasal capsules, but their forebrains are still very short. Comparison with *Romundina* (Fig. 3) indicates that this configuration was produced by shortening the trabecular region anteriorly. In the gnathostome crown group the telencephalon, trabecular region and supraoptic ectomesenchymal territory have all been lengthened substantially, but their relative positions and proportions remain similar to those of arthrodires (Fig. 3).

The transformation from cyclostome to gnathostome facial architecture has been described as a heterotopic phenomenon, involving changes in ectomesenchymal growth trajectories and ectoderm–ectomesenchyme interactions[6]. The data from stem gnathostomes support this interpretation but add evidence for the intermediate steps of the transformation and the decoupling of different processes (Fig. 3). Separate but closely spaced nasal and hypophysial placodes, associated with a short telencephalon, were established before the origin of jaws and retained an unchanged geometry across the transition from agnathans to jawed vertebrates. This shows that the heterotopic shift in ectomesenchymal growth trajectory, which replaced the nasohypophysial duct with a solid trabecular region, was not driven by a lengthening of the telencephalon, leading to wider separation of the nasal and hypophysial placodes. On the contrary, we suggest that the creation of a trabecular region contributed to making this lengthening possible. By forming a horizontal shelf between the nasal sacs and hypophysis, the trabeculae also became a skeletal floor for the telencephalic part of the cranial cavity, which could support and protect a long telencephalon. In cyclostomes the telencephalon has no ventral support[7]. Consonant with this hypothesis, telencephalic lengthening occurred independently at least twice among placoderms (in *Macropetalichthys* and pholidosteid arthrodires[23]) as well as in crown gnathostomes. Even within the crown group, some early members such as the sarcopterygian *Tungsenia*[27] retain a relatively short telencephalon, although the nasal–hypophysial distance is greater than in *Romundina* or primitive arthrodires.

When the trabecular region first evolved, its dimensions were similar to an agnathan upper lip (Figs 3 and 4), suggesting that the change of growth trajectory for the premandibular ectomesenchyme was not

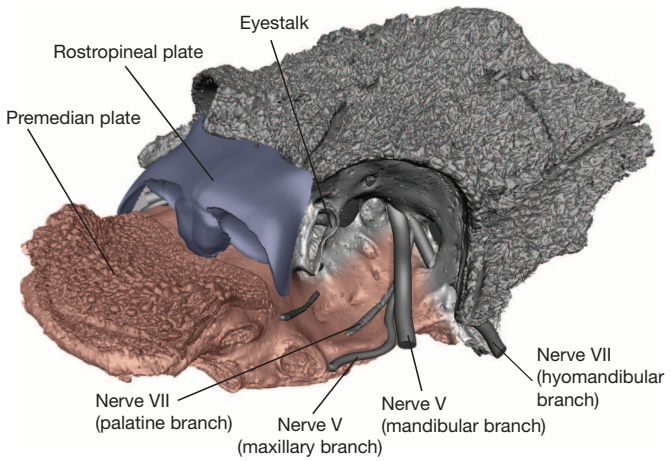


Figure 4 | Inferred neural crest composition of *Romundina*. PPC-SRμCT model of *Romundina* specimen MNHN CPW1 in left anterodorsolateral view, with reconstructed rostronasal capsule, showing inferred extent of premandibular infraoptic ectomesenchyme (pink) and supraoptic ectomesenchyme (blue).

initially associated with a reduction in cell proliferation to produce a smaller structure; relative size reduction of the trabeculae, which produced the spatial configuration characteristic of extant gnathostomes, came later. This implies that trajectories and proliferation patterns must have been regulated separately during the morphological transformation from jawless to jawed facial architecture. We predict that separate regulatory networks for these attributes are also present in extant vertebrates.

METHODS SUMMARY

Specimens of *Romundina stellina* were partly freed from the surrounding limestone matrix by dilute formic acid. One specimen was imaged with PPC-SRμCT at beamline ID19 of the European Synchrotron Radiation Facility (see Supplementary Information and ref. 28 for technical specifications). Mimics v.12.3 and v.13.1 (Materialise) were used for three-dimensional modelling. Pictures and animations were rendered in Maya 2012 (Autodesk).

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

Received 30 July; accepted 30 December 2013.
Published online 12 February 2014.

1. Kuratani, S. & Horigome, N. Development of peripheral nerves in a cat shark, *Scyliorhinus torazame*, with special reference to rombomeres, cephalic mesoderm, and distribution patterns of crest cells. *Zoolog. Sci.* **17,** 893–909 (2000).
2. Kuratani, S., Horigome, N. & Hirano, S. Developmental morphology of the head mesoderm and re-evaluation of segmental theories of the vertebrate head: evidence from embryos of an agnathan vertebrate, *Lampetra japonica*. *Dev. Biol.* **210,** 381–400 (1999).
3. Kuratani, S. Evolution of the vertebrate jaw: comparative embryology and molecular developmental biology reveal the factors behind evolutionary novelty. *J. Anat.* **205,** 335–347 (2004).
4. Kuratani, S. Developmental studies of the lamprey and hierarchical evolutionary steps towards the acquisition of the jaw. *J. Anat.* **207,** 489–499 (2005).
5. Kuratani, S. Evolution of the vertebrate jaw from developmental perspectives. *Evol. Dev.* **14,** 76–92 (2012).
6. Kuratani, S., Nobusada, Y., Horigome, N. & Shigetani, Y. Embryology of the lamprey and evolution of the vertebrate jaw: insights from molecular and developmental perspectives. *Phil. Trans. R. Soc. Lond. B* **356,** 1615–1632 (2001).
7. Oisi, Y., Ota, K. G., Kuraku, S., Fujimoto, S. & Kuratani, S. Craniofacial development of hagfishes and the evolution of vertebrates. *Nature* **493,** 175–180 (2013).
8. Ota, K. G. & Kuratani, S. Cyclostome embryology and early evolutionary history of vertebrates. *Integr. Comp. Biol.* **47,** 329–337 (2007).
9. Brazeau, M. D. The braincase and jaws of a Devonian 'acanthodian' and modern gnathostome origins. *Nature* **457,** 305–308 (2009).
10. Davis, S. P., Finarelli, J. A. & Coates, M. I. Acanthodes and shark-like conditions in the last common ancestor of modern gnathostomes. *Nature* **486,** 247–250 (2012).
11. Gai, Z., Donoghue, M. J., Zhu, M., Janvier, P. & Stampanoni, M. Fossil jawless fish from China foreshadows early jawed vertebrate anatomy. *Nature* **476,** 324–327 (2011).
12. Zhu, M. *et al.* A Silurian placoderm with osteichthyan-like marginal jaw bones. *Nature* (2013).
13. Ørvig, T. in *Problèmes Actuels de Paléontologie: Evolution des Vertébrés* (ed. Lehman, J. P.) 41–72 (Colloques Internationaux Centre National de la Recherche Scientifique, 1975).
14. Couly, G. F., Coltey, P. M. & Le Douarin, N. M. The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. *Development* **117,** 409–429 (1993).
15. Janvier, P. *Early Vertebrates* Vol. 1 (Clarendon, 1996).
16. Goujet, D. & Young, G. C. Interrelationships of placoderms revisited. *Geobios.* **28,** 89–95 (1995).
17. Denison, R. H. *Placodermi* (Gustav Fischer, 1978).
18. Young, G. C. A new Early Devonian placoderm from New South Wales, Australia, with a discussion of placoderm phylogeny. *Palaeontographica* **167,** 10–76 (1980).
19. Young, G. C. Reconstruction of the jaws and braincase in the Devonian placoderm fish *Bothriolepis*. *Palaeontology* **27,** 635–661 (1984).
20. Goujet, D. & Young, G. C. in *Recent Advances in the Origin and Early Radiation of Vertebrates* (eds Arratia, G., Wilson, M. V. H. & Cloutier, R.) 109–126 (Dr. Friedlich Pfeil, 2004).
21. Dupret, V. Revision of the genus *Kujdanowiaspis* Stensiö, 1942 (Placodermi, Arthrodira, "Actinolepida") from the Lower Devonian of Podolia (Ukraine). *Geodiversitas* **32,** 5–63 (2010).
22. Goujet, D. *Les Poissons Placodermes du Spitsberg: Arthrodires Dolichothoraci de la Formation de Wood Bay* (*Devonien Inferieur*) (CNRS, 1984).
23. Stensiö, E. in *Traité de Paléontologie* Vol. 4 (ed. Piveteau, J.) 71–693 (Masson, 1969).
24. Osumi-Yamashita, N. *et al.* Cranial anomaly of homozygous *rSey* rat is associated with a defect in the migration pathway of midbrain crest cells. *Dev. Growth Differ.* **39,** 53–67 (1997).
25. Wada, N., Nohno, T. & Kuratani, S. Dual origin of the prechordal cranium in the chicken embryo. *Dev. Biol.* **356,** 529–540 (2011).
26. Zhu, M. & Janvier, P. A small antiarch, *Minicrania lirouyii* gen. et sp. nov from the early Devonian of Qujing, Yunnan (China), with remarks on the antiarch phylogeny. *J. Vertebr. Paleontol.* **16,** 1–15 (1996).
27. Lu, J. *et al.* The earliest known stem-tetrapod from the Lower Devonian of China. *Nature Commun.* **3,** 1160 (2012).
28. Dupret, V., Sanchez, S., Goujet, D., Tafforeau, P. & Ahlberg, P. Bone vascularization and growth in placoderms (Vertebrata): the example of the premedian plate of *Romundina stellina* Ørvig, 1975. *C. R. Palevol* **9,** 369–375 (2010).
29. Collin, S. P. in *Fish Physiology* (eds McKenzie, D. J., Farell, A. P. & Brauner, C. J.) 121–179 (Elsevier, 2007).
30. De Iuliis, G. & Pulerà, D. *The Dissection of Vertebrates—A Laboratory Manual* (Academic, 2007).

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

Acknowledgements We thank the European Synchrotron Radiation Facility for granting us beam time at ID19 (proposal EC-203). P.E.A., V.D. and S.S. acknowledge the support of European Research Council Advanced Investigator Grant 233111 and a Wallenberg Scholarship from the Knut and Alice Wallenberg Foundation, both awarded to P.E.A. We thank B. Ryll, M. Kundrat and H. Blom at Uppsala University for discussions. Specimen MNHN CPW 1 photographs were taken by P. Loubry (Centre National pour la Recherche Scientifique, Muséum national d'Histoire naturelle, Paris).

Author Contributions The project was conceived by V.D., P.E.A. and S.S. Specimens of *Romundina* were collected by D.G. Scanning and reconstruction of data sets were carried out by S.S. and P.T., with minor contributions to the scanning by V.D. and P.E.A. Three-dimensional modelling, rendering and animation were done by V.D., who also carried out phylogenetic analysis and anatomical interpretation with input from P.E.A. P.E.A. led the comparative developmental interpretation. P.E.A. and V.D. wrote the text. All authors critically reviewed the manuscript and approved the final draft.

Author Information Data are deposited online at http://paleo.esrf.eu. 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 V.D. (vincent.dupret@ebc.uu.se) or P.E.A. (per.ahlberg@ebc.uu.se).

LETTER

doi:10.1038/nature13149

Visual space is compressed in prefrontal cortex before eye movements

Marc Zirnsak[1,2], Nicholas A. Steinmetz[1], Behrad Noudoost[1], Kitty Z. Xu[1] & Tirin Moore[1,2]

We experience the visual world through a series of saccadic eye movements, each one shifting our gaze to bring objects of interest to the fovea for further processing. Although such movements lead to frequent and substantial displacements of the retinal image, these displacements go unnoticed. It is widely assumed that a primary mechanism underlying this apparent stability is an anticipatory shifting of visual receptive fields (RFs) from their presaccadic to their postsaccadic locations before movement onset[1]. Evidence of this predictive 'remapping' of RFs has been particularly apparent within brain structures involved in gaze control[2–4]. However, critically absent among that evidence are detailed measurements of visual RFs before movement onset. Here we show that during saccade preparation, rather than remap, RFs of neurons in a prefrontal gaze control area massively converge towards the saccadic target. We mapped the visual RFs of prefrontal neurons during stable fixation and immediately before the onset of eye movements, using multi-electrode recordings in monkeys. Following movements from an initial fixation point to a target, RFs remained stationary in retinocentric space. However, in the period immediately before movement onset, RFs shifted by as much as 18 degrees of visual angle, and converged towards the target location. This convergence resulted in a threefold increase in the proportion of RFs responding to stimuli near the target region. In addition, like in human observers[5,6], the population of prefrontal neurons grossly mislocalized presaccadic stimuli as being closer to the target. Our results show that RF shifts do not predict the retinal displacements due to saccades, but instead reflect the overriding perception of target space during eye movements.

We recorded from neurons within the frontal eye field (FEF) of monkeys (*Macaca mulatta*) using linear electrode arrays (Fig. 1a and Methods). The FEF is an area of prefrontal cortex with a known involvement in gaze control[7] and visual attention[8–11]. Previous studies have found evidence that visual RFs of FEF neurons predictively remap before saccades[1,4]. That is, this body of evidence suggests that FEF RFs shift from their presaccadic locations to their anticipated, postsaccadic locations before the onset of each saccade (Extended Data Fig. 1). However, because these studies inferred RF shifts from visual responses to stimuli presented at only a few locations, the validity of the remapping framework remains uncertain. We therefore mapped the RFs of simultaneously recorded FEF neurons with flashed (25 ms) 'probe' stimuli while monkeys performed a standard saccade task[4] (Methods and Extended Data Fig. 2). To obtain detailed measurements of RFs during the task, we used a dense array of visual probes: 10×9 positions covering an area of 36×32 degrees of visual angle (dva) (Fig. 1b). Using this arrangement, we mapped RFs during three separate periods: during fixation at each of two fixation points (FP1 and FP2) and just prior (69 ms, s.d. = 35) to a saccade from FP1 to FP2. Figure 1c shows four examples of FEF RFs mapped during fixation at FP1 and FP2, and their corresponding RF centres (RF1s and RF2s) (Methods). In each example, the change in fixation from one location to the other was accompanied by a RF displacement that was equivalent to the displacement of fixation, reflecting the retinocentric property of FEF RFs. RFs measured at FP2 also served as empirical estimates of the expected shifts due to predictive remapping. When measured while monkeys were still fixating at FP1, but preparing saccades to FP2, RFs were considerably different from those measured during stable fixation. As shown in the examples, presaccadic RF centres (PRFs) differed both from the RF1s and the RF2s (Fig. 1d). Furthermore, PRFs tended to be much closer to the saccade target (FP2), in some cases shifting from their RF1 location in a direction orthogonal (Fig. 1c, example 3) or opposite (Fig. 1c, example 4) to the saccade direction, and thus were inconsistent with remapping.

We measured the changes in RF2s and PRFs from their corresponding RF1s (Δ_{FIX} and Δ_{PRE}, respectively) for a population of 179 RFs mapped in two monkeys (Fig. 2a and Methods). As expected, for a fixed saccade vector, the average displacement of RF2s (11.2 dva, s.d. = 1.5) was approximately equal to the saccade amplitude (12 dva, s.d. = 0.2), and was independent of the RF1 distance to the saccade target (ε'_{RF1}) ($r = 0.05$, $P = 0.49$) (Methods). Without presaccadic RF modulation, we would expect PRFs to be identical to their corresponding RF1s, given that retinal stimulation is essentially identical in these two conditions. However, FEF RFs are considerably altered during saccade preparation, and are shifted by an average of 8.4 dva (s.d. = 3.6) and as much as 18 dva (Methods and Extended Data Fig. 3). This shift was greater ($P < 10^{-10}$) than the small variations in RFs measured between the fixation conditions (Extended Data Fig. 4). Furthermore, in contrast to RF2s, PRFs depended on the RF1 distance to the saccade target ($r = 0.44$, $P < 10^{-9}$), with larger PRF shifts occurring for more distant RFs. Note that this dependence should not occur if the presaccadic RFs predictively shift from RF1s to their postsaccadic RF2s. Moreover, the angular differences (ϕ') between the RF2 displacement vectors and the PRF shift vectors were not uniform as expected with remapping, but depended on the angular deviation (θ) of RF1 from the saccade vector ($r = 0.75$, $P < 10^{-10}$) (Fig. 2b). Thus, we observed substantial presaccadic shifts that were inconsistent with the remapping prediction (Extended Data Fig. 5). Furthermore, the overall pattern of RF shifts reveals how individual shifts can appear consistent with remapping at some locations in space (Methods and Extended Data Fig. 6).

To further understand the presaccadic RF modulation, we examined PRF shift vectors across the entire distribution of measured RF locations (Fig. 2c, Methods and Extended Data Fig. 7). We observed that PRFs shifted in the direction of the saccade target independent of their corresponding RF1 location. On average, PRFs were 6.1 dva closer to the saccade target than the RF1s ($P < 10^{-10}$). We further compared PRFs to the remapping prediction, using the RF2s as an empirical estimate, and found PRFs to be closer to the saccade target than expected with remapping ($P < 10^{-10}$). Thus, PRFs deviated from the remapping prediction and instead converged towards the saccade target. This pattern of results was the same when considering only single neurons, indicating that the convergence reflected the shifting of individual neuronal RFs rather than a differential gain change across multiple neurons (Methods and Extended Data Figs 8 and 9).

We next examined how presaccadic changes in RFs altered the representation of visual space by the population of recorded FEF neurons.

[1]Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA.

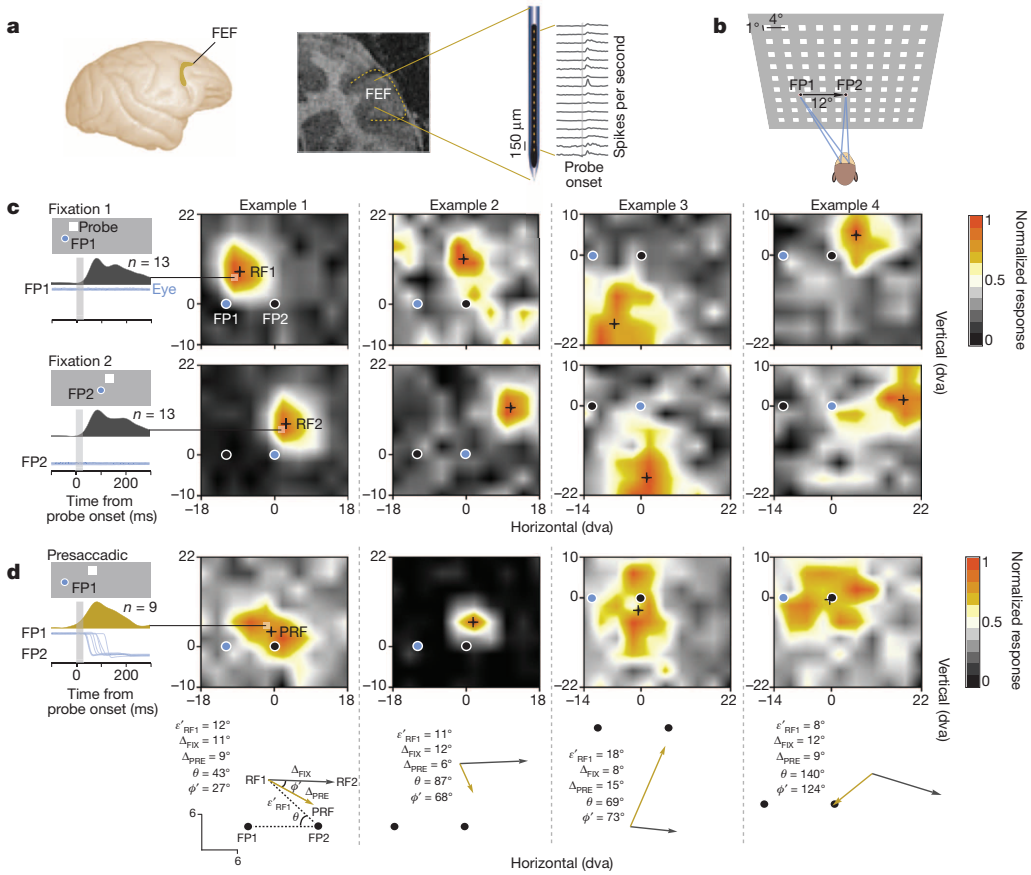


Figure 1 | Multi-electrode recordings and high-density mapping of FEF neuronal RFs. **a**, The FEF in the macaque cortex (left) and in a coronal magnetic resonance image from monkey N (middle). Right, linear array microelectrode and traces of FEF visual responses recorded simultaneously across 16 electrode contacts. **b**, FEF RFs were mapped with a 10 × 9 array (36 × 32 dva) of probe stimuli (squares) flashed during fixation at FP1 and FP2 and immediately before saccades from FP1 to FP2. **c**, Four example neuronal RF maps. Left, mean peri-stimulus response histogram for the most effective probe location during fixation at FP1 (top) and FP2 (bottom). Blue traces and circles indicate eye position and fixation location, respectively. **d**, Presaccadic RF maps of the same neurons mapped immediately before saccades from FP1 to FP2. Left, mean peri-stimulus response histogram for the most effective probe location. Probe presentation (vertical grey shading) was completed before saccadic onset on all trials. Bottom, measurement of changes in RF centres (black crosses, RF1, RF2 and PRF) during fixation (grey vectors) and saccade preparation (gold vectors).

Rather than focus solely on RF centres, however, we instead measured how saccade preparation altered the degree to which each visual probe elicited responses from the neuronal population (Methods). Figure 3a shows the percentage of 'population RFs' yielding responses across all probe locations during fixation and saccade preparation. These distributions of 'RF density' illustrate how the representation of visual space is displaced by the change in fixation from FP1 to FP2 in the population. Moreover, these RF density distributions revealed a substantial

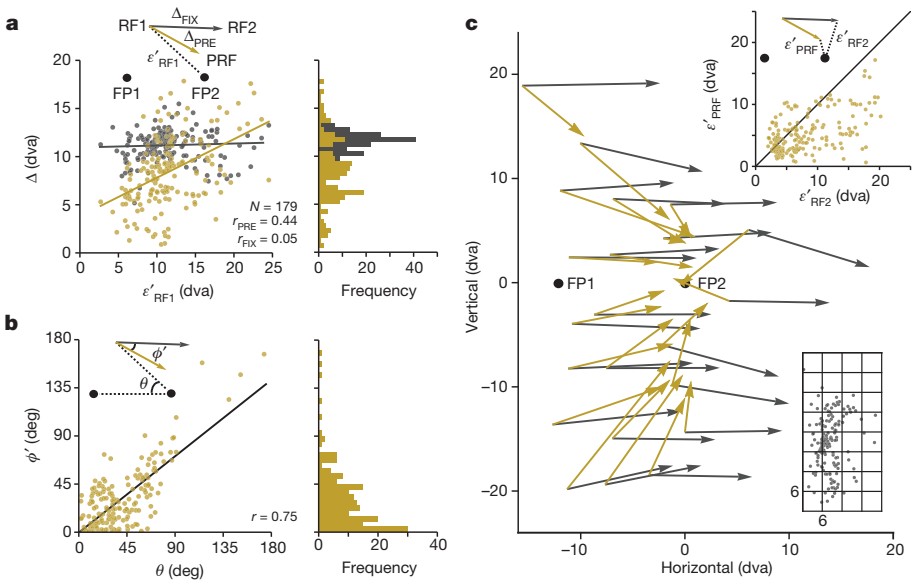


Figure 2 | Shifting of FEF neuronal RFs before saccade onset. **a**, The presaccadic shift amplitude (Δ_{PRE}) as a function of the distance of RF1 from the saccade target (FP2) (ε'_{RF1}) for 179 FEF RFs. The distance between RF1 and RF2 (Δ_{FIX}) as a function of ε'_{RF1} is shown for comparison. Lines denote linear regression fits. **b**, The angular deviation of the presaccadic RF shift from the remapping prediction (ϕ') as a function of θ. **c**, Comparison of the population of presaccadic RF shifts (gold vectors) with the remapping prediction (grey vectors). Vectors were averaged within the 6 × 6 dva spatial bins depicted in the lower right along with the distribution of individual RF1s. Top right plot shows the distances between the PRF and the saccade target compared with the distances between RF2 and the saccade target. Black line denotes the line of unity. Vector origins and end points are based on, respectively, RF1 and RF2 and RF1 and PRF.

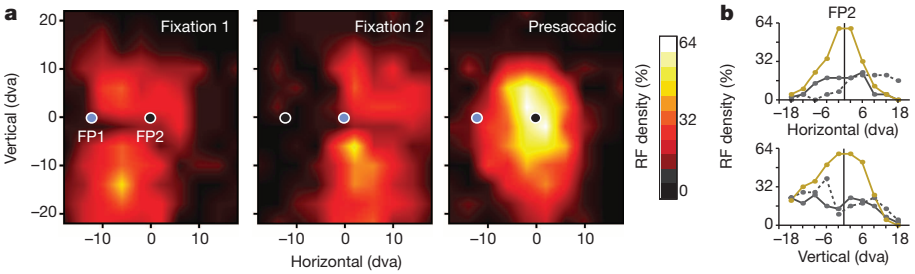


Figure 3 | Fixation and presaccadic population RF densities of recorded FEF neurons. a, Panels plot the percentage of population RFs responsive for a given probe location. Blue circles denote the location of fixation. **b**, Panels show horizontal and vertical cross-sections through RF density plots centred on the saccade target (FP2) for fixation 1 (solid grey), fixation 2 (dotted grey) and presaccadic (gold) conditions.

effect of saccade preparation on the representation of space. Specifically, we observed an increase in RF density centred on the saccade target. Within a 20 × 20 dva region around the saccade target, RF density increased by more than twofold (106.4%, $P < 10^{-3}$). Note the more than threefold increase in the proportion of RFs with responses at locations nearest the target (Fig. 3b). These changes in RF density were accompanied by alterations in the presaccadic spike count correlations of simultaneously recorded neurons (Methods and Extended Data Fig. 10), suggesting changes in effective connectivity[12,13] during saccade preparation. The above observations point towards a substantial enhancement in the representation of visual space at the saccade target.

Finally, we considered whether the presaccadic convergence of RFs to the saccade target results in a distorted 'read-out' of stimulus location. In human observers, stimulus location during saccade execution is grossly misjudged. This mislocalization results in a 'compression of visual space', with observers reporting stimuli as appearing much closer to the saccade target than they actually are[5,6]. Therefore, we decoded probe positions from the full population of recorded FEF neurons during fixation at FP1 (Fig. 4a) and during saccade preparation (Fig. 4b) (Methods). During fixation, the location of probe stimuli near FP1 ($x = [-10, -2]$, $y = [-10, 6]$), where most of the RFs were sampled, could be accurately decoded from the population response (average error = 1.3 dva). By contrast, during saccade preparation, probe locations within the same region of space were grossly mislocalized by the population of neurons (average error = 7 dva). Furthermore, the error was systematic in that the population response consistently mislocalized probes as being closer to the saccade target (Fig. 4c). The distance between the population estimate of probe location and the saccade target was reduced by 47% compared to fixation (−3.7 dva, $P < 10^{-5}$) and by 46% compared to veridical (−3.6 dva, $P < 10^{-4}$). Thus, the convergence of RFs resulted in a compression of visual space towards the target.

Predictive remapping of RFs is widely assumed to be the mechanism by which perceptual stability is achieved during saccades[1], specifically by a global anticipatory updating of visual space. Although our results demonstrate robust presaccadic shifts of FEF RFs, those shifts clearly violate the remapping prediction, and instead reveal a compression of visual space towards the saccade target. This observation raises an important question about the role of RF shifts in maintaining stability across eye movements. Specifically, how, if at all, might the convergence of RFs, rather than remapping, contribute to stable vision? It has been proposed by some that visual stability may be due to a strong bias of perceptual processing towards the targets of saccades[14]. This hypothesis argues that the failure to detect retinal image displacements results from a reduced representation of non-target locations, as compared to the overriding perception of target space[15]. Consistent with this hypothesis is psychophysical evidence of enhanced perception at the saccade target before movement onset[16,17], as well as enhanced visual cortical signals[18–21]. Furthermore, the perception of visual space is massively compressed before saccades[5,6]. Our results reveal a neuronal correlate of these perceptual effects. In particular, we found that populations of FEF neurons grossly mislocalize stimuli as being closer to the target, resembling psychophysical compression. Thus, regardless of the role of the above perceptual phenomena in visual stability, the representation within the FEF mirrors them.

FEF neurons have been causally implicated in the control of visual attention[8] and the corresponding modulation of stimulus-driven activity in posterior visual cortex[9,10,22–24]. Several recent studies suggest that the influence exerted by FEF neurons on visual cortex during attention originates from predominantly visual signals[11,25]. Our results indicate that FEF visual signals conveyed to posterior areas before saccades grossly over-represent the space occupied by the target. Thus, before each eye movement, or during covert attention[26,27], feedback from FEF neurons may impose the same distortion onto visual cortex[28–30], and this biased representation of target space could result in the aforementioned attentional enhancement within that space.

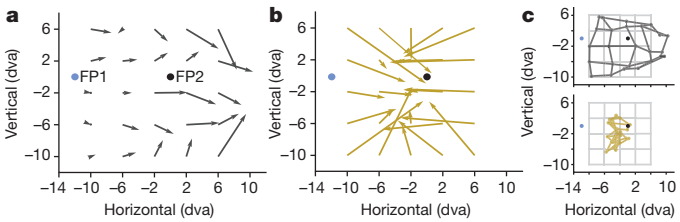


Figure 4 | Localization of probe stimuli by the population of recorded FEF neurons. a, Population estimate of probe locations within a 16 × 16 dva area of visual space during fixation at FP1. Vector origins denote the veridical probe location and the end points denote the average population estimate of that location. Vector lengths denote the errors, measured as the average Euclidean distance to the veridical probe location. **b**, Population estimate of probe locations during saccade preparation to FP2. **c**, Reconstruction of probe stimulus grid from population estimates during fixation (top, dark grey) and during saccade preparation (bottom, gold).

METHODS SUMMARY

All experimental procedures were in compliance with the US Public Health Service policy on the humane care and use of laboratory animals and the Stanford University Animal Care and Use Committee.

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 July 2013; accepted 13 February 2014.

1. Sommer, M. A. & Wurtz, R. H. Brain circuits for the internal monitoring of movements. *Annu. Rev. Neurosci.* **31,** 317–338 (2008).
2. Duhamel, J. R., Colby, C. L. & Goldberg, M. E. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* **255,** 90–92 (1992).
3. Walker, M. F., Fitzgibbon, E. J. & Goldberg, M. E. Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *J. Neurophysiol.* **73,** 1988–2003 (1995).
4. Sommer, M. A. & Wurtz, R. H. Influence of the thalamus on spatial visual processing in frontal cortex. *Nature* **444,** 374–377 (2006).
5. Ross, J., Morrone, M. C. & Burr, D. C. Compression of visual space before saccades. *Nature* **386,** 598–601 (1997).
6. Kaiser, M. & Lappe, M. Perisaccadic mislocalization orthogonal to saccade direction. *Neuron* **41,** 293–300 (2004).

7. Robinson, D. A. & Fuchs, A. F. Eye movements evoked by stimulation of frontal eye fields. *J. Neurophysiol.* **32,** 637–648 (1969).
8. Moore, T. & Fallah, M. Control of eye movements and spatial attention. *Proc. Natl Acad. Sci. USA* **98,** 1273–1276 (2001).
9. Buschman, T. J. & Miller, E. K. Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron* **63,** 386–396 (2009).
10. Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324,** 1207–1210 (2009).
11. Gregoriou, G. G., Gotts, S. J. & Desimone, R. Cell-type-specific synchronization of neural activity in FEF with V4 during attention. *Neuron* **73,** 581–594 (2012).
12. Averbeck, B. B. & Lee, D. Coding and transmission of information by neural ensembles. *Trends Neurosci.* **27,** 225–230 (2004).
13. Ecker, A. S. *et al.* Decorrelated neuronal firing in cortical microcircuits. *Science* **327,** 584–587 (2010).
14. Dodge, R. The illusion of clear vision during eye movement. *Psychol. Bull.* **2,** 193–199 (1905).
15. Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A. & Irwin, D. E. The role of the saccade target object in the perception of a visually stable world. *Percept. Psychophys.* **62,** 673–683 (2000).
16. Deubel, H. & Schneider, W. X. Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Res.* **36,** 1827–1837 (1996).
17. Rolfs, M. & Carrasco, M. Rapid simultaneous enhancement of visual sensitivity and perceived contrast during saccade preparation. *J. Neurosci.* **32,** 13744–13752a (2012).
18. Sheinberg, D. L. & Logothetis, N. K. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.* **21,** 1340–1350 (2001).
19. Khayat, P. S., Spekreijse, H. & Roelfsema, P. R. Correlates of transsaccadic integration in the primary visual cortex of the monkey. *Proc. Natl Acad. Sci. USA* **101,** 12712–12717 (2004).
20. Khavat, P. S., Spekreijse, H. & Roelfsema, P. R. Visual information transfer across eye movements in the monkey. *Vision Res.* **44,** 2901–2917 (2004).
21. Bichot, N. P., Rossi, A. F. & Desimone, R. Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* **308,** 529–534 (2005).
22. Moore, T. & Armstrong, K. M. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* **421,** 370–373 (2003).
23. Ekstrom, L. B., Roelfsema, P. R., Arsenault, J. T., Bonmassar, G. & Vanduffel, W. Bottom-up dependent gating of frontal signals in early visual cortex. *Science* **321,** 414–417 (2008).
24. Noudoost, B. & Moore, T. Control of visual cortical signals by prefrontal dopamine. *Nature* **474,** 372–375 (2011).
25. Thompson, K. G., Bichot, N. P. & Sato, T. R. Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. *J. Neurophysiol.* **93,** 337–351 (2005).
26. Connor, C. E., Preddie, D. C., Gallant, J. L. & Van Essen, D. C. Spatial attention effects in macaque area V4. *J. Neurosci.* **17,** 3201–3214 (1997).
27. Womelsdorf, T., Anton-Erxleben, K., Pieper, F. & Treue, S. Dynamic shift of visual receptive fields in cortical area MT by spatial attention. *Nature Neurosci.* **9,** 1156–1160 (2006).
28. Tolias, A. S. *et al.* Eye movements modulate visual receptive fields of V4 neurons. *Neuron* **29,** 757–767 (2001).
29. Nakamura, K. & Colby, C. L. Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proc. Natl Acad. Sci. USA* **99,** 4026–4031 (2002).
30. Hamker, F. H., Zirnsak, M., Calow, D. & Lappe, M. The peri-saccadic perception of objects and space. *PLOS Comput. Biol.* **4,** e31 (2008).

Acknowledgements This work was supported by National Institutes of Health grant EY014924 and the Howard Hughes Medical Institute (T.M.). We thank D.S. Aldrich for technical assistance.

Author Contributions M.Z. and T.M. designed the study. M.Z., B.N. and K.Z.X. performed the experiments. M.Z. and N.A.S. analysed the data. M.Z., N.A.S. and T.M. wrote the manuscript.

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

LETTER

doi:10.1038/nature12998

The E3 ligase Cbl-b and TAM receptors regulate cancer metastasis via natural killer cells

Magdalena Paolino[1], Axel Choidas[2], Stephanie Wallner[3], Blanka Pranjic[1], Iris Uribesalgo[1], Stefanie Loeser[1], Amanda M. Jamieson[4], Wallace Y. Langdon[5], Fumiyo Ikeda[1], Juan Pablo Fededa[1], Shane J. Cronin[1], Roberto Nitsch[1], Carsten Schultz-Fademrecht[2], Jan Eickhoff[2], Sascha Menninger[2], Anke Unger[2], Robert Torka[6], Thomas Gruber[3], Reinhard Hinterleitner[3], Gottfried Baier[3], Dominik Wolf[3,7], Axel Ullrich[6], Bert M. Klebl[2] & Josef M. Penninger[1]

Tumour metastasis is the primary cause of mortality in cancer patients and remains the key challenge for cancer therapy[1]. New therapeutic approaches to block inhibitory pathways of the immune system have renewed hopes for the utility of such therapies[2]. Here we show that genetic deletion of the E3 ubiquitin ligase Cbl-b (casitas B-lineage lymphoma-b) or targeted inactivation of its E3 ligase activity licenses natural killer (NK) cells to spontaneously reject metastatic tumours. The TAM tyrosine kinase receptors Tyro3, Axl and Mer (also known as Mertk) were identified as ubiquitylation substrates for Cbl-b. Treatment of wild-type NK cells with a newly developed small molecule TAM kinase inhibitor conferred therapeutic potential, efficiently enhancing anti-metastatic NK cell activity *in vivo*. Oral or intraperitoneal administration using this TAM inhibitor markedly reduced murine mammary cancer and melanoma metastases dependent on NK cells. We further report that the anticoagulant warfarin exerts anti-metastatic activity in mice via Cbl-b/TAM receptors in NK cells, providing a molecular explanation for a 50-year-old puzzle in cancer biology[3]. This novel TAM/Cbl-b inhibitory pathway shows that it might be possible to develop a 'pill' that awakens the innate immune system to kill cancer metastases.

Genetic ablation of *Cbl-b* or functional inactivation of its E3 ligase activity in mice results in $CD8^+$ T-cell-mediated rejection of primary tumours in several different models[4–6]. In a control experiment subcutaneous injection of TC-1 tumour cells into $Cbl\text{-}b^{-/-}$ $Rag2^{-/-}$ mice, which lack B and T cells, caused a significant delay in tumour growth compared to $Cbl\text{-}b^{+/+}$ $Rag2^{-/-}$ littermates (Fig. 1a), suggesting that innate immune cells are also involved. Immunohistochemistry revealed increased numbers of NK cells infiltrating the tumours of $Cbl\text{-}b^{-/-}$ mice (Extended Data Fig. 1a). Cbl-b is expressed in murine and human NK cells; expression levels were not altered in NK cells from E3 ligase-defective mice ($C373A^{KI/KI}$), which carry a cysteine to alanine knock-in (KI) mutation at position 373 (ref. 7) (Fig. 1b and Extended Data Fig. 1b). Loss of Cbl-b expression ($Cbl\text{-}b^{-/-}$) or the $C373A^{KI/KI}$ mutation had no overt effects on NK cell development (Extended Data Fig. 1c, d).

$Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ NK cells exhibited significantly increased proliferation and interferon (IFN)-γ production when activated *in vitro* and were also more efficient in killing the prototypic NK cell target tumour cell line RMA-S *in vivo* (Fig. 1c, d and Extended Data Fig. 1e–j). In contact with YAC-1 target tumour cells, $Cbl\text{-}b^{-/-}$ NK cells also displayed a higher capacity to kill, produce IFN-γ, degranulate, secrete granzyme B, and to express higher levels of the cytotoxic mediator perforin; knockdown of *CBL-B* in the human NK cell line NKL also resulted in enhanced cytotoxicity towards Jurkat cells (Extended Data Fig. 2a–h). NK cell immunodepletion using anti-NK1.1 antibodies and functional blockade of NKG2D (also known as Klrk1) receptors abolished anti-TC-1 tumour responses in $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice (Extended Data Fig. 3a–c). Moreover, subcutaneous B16F10 melanomas were slower growing in $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice; depletion of $NK1.1^+$ cells reduced this increased survival of melanoma-bearing mice (Extended Data Fig. 3d–i). Thus, Cbl-b, via its E3 ligase activity, negatively regulates NK cell functions and controls NK-cell anti-tumour responses.

We next analysed whether the absence of Cbl-b can potentiate the anti-metastatic activity of NK cells. Three weeks after intravenous B16F10 melanoma challenge, $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice exhibited reduced lung metastases and increased survival (Fig. 1e and Extended Data Fig. 4a–e). Immunodepletion of $NK1.1^+$ cells caused uncontrollable tumour growth in all $Cbl\text{-}b^{+/+}$, $Cbl\text{-}b^{+/-}$, $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice (Fig. 1f, g). NKG2D blockade in $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice also prevented the reduction of lung melanomas (Extended Data Fig. 4f–i). Of note, B16F10 melanomas by themselves do not express the NKG2D ligand Rae1, suggesting that this ligand is expressed on the tumour microenvironment. $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice also exhibited significantly reduced metastases to extrapulmonary organs (Extended Data Fig. 5a–d). In the absence of NK cells, all $Cbl\text{-}b^{+/+}$, $Cbl\text{-}b^{+/-}$, $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice displayed secondary metastases in multiple organs (Fig. 1h, i), even at a lower tumour dose (Extended Data Fig. 5e–g). Immunodepletion of $CD8^+$ T cells had no overt effect on the anti-metastases response of $Cbl\text{-}b^{-/-}$ and $C373A^{KI/KI}$ mice (Extended Data Fig. 5h, i). When backcrossed to perforin mutant mice ($Prf1^{-/-}$), $Cbl\text{-}b^{-/-}$ $Prf1^{-/-}$ double mutants were unable to reduce melanoma metastases (Fig. 1j, k and Extended Data Fig. 5j). Cbl-b has been implicated in anergic responses of natural killer T (NKT) cells[8], a cell type that also expresses NK1.1 (ref. 9). To provide definitive proof that *Cbl-b* mutant NK cells directly reject metastatic melanomas, we set up a therapeutic transplantation model. Significantly fewer B16F10 metastases were observed in the lungs of wild-type mice that received $Cbl\text{-}b^{-/-}$ or $C373A^{KI/KI}$ NK cells, compared to untreated mice and mice receiving $Cbl\text{-}b^{+/+}$ NK cells (Fig. 2a–e). Thus, NK cells lacking Cbl-b, or defective in its E3 catalytic activity, are sufficient to inhibit the progression of lung melanomas and distant metastases.

To extend our findings to a spontaneous tumour metastasis model, *Cbl-b* mutant animals were crossed to $NeuT^+$ transgenic mice that develop metastatic breast tumours[10]. Although the onset of palpable mammary tumours was comparable, loss of Cbl-b markedly suppressed the growth of mammary tumours resulting in significantly prolonged overall survival (Extended Data Fig. 6a–d). Although the number of metastatic lung tumour foci was comparable between $Cbl\text{-}b^{+/-}$ $NeuT^+$ and $Cbl\text{-}b^{-/-}$ $NeuT^+$ mice, $Cbl\text{-}b^{-/-}$ $NeuT^+$ mice showed a significant reduction in the tumour-to-lung area (Fig. 2f and Extended Data Fig. 6e–g). In $Cbl\text{-}b^{-/-}$ $NeuT^+$ mice we observed a significant accumulation of NK cells within the metastatic foci (Fig. 2g and Extended Data 6h, i); depletion of NK cells with anti-asialo GM1 antibodies markedly enhanced

[1]IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, 1030 Vienna, Austria. [2]Lead Discovery Center GmbH, D-44227 Dortmund, Germany. [3]Medical University Innsbruck, 6020 Innsbruck, Austria. [4]Department of Microbiology and Immunology, Brown University, Providence, Rhode Island 02912, USA. [5]School of Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia 6009, Perth, Australia. [6]Max-Planck, Institute for Biochemistry, Department of Molecular Biology, D-82152 Martinsried, Germany. [7]Internal Medicine III, University Hospital Bonn, 53127 Bonn, Germany.

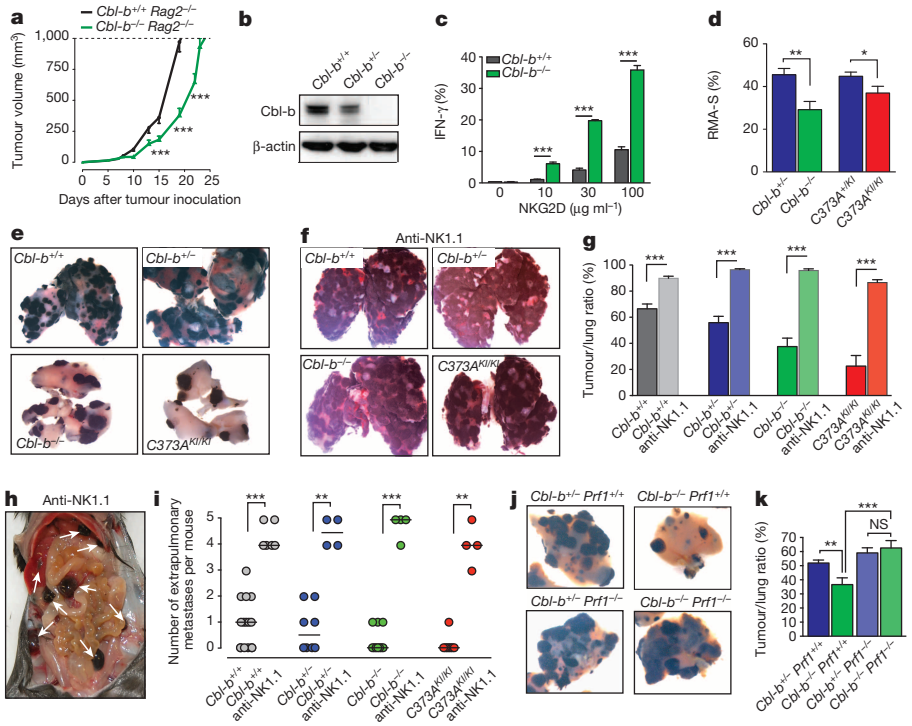
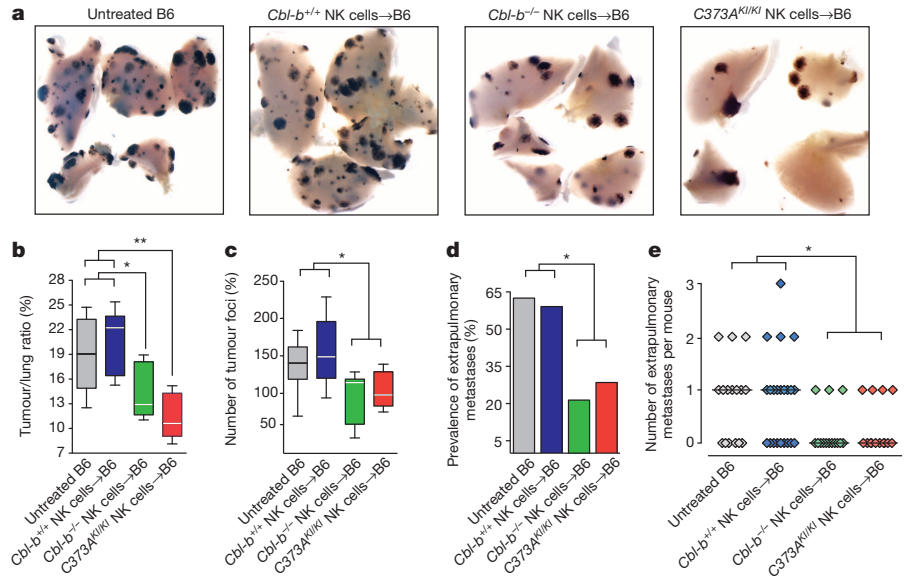


Figure 1 | *Cbl-b* mutant NK cells control metastatic melanomas. **a**, TC-1 tumour growth in *Cbl-b*$^{+/+}$ *Rag2*$^{-/-}$ and *Cbl-b*$^{-/-}$ *Rag2*$^{-/-}$ mice (mean ± s.e.m., $n = 10$ each). ***$P < 0.001$ (two-way analysis of variance (ANOVA), Bonferroni's post-hoc test). **b**, Cbl-b and β-actin protein expression in *Cbl-b*$^{+/+}$, *Cbl-b*$^{+/-}$ and *Cbl-b*$^{-/-}$ NK cells. **c**, IFN-γ^{+} *Cbl-b*$^{+/+}$ and *Cbl-b*$^{-/-}$ NK cells (%) following anti-NKG2D stimulation (mean ± s.e.m., $n = 3$). ***$P < 0.001$ (two-way ANOVA, Bonferroni's post-hoc test). **d**, *In vivo* NK cell cytotoxicity towards RMA-S cells (mean ± s.e.m., $n = 16/10/15/14$). *$P < 0.05$, **$P < 0.01$ (Student's *t*-test). **e**, **f**, Representative lung melanoma metastases in control (**e**) or NK.1.1-depleted *Cbl-b*$^{+/+}$, *Cbl-b*$^{+/-}$, *Cbl-b*$^{-/-}$ and *C373A*$^{KI/KI}$ mice (**f**) at day +21. **g**, B16F10 tumour-to-lung ratios (mean ± s.e.m.) of control and NK1.1^{+}-depleted *Cbl-b*$^{+/+}$ ($n = 12/6$), *Cbl-b*$^{+/-}$ ($n = 6/4$), *Cbl-b*$^{-/-}$ ($n = 9/4$) and *C373A*$^{KI/KI}$ ($n = 7/4$) mice. ***$P < 0.001$ (Student's *t*-test). **h**, Representative B16F10 extrapulmonary metastases in a NK1.1^{+} cell-depleted mouse. **i**, Extrapulmonary metastases in control or NK1.1-depleted mice (lines are median, day +16–21) **$P < 0.01$, ***$P < 0.001$ (Mann–Whitney *U*-test). **j**, **k**, Representative B16F10 lung metastases (**j**) and tumour-to-lung ratios (**k**) (mean ± s.e.m., day +21) in *Cbl-b*$^{+/-}$ *Prf1*$^{+/+}$, *Cbl-b*$^{-/-}$ *Prf1*$^{+/+}$, *Cbl-b*$^{+/-}$ *Prf1*$^{-/-}$ and *Cbl-b*$^{-/-}$ *Prf1*$^{-/-}$ mice. $n = 5/7/8/8$. **$P < 0.01$, ***$P < 0.001$; NS, not significant (one-way ANOVA, Tukey's post-hoc test).



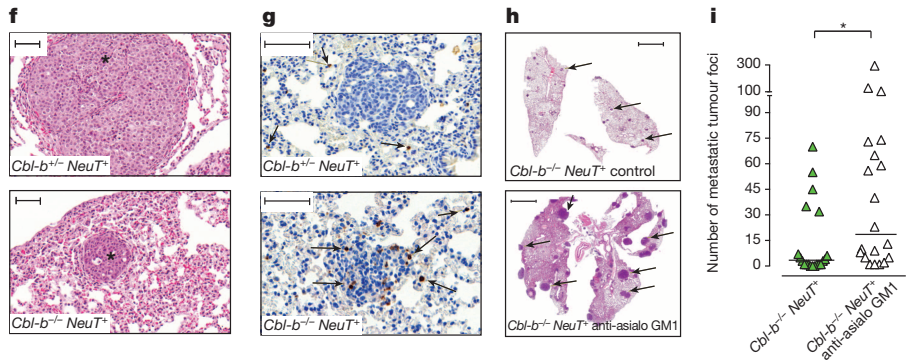


Figure 2 | Therapeutic anti-tumour activity of *Cbl-b* mutant NK cells. **a**–**d**, Representative lung metastases (**a**), tumour-to-lung ratios (**b**), numbers of tumour foci (**c**), prevalence (**d**) and numbers of extrapulmonary metastases (**e**) at day +14 in C57BL/6J (B6) mice untreated or adoptively transferred with *Cbl-b*$^{+/+}$, *Cbl-b*$^{-/-}$ and *C373A*$^{KI/KI}$ NK cells. For **b** and **c**, $n = 6/6/6/7$, *$P < 0.05$, **$P < 0.01$ (Mann–Whitney *U*-test). For **d** and **e**, $n = 16/22/14/15$, *$P < 0.05$ (Chi-squared test in **d** and Mann–Whitney *U*-test in **e**). **f**, **g**, Metastases (**f**, haematoxylin and eosin, asterisk) and NKp46^{+} NK cells (**g**, arrows) in lungs of *Cbl-b*$^{+/-}$ *NeuT*$^{+}$ and *Cbl-b*$^{-/-}$ *NeuT*$^{+}$ mice. $n = 17/20$. **h**, **i**, Representative haematoxylin and eosin-stained lung metastases (**h**, arrows) and number of metastatic foci (**i**) in control and anti-asialo GM1-treated *Cbl-b*$^{-/-}$ *NeuT*$^{+}$ mice. $n = 11/10$. *$P < 0.05$ (Mann–Whitney *U*-test, lines are median). Scale bars, 50 μM (**f**, **g**) and 2,000 μM (**h**).

metastases of mammary tumours into the lungs of *Cbl-b*$^{-/-}$ *NeuT*$^{+}$ and *Cbl-b*$^{+/-}$ *NeuT*$^{+}$ mice (Fig. 2h, i and Extended Data Fig. 7a–e). Deletion of both or even one *Prf1* allele again abrogated the capacity of *Cbl-b*$^{-/-}$ *NeuT*$^{+}$ mice to control mammary metastases to the lung (Extended Data Fig. 7f–h). Thus, *Cbl-b*-deficient NK cells can also control metastatic tumours in oncogene-driven mammary cancer.

To identify ubiquitylation substrates, we performed *in vitro* ubiquitylation reactions on 9,000 human proteins. The transmembrane receptor tyrosine kinase TYRO3 was identified as the top candidate for CBL-B-mediated ubiquitylation (Extended Data Fig. 8a). Tyro3, together with Axl and Mer, constitute the TAM family of cell-surface tyrosine kinase receptors[11]. In additional *in vitro* assays using recombinant proteins, we confirmed that CBL-B ubiquitylates all three TAM family proteins (Fig. 3a and Extended Data Fig. 8b). Activation of HeLa cells with GAS6, a known ligand for TAM receptors[12], resulted in recruitment of CBL-B to TAM receptors and AXL ubiquitylation; *CBL-B* knock-down using short interfering RNA (siRNA) reduced GAS6-induced AXL ubiquitylation (Extended Data Fig. 8c, d). Thus, TAM receptors are molecular substrates for CBL-B ubiquitylation *in vitro* and *in vivo*.

In contrast to TAM triple-mutant mice[13], loss of Cbl-b had no overt effect on NK cell development (Extended Data Fig. 1c, d). The absence of Cbl-b, or its enzymatic E3 ligase activity, also had no apparent effect on basal Tyro3, Axl or Mer surface expression in splenic NK cells (Extended Data Fig. 8e). Because TAM receptors can inhibit macrophage and dendritic cell activation[14], we proposed that TAM receptors might also negatively regulate NK cells. Gas6 indeed suppressed proliferation and IFN-γ production of NKG2D-activated NK cells; *Cbl-b*$^{-/-}$ NK cells were resistant to this inhibition (Fig. 3b and Extended Data Fig. 8f). Mechanistically, whereas Gas6 stimulation resulted in downregulation of all three TAM receptors from the cell surface, this downregulation was impaired in NK cells lacking Cbl-b (Fig. 3c). Impaired receptor downregulation was confirmed in HeLa cells using siRNA-mediated *CBL-B* knockdown (Extended Data Fig. 8g, h). Thus, Cbl-b/TAM receptors constitute an inhibitory pathway for NK cell activation.

To assess whether modulation of this pathway could be used for anti-tumour immunotherapy, we developed a highly selective TAM kinase inhibitor, termed LDC1267. LDC1267 preferentially inhibits Tyro3, Axl and Mer at low nanomolarity, as determined by tracer-based binding assays (Fig. 3d, e). Selectivity of LDC1267 was further assessed using a cell-free KINOMEscan assay covering 456 kinases (Fig. 3f and Supplementary Table 1). Cellular selectivity was confirmed using a quantitative chemical proteomics approach, cell-based phosphorylation and proliferation assays (Extended Data Fig. 9a–d). Treatment of NKG2D-activated NK cells with LDC1267 indeed abolished the inhibitory effects of Gas6 stimulation; LDC1267 had no apparent additional effect in *Cbl-b*-deficient NK cells (Fig. 3g). *In vivo*, wild-type mice treated

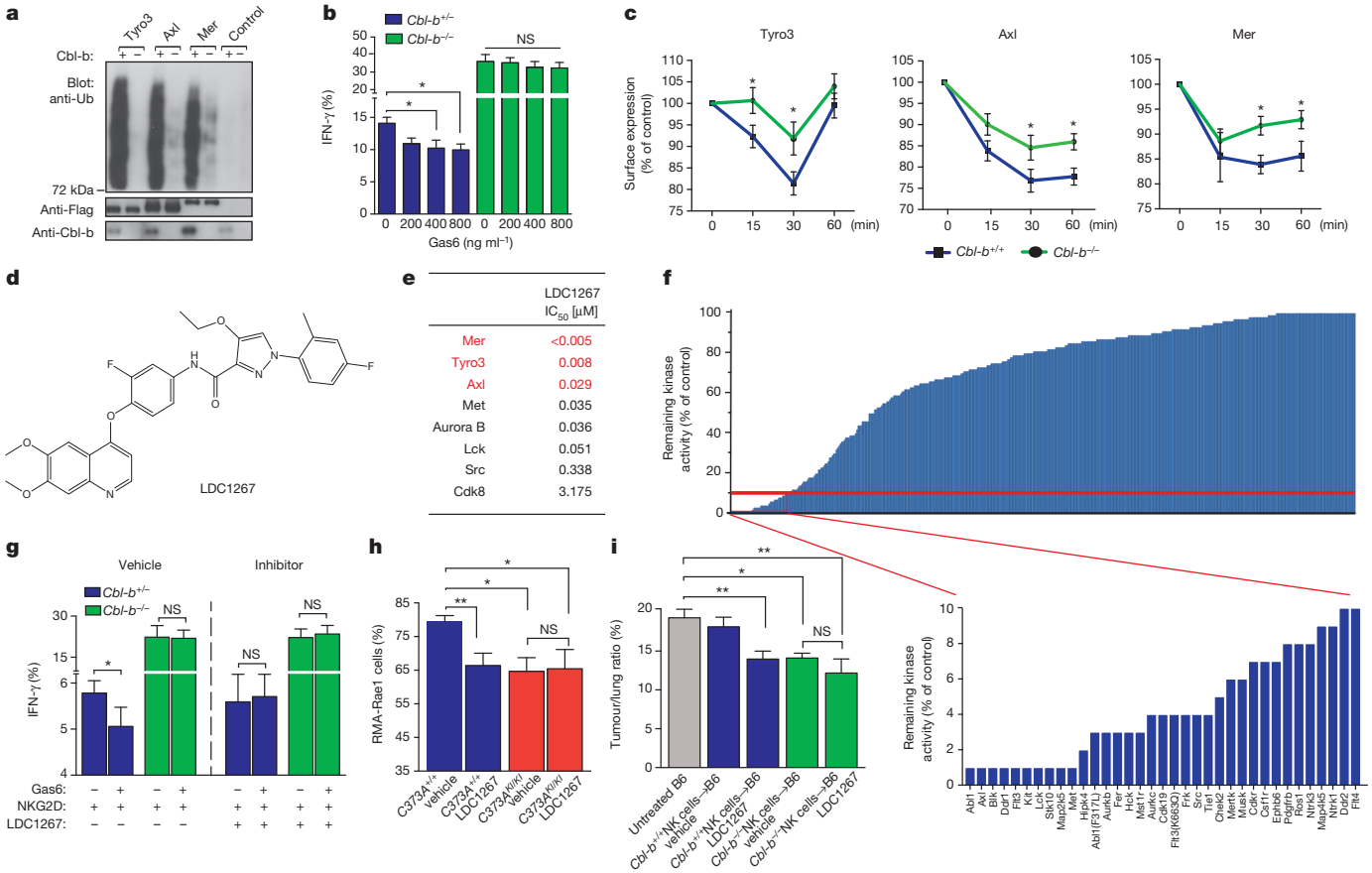


Figure 3 | TAM receptors are Cbl-b targets and characterization of a TAM receptor blocker. **a**, *In vitro* Cbl-b-dependent ubiquitylation of Tyro3, Axl and Mer (anti-Ub). Control, no TAM receptors. Loading controls are shown. **b**, IFN-γ^{+} *Cbl-b*$^{+/-}$ and *Cbl-b*$^{-/-}$ NK cells (%) stimulated with anti-NKG2D antibodies in the presence of soluble Gas6. $^{*}P < 0.05$ (one-way ANOVA, Dunnett's post-hoc test, $n = 6/5$). **c**, TAM receptor surface expression in NK cells after stimulation with Gas6. $^{*}P < 0.05$ (two-way ANOVA, Bonferroni's post-hoc test, $n = 10–13$). **d**, Chemical structure of the TAM receptor kinase inhibitor LDC1267. **e**, Half maximal inhibitory concentrations (IC_{50}) for the indicated protein kinases as determined by tracer assays. **f**, Remaining activity (compared to dimethylsulphoxide (DMSO) control) in 456 kinases treated with 1 μM LDC1267. **g**, IFN-γ^{+} *Cbl-b*$^{+/-}$ and *Cbl-b*$^{-/-}$ NK cells (%) pre-treated with vehicle or LDC1267 and then stimulated with anti-NKG2D antibodies and Gas6. $^{*}P < 0.05$; NS, not significant (Student's *t*-test, $n = 11/8$). **h**, *In vivo* NK cytotoxicity towards RMA-Rae1 cells in mice treated with vehicle or LDC1267. $^{*}P < 0.05$, $^{**}P < 0.01$ (one-way ANOVA, Tukey's post hoc test, $n = 16/14/9/10$). **i**, Tumour-to-lung areas in B16F10 tumour-bearing B6 mice untreated or adoptively transplanted with NK cells *ex vivo*-treated with vehicle or LDC1267. $^{*}P < 0.05$ and $^{**}P < 0.01$ (one-way ANOVA, Tukey's post-hoc test, $n = 16/16/16/10/10$). Data in **b**, **c**, **g**–**i** are mean ± s.e.m.

with LDC1267 showed enhanced cytotoxicity towards RMA cells over-expressing the NKG2D ligand Rae-1 (RMA-Rae1) to the same extent as *C373A*$^{KI/KI}$ mice, but had no effect on the already enhanced NK cytotoxicity in *Cbl-b*-mutant mice (Fig. 3h and Extended Data 9e). To provide definitive proof that LDC1267 treatment can license NK cells to kill, we treated B16F10 melanoma-bearing mice with NK cells that were either untreated or treated *ex vivo* with our TAM inhibitor (Extended Data Fig. 9f). Adoptive transfer of LDC1267-treated wild-type NK cells significantly increased the anti-tumour response to levels observed in mice transplanted with *Cbl-b*$^{-/-}$ NK cells, but did not increase the anti-metastatic efficacy of *Cbl-b*-knockout NK cells (Fig. 3i), reinforcing the notion that Cbl-b acts downstream of TAM receptors for NK cell inhibition. These results indicate that TAM receptor inhibition using LDC1267 unleashes NK cells to kill tumour cells *in vivo*.

To test whether our novel TAM inhibitor can be used in an *in vivo* therapeutic setting, we challenged mice with B16F10 melanoma followed by intraperitoneal LDC1267 treatment. LDC1267 markedly reduced metastatic spreading of melanomas; NK1.1 depletion abolished the therapeutic benefits of LDC1267 (Fig. 4a–c). We next set-up an orthotopic metastatic breast cancer model, that is, injection of 4T1 breast cancer cells into the mammary fat pad of BALB/c mice[15]. LDC1267 treatment significantly reduced the numbers and sizes of 4T1 micro-metastases in the liver, without any apparent effect on the primary mammary tumour (Fig. 4d–f and Extended Data Fig. 10a–c). NK cell depletion using anti-asialo GM1 antibodies resulted in markedly enhanced 4T1 liver metastases and abolished the therapeutic benefits of LDC1267 (Fig. 4d–f). We finally tested whether LDC1267 could be also effective in the 4T1 liver metastasis model, when administrated orally (Extended Data Fig. 10d). Oral LDC1267 significantly reduced 4T1 liver micro-metastases (Extended Data Fig. 10e–g). Thus, using different model systems and different routes of administration, *in vivo* therapy with the TAM inhibitor LDC1267 consistently reduced metastases dependent on NK cells.

Warfarin is the most widely used oral anticoagulant worldwide[16] and has been known to reduce tumour metastases in mouse and rat models[3,17]. The underlying mechanisms were largely speculative[18,19]. Besides its effects on coagulation, warfarin prevents γ-carboxylation of TAM ligands, rendering Gas6 unable to activate TAM receptors[20]. We therefore proposed that warfarin could exhibit anticancer effects via modulation of the Gas6/TAM/Cbl-b pathway in NK cells. Administration of low-dose warfarin, which inhibits TAM receptor activity *in vivo*[21,22] without affecting coagulation, markedly reduced melanoma metastases to the lungs and distal organs of *Cbl-b*$^{+/+}$ and *Cbl-b*$^{+/-}$ mice but had no apparent effect on *Cbl-b*-defective mice; the absence of NK1.1^{+} cells abolished the effects of warfarin on metastatic melanomas (Fig. 4g–j and Extended Data Fig. 10h–j). Moreover, warfarin treatment resulted

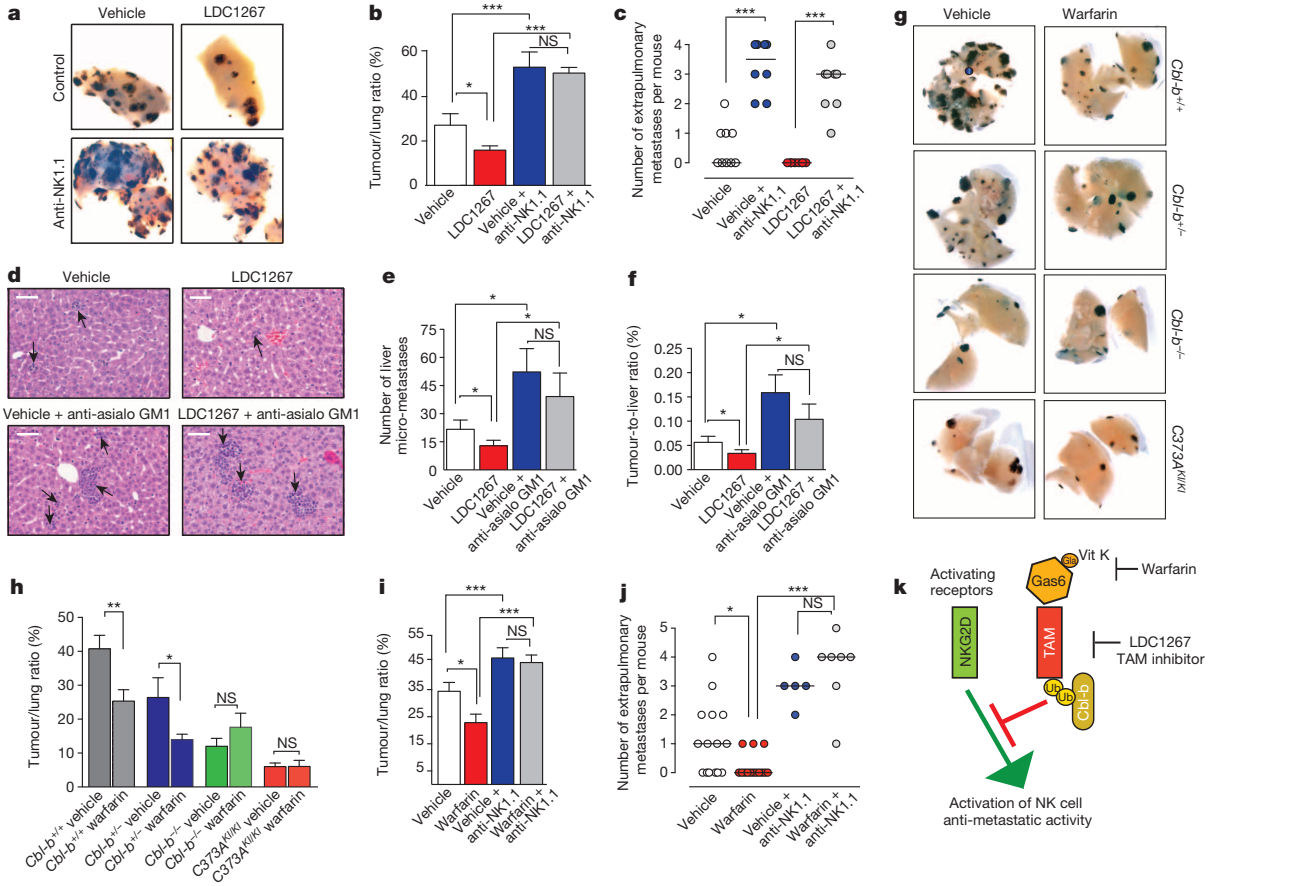


Figure 4 | Control of metastases by TAM receptor inhibition. **a**–**c**, Lung melanoma metastases (**a**), tumour-to-lung ratios (**b**, mean ± s.e.m.) and extrapulmonary metastases (**c**, lines are median) in vehicle- and LDC1267-treated wild-type control or NK1.1-depleted mice. $n = 8$ each. $^{*}P < 0.05$, $^{***}P < 0.001$; NS, not significant (Student's *t*-test in **b** and Mann–Whitney *U*-test in **c**). **d**–**f**, Representative images (**d**, arrows; scale bars, 50 μm), and numbers (**e**) and relative sizes (**f**) of 4T1 liver micro-metastasis in control or asialo-GM1 immunodepleted mice treated with vehicle or LDC1267 (20 mg kg^{-1}, daily intraperitoneal injection). Mean ± s.e.m., $n = 7/9/8/8$. $^{*}P < 0.05$; NS, not significant (Mann–Whitney *U*-test). **g**, **h**, Representative lung B16F10 melanoma metastases (**g**) and tumour-to-lung ratios (**h**) in vehicle- and warfarin-treated (0.5 mg l^{-1}) *Cbl-b*$^{+/+}$ ($n = 12/12$), *Cbl-b*$^{+/-}$ ($n = 6/7$), *Cbl-b*$^{-/-}$ ($n = 4/7$) and *C373A*$^{KI/KI}$ ($n = 8/8$) mice. Mean ± s.e.m. $^{**}P < 0.01$, $^{*}P < 0.05$; NS, not significant (Student's *t*-test). **i**, **j**, Relative lung tumour area (**i**, mean ± s.e.m.) and extrapulmonary melanoma metastases (**j**, lines are median) in *Cbl-b*$^{+/+}$ mice treated with vehicle ($n = 15$), warfarin ($n = 14$), vehicle and anti-NK1.1 antibodies ($n = 5$) or warfarin and anti-NK1.1 antibodies ($n = 7$). $^{*}P < 0.05$, $^{***}P < 0.001$; NS, not significant (one-way ANOVA, Tukey's post-hoc test for **i**; Mann–Whitney *U*-test for **j**). **k**, Schematic illustrating the TAM receptor/Cbl-b inhibitory pathway controlling NK cell anti-metastatic activities.

in enhanced *in vivo* NK cell cytotoxicity towards their prototypic targets RMA-S and RMA-Rae1 (Extended Data Fig. 10k). Our data support a coagulation-independent role of warfarin in tumour metastasis that relies on functional TAM receptors/Cbl-b in NK cells.

Our experiments uncover a novel inhibitory role of the E3 ligase Cbl-b in the regulation of NK cell functions via TAM receptors (Fig. 4k). Releasing the inhibition imposed by this TAM/Cbl-b pathway renders NK cells capable of rejecting tumour metastases. There is high redundancy for TAM receptor functions in immune cells[23]. Because all three TAM receptors are expressed on NK cells[13], they most likely, as they do in NK cell development[13], also exert redundant functions in NK cell activation. Indeed, genetic deletion of *Tyro3* alone affected neither NK cell activation nor *in vivo* metastases control by NK cells (not shown). Mechanistically, upon ligand stimulation, Cbl-b is recruited to and ubiquitylates TAM receptors, which in turn appears to control TAM receptor internalization, corroborating the role of Cbl-b and its homologue c-Cbl in receptor tyrosine kinase endocytosis[24]. Of note, c-CBL has been implicated in the ubiquitylation of AXL in human epithelial HLEC cells[25]. Whether Cbl-b affects NK cell activation solely via TAM receptors and/or also directly impinges on other receptor systems needs to be determined.

Current tumour immunotherapies face the challenge of severe side effects[26]. As *Cbl-b*-knockout mice and mice treated with our TAM blocker do not exert any severe signs of inflammation or autoimmunity, there might be a therapeutic window for Cbl-b/TAM receptor inhibition to mediate tumour rejection without serious cytotoxicity. We also provide a molecular explanation for the longstanding conundrum on how warfarin confers anti-metastatic activity, offering the possibility to re-assess the use of vitamin K antagonists such as warfarin in cancer therapy. Importantly, the anti-tumour effects of our TAM inhibitor provide a proof of principle that it might be possible to develop a 'pill' that awakens the innate immune system to kill cancer metastases.

METHODS SUMMARY

Cbl-b and *C373A*$^{KI/KI}$ mice have been reported previously[7,27]. Subcutaneous TC-1 and B16F10 tumours were induced by injection of 2.5×10^5 tumour cells in the left flank of mice. For 4T1 tumours, 5×10^5 4T1 cells were inoculated directly into the third mammary fat pad. Mammary tumours and lung metastases in *NeuT*$^+$ mice arise spontaneously[10]. For melanoma metastases, 2.5×10^5 B16F10 cells were injected intravenously into the tail veins. Lungs were dissected for tumour quantification and the presence of metastases in any other organ was recorded. For adoptive transfer of NK cells, wild-type C57BL/6J recipient mice were inoculated intravenously with 2.5×10^5 B16F10 at day 0, and one day later were either left untreated or inoculated intravenously with 1×10^5 purified *Cbl-b*$^{+/+}$, *Cbl-b*$^{-/-}$ or *C373A*$^{KI/KI}$ *ex vivo*-expanded NK cells. For blocking TAM receptors, mice were orally treated in their drinking water with low doses of warfarin (0.5 mg ml^{-1}). The TAM small molecule kinase inhibitor LDC1267 was developed applying rationale design based on the homology modelling and known non-specific TAM receptor kinase inhibitors. To study the therapeutic potential of TAM inhibition, *ex vivo*-expanded and purified NK cells were treated for 2.5 h with LDC1267 (2.5 μM) or vehicle and intravenously injected into wild-type C57BL/6J recipient mice bearing B16F10 tumours. For 4T1 and melanoma metastases experiments, mice were treated with LDC1267 using daily oral gavage (100 mg kg^{-1}) or intraperitoneal injections (20 mg kg^{-1}). All tumours were quantified blindly on the basis of a histomorphometric analysis using the Definiens Developer XD image analyser software.

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 November 2012; accepted 3 January 2014.
Published online 19 February 2014.

1. Brabletz, T., Lyden, D., Steeg, P. S. & Werb, Z. Roadblocks to translational advances on metastasis research. *Nature Med.* **19,** 1104–1109 (2013).
2. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nature Rev. Cancer* **12,** 252–264 (2012).
3. Lacour, F., Oberling, C. & Guerin, M. Effect of dicoumarol on the development of metastases of the T8 epithelioma in the rat; new research. *Bull. Assoc. Fr. Etud. Cancer* **44,** 88–91 (1957).
4. Loeser, S. *et al.* Spontaneous tumor rejection by *cbl-b*-deficient CD8$^+$ T cells. *J. Exp. Med.* **204,** 879–891 (2007).
5. Chiang, J. Y., Jang, I. K., Hodes, R. & Gu, H. Ablation of Cbl-b provides protection against transplanted and spontaneous tumors. *J. Clin. Invest.* **117,** 1029–1036 (2007).
6. Paolino, M. *et al.* Essential role of E3 ubiquitin ligase activity in *Cbl-b*-regulated T cell functions. *J. Immunol.* **186,** 2138–2147 (2011).
7. Oksvold, M. P., Dagger, S. A., Thien, C. B. & Langdon, W. Y. The Cbl-b RING finger domain has a limited role in regulating inflammatory cytokine production by IgE-activated mast cells. *Mol. Immunol.* **45,** 925–936 (2008).
8. Kojo, S. *et al.* Mechanisms of NKT cell anergy induction involve Cbl-b-promoted monoubiquitination of CARMA1. *Proc. Natl Acad. Sci. USA* **106,** 17847–17851 (2009).
9. Makino, Y., Kanno, R., Ito, T., Higashino, K. & Taniguchi, M. Predominant expression of invariant $V_\alpha 14^+$ TCR α chain in NK1.1$^+$ T cell populations. *Int. Immunol.* **7,** 1157–1161 (1995).
10. Pannellini, T., Forni, G. & Musiani, P. Immunobiology of her-2/neu transgenic mice. *Breast Dis.* **20,** 33–42 (2004).
11. Linger, R. M., Keating, A. K., Earp, H. S. & Graham, D. K. TAM receptor tyrosine kinases: biologic functions, signaling, and potential therapeutic targeting in human cancer. *Adv. Cancer Res.* **100,** 35–83 (2008).
12. Nagata, K. *et al.* Identification of the product of growth arrest-specific gene *6* as a common ligand for Axl, Sky, and Mer receptor tyrosine kinases. *J. Biol. Chem.* **271,** 30022–30027 (1996).
13. Caraux, A. *et al.* Natural killer cell differentiation driven by Tyro3 receptor tyrosine kinases. *Nature Immunol.* **7,** 747–754 (2006).
14. Lemke, G. & Rothlin, C. V. Immunobiology of the TAM receptors. *Nature Rev. Immunol.* **8,** 327–336 (2008).
15. Pulaski, B. A. & Ostrand-Rosenberg, S. Mouse 4T1 breast tumor model. *Curr. Protoc. Immunol.* **20,** Unit 20 22 (2001).
16. Pirmohamed, M. Warfarin: almost 60 years old and still causing problems. *Br. J. Clin. Pharmacol.* **62,** 509–511 (2006).
17. Ryan, J. J., Ketcham, A. S. & Wexler, H. Reduced incidence of spontaneous metastases with long-term Coumadin therapy. *Ann. Surg.* **168,** 163–168 (1968).
18. McCulloch, P. & George, W. D. Warfarin inhibits metastasis of *Mtln3* rat mammary carcinoma without affecting primary tumour growth. *Br. J. Cancer* **59,** 179–183 (1989).
19. Brown, J. M. A study of the mechanism by which anticoagulation with warfarin inhibits blood-borne metastases. *Cancer Res.* **33,** 1217–1224 (1973).
20. Hasanbasic, I., Rajotte, I. & Blostein, M. The role of γ-carboxylation in the anti-apoptotic function of gas6. *J. Thromb. Haemost.* **3,** 2790–2797 (2005).
21. Yanagita, M. *et al.* Gas6 regulates mesangial cell proliferation through Axl in experimental glomerulonephritis. *Am. J. Pathol.* **158,** 1423–1432 (2001).
22. Nagai, K. *et al.* Growth arrest-specific gene 6 is involved in glomerular hypertrophy in the early stage of diabetic nephropathy. *J. Biol. Chem.* **278,** 18229–18234 (2003).
23. Lu, Q. & Lemke, G. Homeostatic regulation of the immune system by receptor tyrosine kinases of the Tyro 3 family. *Science* **293,** 306–311 (2001).
24. Haglund, K. & Dikic, I. The role of ubiquitylation in receptor endocytosis and endosomal sorting. *J. Cell Sci.* **125,** 265–275 (2012).
25. Valverde, P. Effects of Gas6 and hydrogen peroxide in Axl ubiquitination and downregulation. *Biochem. Biophys. Res. Commun.* **333,** 180–185 (2005).
26. Caspi, R. R. Immunotherapy of autoimmunity and cancer: the penalty for success. *Nature Rev. Immunol.* **8,** 970–976 (2008).
27. Bachmaier, K. *et al.* Negative regulation of lymphocyte activation and autoimmunity by the molecular adaptor Cbl-b. *Nature* **403,** 211–216 (2000).

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

Acknowledgements We thank T. Hanada, R. Hanada, R. Karim and all other members of the Penninger laboratory for discussions and technical support. We thank all members of the IMP-IMBA BioOptics service facility for assistance in cell sorting and image quantification, and S. Soto and H. Popper for pathology analysis of tumour metastases. We thank A. L. Prieto, E. Vivier, D. Raulet and C. Melief for providing us with critical reagents and C. Martinez, A. Majoros and P. C. Esk for sharing reagents and for discussions. We also thank G. Kéri, L. Örfi and colleagues from Vichem Kft., Budapest, for their initial synthetic work on the quinoline-based Axl inhibitors, which served as the basis for the current rationale design of LDC1267. M.P. is supported by the European Research Council (ERC) and Era of Hope/DoD Innovator Award. J.M.P. is supported by grants from the Institute of Molecular Biotechnology (IMBA), the Austrian National Foundation, the Austrian Academy of Sciences, GEN-AU (AustroMouse), an Era of Hope/DoD Innovator Award and an EU ERC Advanced Grant.

Author Contributions All experiments, with the exceptions listed below, were performed by M.P. B.P. helped with mouse dissections, *in vivo* administration of LDC1267 and melanoma lung tumour quantifications. I.U. helped with 4T1 tumour experiments and performed Rae1 qPCR analysis on lung melanoma tumour samples. S.L. performed TC1 tumour experiments on *Rag2*$^{-/-}$ mice and the histological analysis of *NeuT*$^+$ mammary gland tumours. A.C., C.S.-F., J.E., S.M., A.U., R.T., A.U. and B.M.K. developed the TAM small molecule inhibitor LDC1267 and characterized its kinase selectivity. S.W. and D.W. performed Cbl-b qRT–PCR, ^{51}Cr-release NK cytotoxic assay, and perforin FACS analysis. R.N., S.J.C., T.G., R.H. and G.B. performed cell-free *in vitro* ubiquitylation experiments. W.Y.L. provided Cbl-b RING finger mutant mice. F.I., A.M.J. and J.P.F. provided reagents and expertise in key experiments. M.P. together with J.M.P. designed the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M.P. (josef.penninger@imba.oeaw.ac.at).

LETTER

doi:10.1038/nature12910

Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development

Xindong Liu[1,2], Xin Chen[1], Bo Zhong[2]†, Aibo Wang[1,2], Xiaohu Wang[1,2], Fuliang Chu[3], Roza I. Nurieva[2], Xiaowei Yan[4], Ping Chen[5], Laurens G. van der Flier[6]†, Hiroko Nakatsukasa[7], Sattva S. Neelapu[3], Wanjun Chen[7], Hans Clevers[6], Qiang Tian[4], Hai Qi[1], Lai Wei[8] & Chen Dong[1,2]

In immune responses, activated T cells migrate to B-cell follicles and develop into follicular T-helper (T_{FH}) cells, a recently identified subset of $CD4^+$ T cells specialized in providing help to B lymphocytes in the induction of germinal centres[1,2]. Although Bcl6 has been shown to be essential in T_{FH}-cell function, it may not regulate the initial migration of T cells[3] or the induction of the T_{FH} program, as exemplified by C-X-C chemokine receptor type 5 (CXCR5) upregulation[4]. Here we show that expression of achaete-scute homologue 2 (Ascl2)—a basic helix–loop–helix (bHLH) transcription factor[5]—is selectively upregulated in T_{FH} cells. Ectopic expression of *Ascl2* upregulates CXCR5 but not Bcl6, and downregulates C-C chemokine receptor 7 (CCR7) expression in T cells *in vitro*, as well as accelerating T-cell migration to the follicles and T_{FH}-cell development *in vivo* in mice. Genome-wide analysis indicates that Ascl2 directly regulates T_{FH}-related genes whereas it inhibits expression of T-helper cell 1 (T_H1) and T_H17 signature genes. Acute deletion of Ascl2, as well as blockade of its function with the Id3 protein in $CD4^+$ T cells, results in impaired T_{FH}-cell development and germinal centre response. Conversely, mutation of Id3, known to cause antibody-mediated autoimmunity, greatly enhances T_{FH}-cell generation. Thus, Ascl2 directly initiates T_{FH}-cell development.

The development of T_{FH} cells is initiated by and dependent on their movement out of the T-cell zone and into the B-cell follicle. This migration process is regulated by upregulation of CXCR5 as well as downregulation of both CCR7 and P-selectin glycoprotein ligand 1 (PSGL1)[2,6]. T_{FH} cells have unique developmental regulation and Bcl6 was reported to be selectively expressed in T_{FH} cells[7–9]. However, although Bcl6 potentiates T_{FH}-cell generation *in vivo*, recent data suggest that it may not regulate CXCR5 upregulation by activated T cells or their migration to B-cell follicles *in vivo*[3,4]. Hence, the transcriptional mechanisms underlying initial T_{FH}-cell commitment remain unclear[4].

Recently, we observed that a bHLH-domain-containing transcription factor, Ascl2, was highly expressed in $CXCR5^{hi}Bcl6^{hi}$ cells in comparison with $CXCR5^-Bcl6^-$ T cells[4]. Interestingly, $CXCR5^+Bcl6^{lo}$ T cells also exhibited upregulation of *Ascl2* messenger RNA expression[4], suggesting that its upregulation may precede that of Bcl6. Moreover, the *Ascl2* gene locus was marked with active chromatin marker trimethylated histone H3 lysine 4 (H3K4me3) in T_{FH} and, to a much less extent, T_H2 cells, but not in other T-cell subsets, whereas the other T_{FH}-regulating genes *Bcl6* (refs 7–9), *Maf*[10], *Batf*[11,12] and *Irf4* (ref. 13) were uniformly associated with H3K4me3 in all T-cell subsets (Extended Data Fig. 1a). To validate these results, we sorted three subpopulations of cells ($CXCR5^-RFP^-$, $CXCR5^+RFP^{lo}$ and $CXCR5^{hi}RFP^{hi}$; RFP, red fluorescent protein) from *Bcl6-RFP* reporter mice immunized with keyhole limpet haemocyanin (KLH) emulsified with complete Freund's adjuvant (CFA) (Fig. 1a), and found that Ascl2 was highly expressed in T_{FH} cells at both mRNA and protein levels (Fig. 1b and Extended Data Fig. 1b). Also, Ascl2 expression was closely correlated with that of CXCR5 (Fig. 1b) and was higher in T_{FH} cells than in other T-cell subsets (Fig. 1c). In human T cells, expression of Ascl2 as well as CXCR5 and Bcl6 was found in human tonsil $CXCR5^{hi}PD1^{hi}$ T_{FH} cells (Fig. 1d, e). Collectively, these results suggest that Ascl2 is highly expressed in T_{FH} cells and that its expression may precede that of Bcl6.

Bcl6 and Batf are necessary in T_{FH}-cell development[6,12], whereas Stat5 inhibits T_{FH}-cell development[14,15]. Overexpression of Bcl6 or Batf, or Stat5 deficiency, failed to increase Ascl2 expression (Extended Data Fig. 1c). None of the known stimuli, including anti-CD3, anti-CD28, anti-ICOS, interleukin (IL)-6 and IL-21, nor their combination, upregulated Ascl2 expression in T cells (Extended Data Fig. 1d). Ascl2 was previously shown to be a target of canonical Wnt signalling in intestinal stem cells[5], and we also found that Ascl2 and CXCR5 but not Bcl6 expression in $CD4^+$ T cells could be upregulated by TWS119 (ref. 16; Fig. 1f and Extended Data Fig. 1d, e) or other Wnt agonists (data not shown).

As a first step to examine the function of Ascl2 in T_{FH} cells, retroviral overexpression of Ascl2 was conducted in $CD4^+$ T cells, leading to substantial induction of CXCR5 expression in over 30% of transduced cells, whereas overexpression of Bcl6, Batf or Maf in purified T cells did not (Fig. 2a and Extended Data Fig. 2a). Ascl2 overexpression increased *Cxcr5* mRNA expression by ~60 fold (Fig. 2b), without affecting *Bcl6*, *Prdm1*, *Batf*, *Sh2d1a*, *Cd40lg*, *Icos*, *Pdcd1*, *Btla* and *Il21* expression (Fig. 2c). CXCR5 expression was equally induced by Ascl2 in wild-type, $Bcl6^{-/-}$ and $Batf^{-/-}$ $CD4^+$ T cells *in vitro* (Fig. 2d). Thus, our findings suggest that Ascl2 is unique in its ability to induce CXCR5 protein expression in $CD4^+$ T cells *in vitro*. CCR7 and PSGL1 (*Selplg*) as well as CD25 (*Il2ra*) and CD122 (*Il2rb*) expression were downregulated in Ascl2-overexpressing T cells (Fig. 2e and Extended Data Fig. 2b), probably accounting for increased follicular homing ability and decreased IL-2 signalling in T_{FH} cells[14,15,17]. In addition, T_H1- and T_H17-related signature genes were strongly suppressed by Ascl2 (Extended Data Fig. 2c).

We then assayed the role of Ascl2 *in vivo* by transferring Ascl2-transduced OT-II cells. In $Tcrb^{-/-}$ recipient mice, at day 2 after immunization with 4-hydroxy-3-nitrophenyl (NP)- ovalbumin (OVA) in CFA, neither CXCR5 nor Bcl6 expression were detectable in the vector-transduced control group, whereas Ascl2 overexpression strongly increased in $CXCR5^+Bcl6^{lo}$ cells (Fig. 2f, g). By contrast, ectopic expression of Bcl6 did not promote T_{FH}-cell generation at this time point (Extended Data Fig. 2d, e). At day 6 after immunization, Ascl2 overexpression induced a higher percentage of $CXCR5^{hi}Bcl6^{hi}$ T_{FH} cells (Fig. 2f, g). Accordingly, germinal centre B cells and the total area of the germinal centre at day 8 in C57BL/6 mice receiving Ascl2-transdued T cells were significantly increased (Fig. 2h–j). Anti-NP immunoglobulin (Ig)M, IgA, IgG1 as well as IgG3 titres were increased in $Tcrb^{-/-}$ mice, whereas

[1]Tsinghua University School of Medicine, Beijing 100084, China. [2]Department of Immunology, MD Anderson Cancer Center, Houston, Texas 77054, USA. [3]Department of Lymphoma and Myeloma, MD Anderson Cancer Center, Houston, Texas 77054, USA. [4]Institute for Systems Biology, Seattle, Washington 98103, USA. [5]Laboratory of Immunology, National Eye Institute, NIH, Bethesda, Maryland 20892-1858, USA. [6]Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands. [7]National Institute of Dental and Craniofacial Research, NIH, Bethesda, Maryland 20892-2190, USA. [8]State Key Laboratory of Ophthalmology, Sun Yat-sen University, Guangzhou 510275, China. †Present addresses: College of Life Sciences, Wuhan University, Wuhan 430072, China (B.Z.); SomantiX B.V., Padualaan 8, 3584 CH Utrecht, the Netherlands (L.G.v.d.F.).

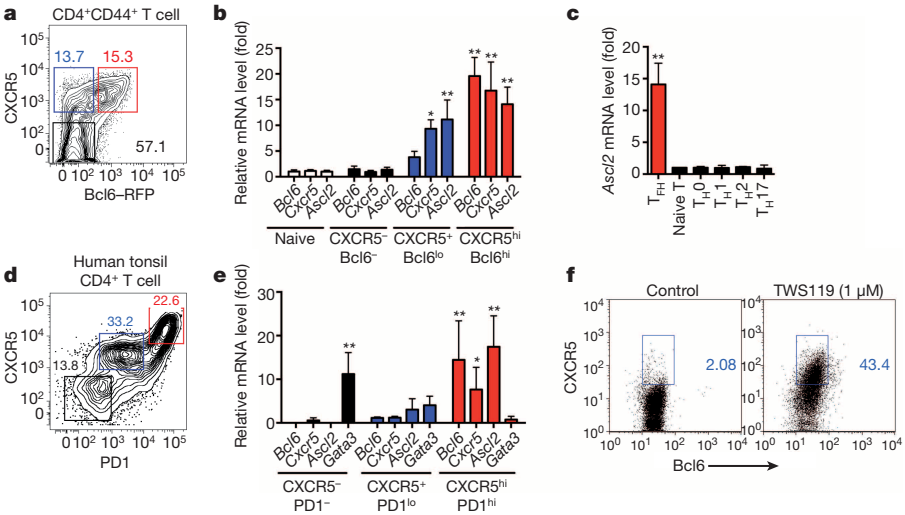


Figure 1 | Ascl2 is selectively expressed in both mouse and human T_{FH} cells. **a**, Three populations of CXCR5^{hi}Bcl6–RFP^{hi} (red), CXCR5^{+}Bcl6–RFP^{lo} (blue) and CXCR5^{-}Bcl6–RFP^{-} (black) cells were sorted from dLNs in *Bcl6-RFP* mice immunized with KLH emulsified in CFA subcutaneously. **b**, Ascl2, CXCR5 and Bcl6 transcriptional expression in sorted cells. **c**, *Ascl2* mRNA expression among *in-vivo*-generated T_{FH}, naive, T_H0, T_H1, T_H2 and T_H17 cells by quantitative polymerase chain reaction with reverse transcription (RT–PCR). **d**, Flow cytometric analysis of human tonsil CD4^{+} T cells by CXCR5 and PD1 staining. **e**, The expression of *Ascl2*, *Cxcr5*, *Bcl6* and *Gata3* mRNA in sorted cells. **f**, CXCR5 expression in CD4^{+} T cells activated by anti-CD3/anti-CD28 in the presence of TWS119 (1 μM) (an inhibitor of glycogen synthase kinase 3β (GSK-3β)) for 3 days. All experiments were repeated at least three times with similar results. **b**, **c**, **e**, Bar graphs show the relative level of mRNA as mean ± standard deviation (s.d.), $n = 3$ per group. $^{*}P < 0.05$, $^{**}P < 0.01$, two-tailed *t*-test.

IgG2a and IgG2b were not affected by Ascl2-overexpressing T cells (Fig. 2k), consistent with IgG2a switching being primarily mediated by extrafollicular T cells[18].

We next assessed whether Ascl2 could promote T-cell follicular homing *in vivo*. Ascl2-overexpressing OT-II cells preferentially accumulated in follicles (Fig. 2l), and even in the germinal centre (Fig. 2j), in comparison with control vector-infected T cells. Given that Bcl6 overexpression does not affect early T_{FH}-cell generation (Extended Data Fig. 2d, e) and T_{FH}-cell migration[3], these observations collectively demonstrate that, by contrast with Bcl6, Ascl2 promotes T-cell migration to the follicles and the initiation of T_{FH}-cell development.

To investigate the mechanism of Ascl2-controlled T_{FH}-cell generation, we performed microarray analysis, and found that the expression of 293 genes was changed by more than twofold by Ascl2 overexpression. Cross-referencing the current data set of Ascl2 versus vector with our previous T_{FH}-cell versus non-T_{FH}-cell data set[4] revealed that 85 of the 293 genes affected by Ascl2 were directly associated with T_{FH}-cell differentiation: 22 genes were upregulated and 63 genes were downregulated[4] (Fig. 3a, b and Supplementary Table 1). The chemokine receptors CXCR5 and CXCR4, which are germinal centre T_{FH}-related receptors[19], were at the top of the upregulated gene list, whereas T_H1-related genes (*Il12rb1*, *Tbx21*, *Ifng* and *Gzmb*) and the T_H17-related aryl hydrocarbon receptor (*Ahr*) gene were greatly suppressed by Ascl2 (Fig. 3a). When comparing Ascl2-RV-GFP-infected (GFP, green fluorescent protein; RV, retroviral vector) CXCR5^{+} T cells and CXCR5^{-} T cells with *in-vivo*-generated T_{FH} and non-T_{FH} cells[4], we found that Ascl2-induced CXCR5^{+} T cells were more similar in gene expression to T_{FH} cells (Extended Data Fig. 2f, g), with ~350 genes commonly expressed in these cells (Extended Data Fig. 2h).

We further examined the effect of Ascl2 on T_H1-, T_H2- and T_H17-cell differentiation. As shown in Extended Data Fig. 3a, overexpression of Ascl2 suppressed both T_H1 and T_H17 differentiation and induced CXCR5 expression. Ascl2 had no effect on TGF-β-induced Foxp3 expression but induced CXCR5^{+} regulatory T (T_{reg}) cells, suggesting that it may also be related to follicular T regulatory (T_{FR})-cell generation[20]. Under T_H2-polarized conditions, Ascl2 enhanced IL-4 expression, whereas it inhibited the expression of Gata3, IL-5 and IL-13 (Extended Data Fig. 3a–c), in agreement with recent studies that showed IL-4 but not Gata3, IL-5 or IL-13 expression in T_{FH} cells[21]. Also, we observed that Ascl2 increased IL-4 but not IL-21 production *in vivo* (Extended Data Fig. 3d, e). Therefore, Ascl2 promotes T_{FH} gene expression and inhibits T_H1-, T_H2- and T_H17-related gene expression.

We next assessed Ascl2 target genes by chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq). The analysis revealed a total of 10,028 Ascl2-binding peaks, among which 41% and 36% were enriched in intronic and intergenic regions, respectively (Fig. 3c). Only 20% of Ascl2-binding sites were located at promoter regions (Fig. 3c). Further comparison of global Ascl2-binding sites with the Ascl2-regulated gene list showed that 145 among 4,374 Ascl2-bound genes were transcriptionally regulated by Ascl2 (Fig. 3d).

As anticipated, analysis of the Ascl2-binding peaks identified an E-box protein-binding site (5′-CANNTG-3′) as the consensus motif[5] (Fig. 3e). Ascl2-binding sites were identified in groups of gene loci including receptor genes (*Cxcr5*, *Cxcr4*, *Ccr7*, *Selplg*, *Il2ra* and *Il2rb*) and inflammatory signature genes (*Ifng*, *Tbx21*, *Il2* and *Rorc*) (Fig. 3f), but not in some T_{FH}-related genes (*Bcl6*, *Prdm1*, *Pdcd1*, *Sh2d1a*, *Icos*, *Il21* and *Cd40lg*) (data not shown).

In particular, the *Cxcr5* locus was found to have multiple Ascl2-binding sites in the conserved non-coding sequence (CNS) regions (Fig. 3f, g). Moreover, these Ascl2-binding sites at the *Cxcr5* locus were confirmed in *in-vivo*-generated T_{FH} cells: two strong binding sites at intronic regions (CNS5 and CNS4-3), and three at distal promoter region (CNS1-1, CNS1 and CNS2) (Fig. 3f–h). Of note, the strongest Ascl2-binding peak was in the CNS5 region, which is consistent with the E47- (bHLH family member) binding site at the *Cxcr5* locus[22] (Fig. 3f–h), implying a potentially redundant role of E2A in transcriptional regulation of the *Cxcr5* gene[23].

To examine the functional significance of Ascl2 binding in regulating CXCR5 expression, we introduced Id3, an inhibitor of the E-box protein[22], into Ascl2-overexpressing OT-II cells by retroviral infection, and observed a substantial reduction in Ascl2-regulated CXCR5 expression (Extended Data Fig. 4a); this reduction was due to the inhibition of Ascl2 binding at the *Cxcr5* locus, as revealed by ChIP assay (Extended Data Fig. 4b). Additionally, a luciferase reporter assay showed that CNS5 and CNS1 were responsive to Ascl2 binding (Extended Data Fig. 4c, d). Together, these results provide evidence that Ascl2 is involved in the direct control of T_{FH}-cell programming.

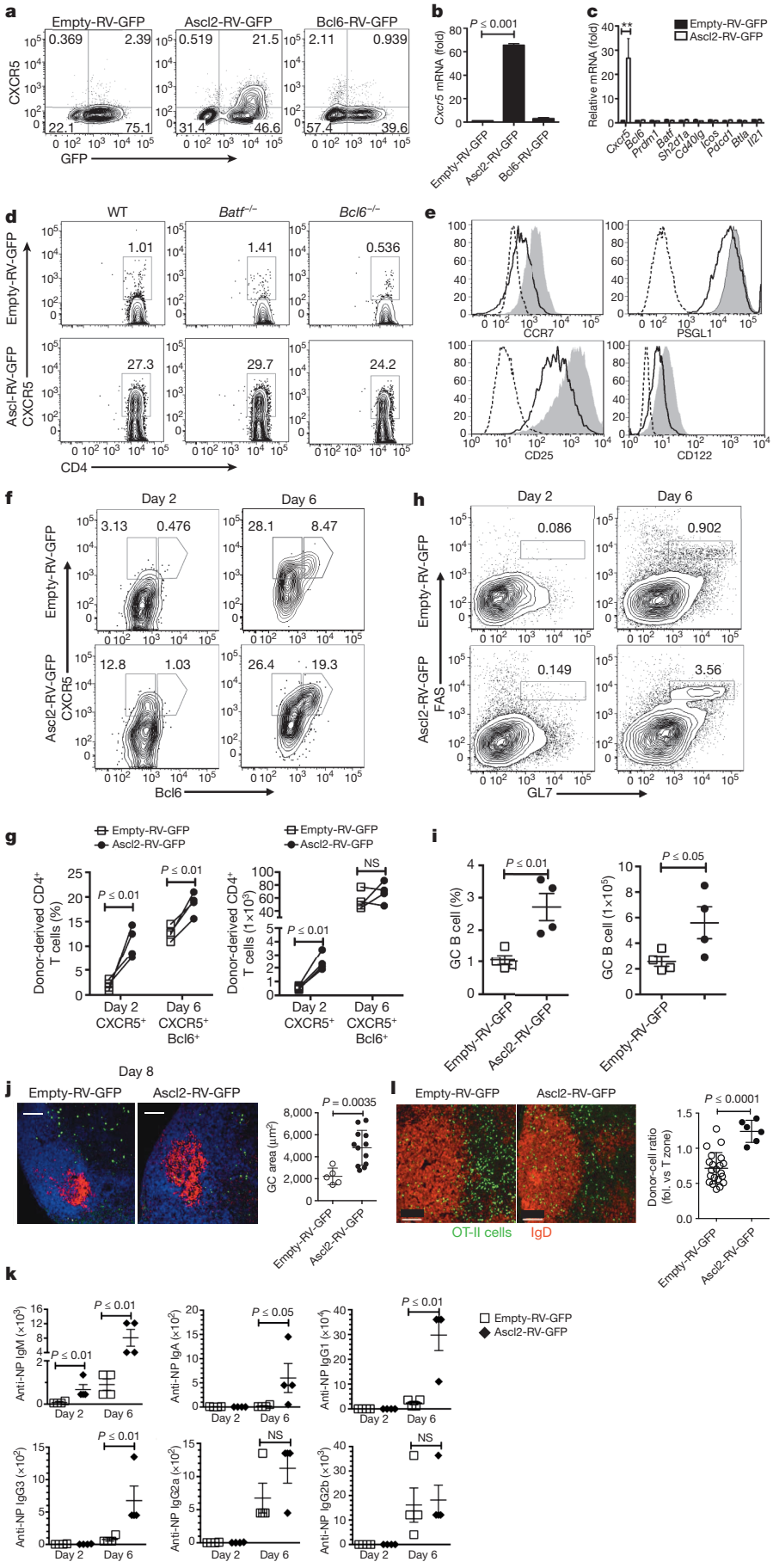
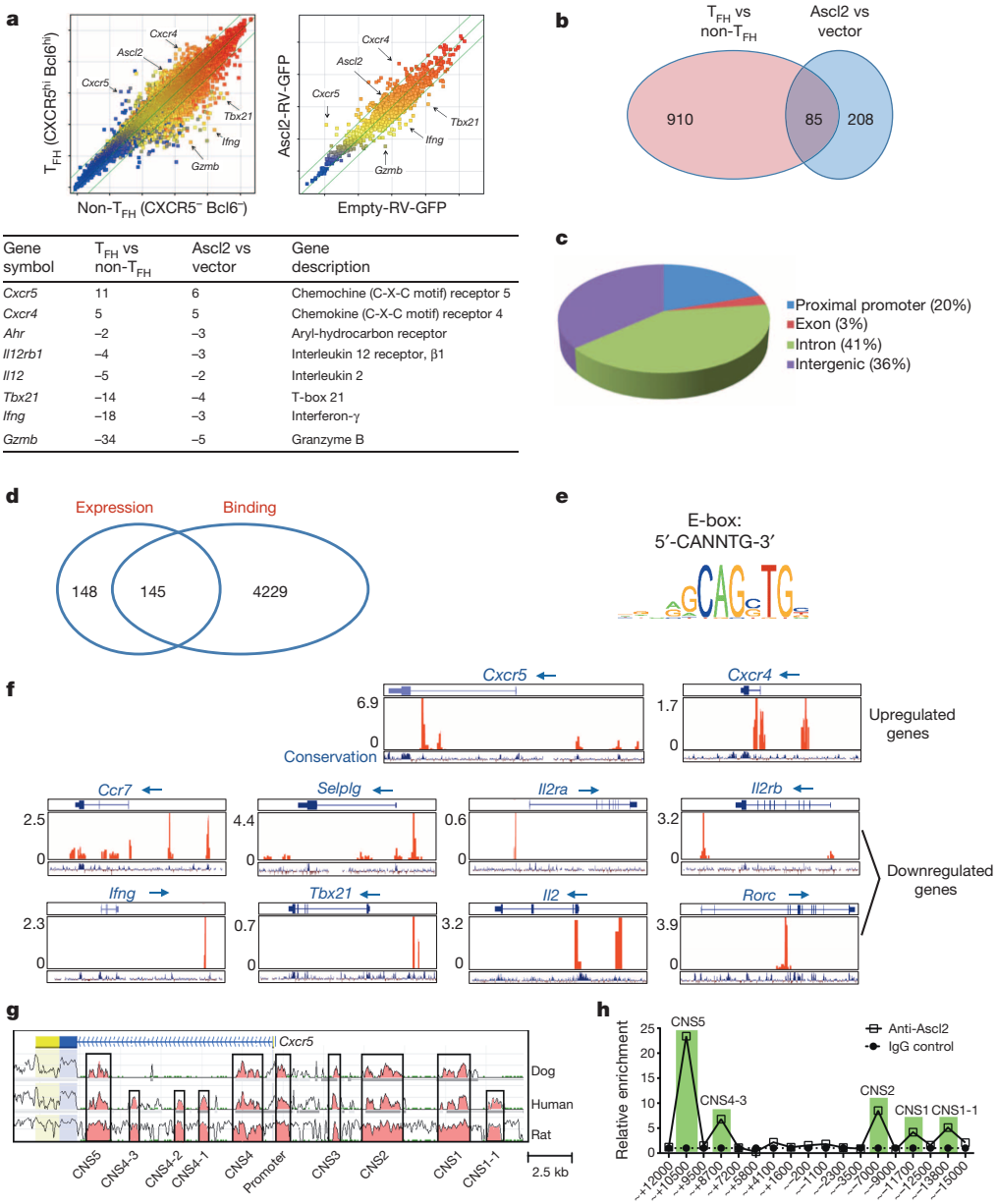


Figure 2 | Ascl2 expression induces the T_{FH} program. **a**, Flow cytometry analysis of surface CXCR5 expression from empty-RV-GFP, Ascl2-RV-GFP and Bcl6-RV-GFP retrovirus-infected T cells, respectively. **b**, Sorted GFP^+ T cells were subjected to the measurement of *Cxcr5* mRNA by quantitative RT–PCR. **c**, Measurement of gene expression, including *Cxcr5*, *Bcl6*, *Prdm1*, *Batf*, *Sh2d1a*, *Cd40lg*, *Icos*, *Pdcd1*, *Btla* and *Il21*. **d**, Surface CXCR5 expression in Ascl2-RV-GFP- and vector-virus-infected wild-type (WT), $Bcl6^{-/-}$ and $Batf^{-/-}$ T cells. **e**, CCR7, PSGL1, CD25 and CD122 expression by flow cytometry analysis. Dashed line, isotype; solid line, Ascl2-RV-GFP; shaded area, empty-RV-GFP. **f**–**i**, **k**, Ascl2-RV-GFP- or empty-RV-GFP-transduced GFP^+ OT-II cells were transferred into naive $Tcrb^{-/-}$ mice subsequently immunized with NP-OVA/CFA. **f**, At day 2 and day 6, flow cytometry analysis of donor cells was undertaken with staining for CXCR5 and Bcl6 ($n = 4$). **g**, Quantification of $CXCR5^+$ and $CXCR5^+Bcl6^+$ donor-derived T cells. **h**, Germinal centre B cells ($GL7^{hi}FAS^{hi}$) in recipient mice ($n = 4$). **i**, Quantification of germinal centre (GC) B cells. **j**, Ascl2-RV-GFP- or empty-RV-GFP-transduced GFP^+ OT-II were transferred into congenic mice. Four days later, each mouse was immunized subcutaneously with OVA (30 μg)/Alum/LPS. At day 8, dLNs were collected and subjected to histochemistry analysis. Green, GFP; red, Bcl6; blue, anti-IgD; scale bar, 100 μm, dot graph represents GC areas per slide view (some slides in Ascl2 group contain > 1 GC), displayed as mean ± s.d., control group ($n = 5$), Ascl2-RV-GFP group ($n = 12$). **k**, Titres of NP-specific antibodies in serum from mice on day 8 after immunization ($n = 4$). **l**, Distribution of Ascl2-RV-GFP- and vector-infected GFP^+ OT-II donor cells in IgD^+ B-cell follicles from dLNs in mice immunized with OVA/Alum/LPS for 4 days. Scale bar, 100 μm. Dot graph represents distribution with the ratio of donor cells in the B-cell follicle (fol.) versus the T zone, displayed as mean ± s.d. Empty-RV-GFP, $n = 21$; Ascl2-RV-GFP, $n = 15$. All experiments were repeated at least two times with similar results. **b**, **c**, **g**, **i**, **j**–**l**, Graphs show data as mean ± s.d., two-tailed *t*-test. NS, not significant.



Gene symbol	T_{FH} vs non-T_{FH}	Ascl2 vs vector	Gene description
Cxcr5	11	6	Chemochine (C-X-C motif) receptor 5
Cxcr4	5	5	Chemokine (C-X-C motif) receptor 4
Ahr	–2	–3	Aryl-hydrocarbon receptor
Il12rb1	–4	–3	Interleukin 12 receptor, β1
Il12	–5	–2	Interleukin 2
Tbx21	–14	–4	T-box 21
Ifng	–18	–3	Interferon-γ
Gzmb	–34	–5	Granzyme B

Figure 3 | Ascl2-dependent transcriptional regulation of T_{FH}-related genes. **a**, Scatterplot of the average signal of T_{FH} versus non-T_{FH} cells, and Ascl2-RV-GFP$^+$ versus empty-RV-GFP$^+$ T-cell gene expression microarray data. The green line indicates gene expression change by a factor of 2; data normalized from two replicates are shown; $n = 2$. Genes with the most transcriptional changes are listed. **b**, Venn diagram of genes regulated by Ascl2 and T_{FH}-related genes. **c**, Distribution of Ascl2 ChIP-seq peaks in Ascl2-overexpressed CD4$^+$ T cells. **d**, Venn diagram of genes regulated by Ascl2 and Ascl2-bound genes. **e**, The Ascl2-binding site is identical to the E-box-binding motif. **f**, Ascl2-binding peaks located at gene loci including *Cxcr5*, *Cxcr4*, *Ccr7*, *Selplg1*, *Il2ra*, *Il2rb*, *Ifng*, *Tbx21*, *Il2* and *Rorc*. **g**, Comparison of the mouse *Cxcr5* genomic sequence with that of dog, human and rat. Red regions denote CNS region. **h**, Ascl2-binding sites at the *Cxcr5* locus in *in-vivo*-generated T_{FH} cells as determined by ChIP-qPCR (primers are listed in Supplementary Table 2).

As Maf[10], Batf[11,12] and IRF4 (ref. 13) are required in T_{FH}-cell differentiation, we compared genome-wide occupancy of these transcriptional factors, and found that Ascl2-bound genes hardly correlated with those bound by Maf (Extended Data Fig. 5a). For instance, IL-21 is directly regulated by Maf[24], but not by Ascl2 (Extended Data Fig. 5b). Additionally, there was no binding site for Maf at the *Ascl2* locus (data not shown), or vice versa (Extended Data Fig. 5b), suggesting that Ascl2 and Maf are functionally independent in T_{FH} cells. A large proportion of Ascl2 occupancy colocalized with Batf/IRF4-binding sites (Extended Data Fig. 5c), including at the CNS of gene loci including *Cxcr5*, *Cxcr4*, *Ccr7*, *Selplg*, *Il2*, *Il2ra* and *Il2rb* (Extended Data Fig. 5d), but not at *Bcl6*, *Prdm1* and *Maf* (Extended Data Fig. 5b). Interestingly, Batf/IRF4 and Ascl2 peaks were found to be independently localized in T_H1, T_H2 and T_H17 signature genes (Extended Data Fig. 5e–g), indicating that Batf/IRF4 acts downstream of the T-cell antigen receptor (TCR) in activation of effector programs[25], whereas Ascl2 mediates suppression. These data strongly support Ascl2 acting as a specific regulator in T_{FH} cells.

T_{FH} cells provide important help to B cells in the induction of efficient anti-virus antibodies during viral infection[6,26]. To address the functional roles of Ascl2, we generated *Ascl2*$^{fl/fl}$/CD4-Cre mice, in which T cells were developmentally intact (data not shown), and assessed the requirement of Ascl2 in T_{FH}-cell development *in vivo* with influenza virus infection. After intranasal infection, *Ascl2*$^{fl/fl}$/CD4-Cre mice developed gradual body weight loss from day 3 to day 9, whereas control littermate mice recovered after day 8 (Extended Data Fig. 6a). At day 9 post-infection (d.p.i.), viral haemagglutinin (HA) mRNA expression in the lungs of *Ascl2*$^{fl/fl}$/CD4-Cre mice was over fivefold higher than that in control mice (Extended Data Fig. 6b), whereas CXCR5$^+$Bcl6$^+$ T_{FH} cells in lung draining lymph nodes (dLNs) had

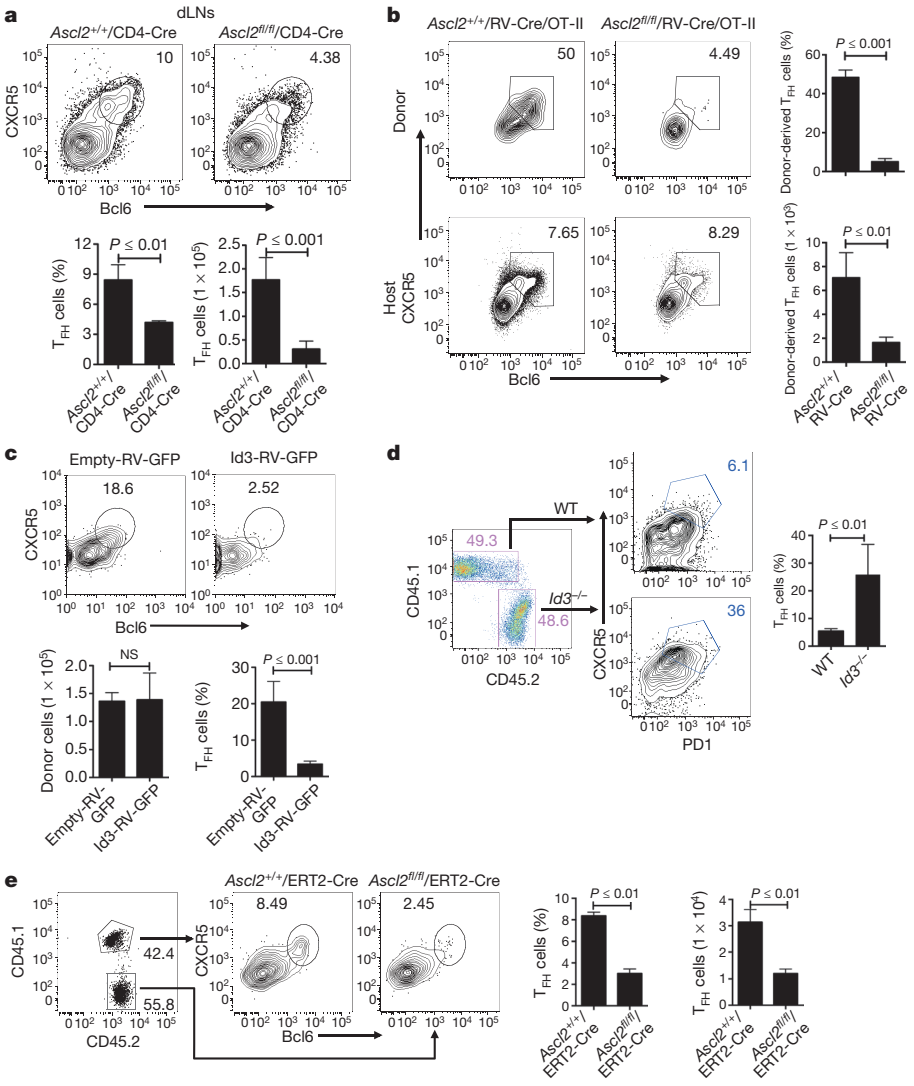


Figure 4 | Loss of Ascl2 in peripheral CD4$^+$ T cells inhibits T_{FH} differentiation. **a**, T_{FH} cells ($CXCR5^{hi}Bcl6^{hi}$) in mediastinal LNs from $Ascl2^{fl/fl}$/CD4-Cre and $Ascl2^{+/+}$/CD4-Cre mice that had received a sublethal dose of influenza virus A/Puerto Rico/8 (H1N1) intranasally 9 days before ($n = 5$). **b**, Flow cytometry analysis of donor cells or host cells in mice receiving Cre-RV-GFP retrovirus-infected $Ascl2^{fl/fl}$/OT-II or $Ascl2^{+/+}$/OT-II GFP$^+$ cells and subcutaneous 5-day immunization with OVA/CFA ($n = 3$). **c**, Flow cytometry analysis of donor-derived T_{FH} cells in mice that had received Id3-RV-GFP or control viral-vector-transduced OT-II cells and subcutaneous OVA/CFA immunization ($n = 3$). **d**, Flow cytometry analysis of donor-derived T_{FH} cells in $Rag^{-/-}$ mice receiving equal numbers of naive $Id3^{-/-}$ and wild-type (WT) CD4$^+$ T cells and subcutaneous KLH/CFA immunization for 7 days ($n = 5$). **e**, Analysis and quantification of T_{FH} cells in chimaeric mice reconstituted with equal numbers of $Ascl2^{fl/fl}$/ERT2-Cre and $Ascl2^{+/+}$/ERT2-Cre bone marrow cells after 5-day tamoxifen treatment and 7-day KLH/CFA immunization ($n = 4$). All data are representative of three independent experiments. Bar graphs show mean ± s.d., two-tailed *t*-test. NS, not significant.

decreased by over 50%, there was a ~3–4-fold decrease in total cell numbers in the absence of Ascl2 (Fig. 4a), probably accounting for a severe reduction in germinal centre B-cell formation and anti-viral IgG production (Extended Data Fig. 6c, d). Also, T_{FH}-cell and germinal centre B-cell development in the spleen was reduced in $Ascl2^{fl/fl}$/CD4-Cre mice (Extended Data Fig. 6e, f). No difference was found between CD8$^+$ T-cell populations in the lung, bronchoalveolar lavage fluid (BALF), spleen and dLNs in $Ascl2^{fl/fl}$/CD4-Cre mice and control mice (Extended Data Fig. 6g), although CD8$^+$ T-cell activation in the dLNs of $Ascl2^{fl/fl}$/CD4-Cre mice was slightly enhanced (Extended Data Fig. 6h). Moreover, we applied the same assay to tamoxifen-treated mixed bone marrow chimaeric mice receiving both $Ascl2^{fl/fl}$/ERT2-Cre and $Ascl2^{+/+}$/ERT2-Cre bone marrow cells, and found a ~50% decrease in T_{FH} cells in the $Ascl2^{fl/fl}$/ERT2-Cre counterpart, whereas no defect was found in Ascl2-ablated CD8$^+$ T cells (Extended Data Fig. 6i, j).

Because Ascl2 can form heterodimers with three other bHLH family members, including E2-2 (also known as TCF4), E47 (also known as TCF3) and HEB (also known as TCF12) in human cells[5,23], this suggests that the partial defect in T_{FH}-cell differentiation in $Ascl2^{fl/fl}$/CD4-Cre mice might be caused by compensation from other bHLH members. Indeed, T_{FH} cells and germinal centre responses were normal in $Ascl2^{fl/fl}$/CD4-Cre mice after immunization with KLH in CFA (data not shown). A substantial enhancement of E47 expression was noticed in naive T, T_{FH} and even non-T_{FH} cells from $Ascl2^{fl/fl}$/CD4-Cre mice compared with control mice (Extended Data Fig. 7a), and expression of E47 also increased CXCR5 expression in CD4$^+$ T cells (Extended Data Fig. 7b), consistent with a recent report that increased E47 activity in $Id3^{-/-}$ mice is associated with increased T_{FH}-like cells[22]. Furthermore, we examined T-cell maturation and T_{FH}-cell differentiation in chimaeric mice receiving both $Ascl2^{fl/fl}$/CD4-Cre and $Ascl2^{+/+}$/CD4-Cre, or only $Ascl2^{fl/fl}$/CD4-Cre bone marrow cells. As shown in Extended Data Fig. 7c, d, in chimaeric mice, Ascl2 deficiency reduced TCR-$\beta^{hi}CD69^{lo}$ mature T cells, whereas mice receiving only $Ascl2^{fl/fl}$/CD4-Cre cells had fewer defects in T-cell maturation. Accordingly, after immunization with KLH/CFA, T_{FH}-cell generation was inhibited in the $Ascl2^{fl/fl}$/CD4-Cre counterpart from mixed chimaeric mice (Extended Data Fig. 7e–g). These data suggest that Ascl2 deletion at an early developmental stage induces compensatory mechanisms to allow T-cell maturation.

To overcome the inducible compensation mechanism, we deleted the *Ascl2* gene acutely using a Cre-expressing retrovirus and examined T_{FH}-cell differentiation. As shown in Fig. 4b, Ascl2 deficiency in this case resulted in an absolute impairment in T_{FH} development *in vivo*. By contrast, Bcl6-deficient OT-II cells showed intact CXCR5 expression and homing ability at day 3 after immunization (Extended Data Fig. 8a–e). Moreover, Id3 overexpression impaired T_{FH}-cell generation and germinal centre responses (Fig. 4c and Extended Data Fig. 9a–c). Conversely, Id3 deficiency enhanced the T_{FH}-cell population (Fig. 4d). Together with the observation that Bcl6 overexpression could not rescue Id3-induced T_{FH} blockage (Extended Data Fig. 9d), Ascl2 seems to have an earlier function than Bcl6 in T_{FH}-cell development.

To confirm this, we examined T_{FH}-cell differentiation in either mixed chimaeric mice receiving equal numbers of *Ascl2*$^{fl/fl}$/ETR2-Cre and *Ascl2*$^{+/+}$/ETR2-Cre bone marrow cells (Fig. 4e), or intact *Ascl2*$^{fl/fl}$/ETR2-Cre mice (data not shown). After tamoxifen treatment and KLH/CFA immunization, Ascl2 deletion inhibited T_{FH}-cell differentiation and germinal centre response *in vivo* (Fig. 4e and data not shown). Therefore, these data verify that Ascl2 is intrinsically necessary for T_{FH}-cell differentiation.

We have identified a new player—the Ascl2 transcription factor—that is crucial for T_{FH}-cell development and function. On the one hand, similar to Bcl6, Ascl2 acts as a novel suppressor of T_H1-, T_H2- and T_H17-cell differentiation. On the other hand, Ascl2 uniquely regulates T_{FH}-cell migration and development by increasing CXCR5 and CXCR4 expression, and suppressing CCR7 and PSGL1 expression, and IL-2 signalling. Our data indicate that Ascl2 and Bcl6 act to program T_{FH}-cell generation. Activated T cells by antigen-presenting cells gain Ascl2 expression, which allows their migration towards B-cell follicles. At the T–B border, cognate B cells provide another signal for precursor T_{FH} cells to increase Bcl6 expression, which completes T_{FH} polarization and germinal centre formation. Therefore, Id3 and Ascl2 may serve as an early checkpoint during T_{FH}-cell development in promoting appropriate antibody responses to infection while keeping autoimmune diseases in check (Extended Data Fig. 10). Further investigation into this axis may offer new ways to modulate antibody responses in infection and autoimmunity.

METHODS SUMMARY

Mice, immunization and infection. *Ascl2*$^{fl/fl}$ mice were generated previously[5] and had been backcrossed with C57BL/6 for at least six generations. Mice were immunized with antigens emulsified in CFA subcutaneously (100 μl each mouse). Influenza virus A/Puerto Rico/8 (PR8, H1N1) was purchased from Charles River Laboratories. Mice were anaesthetized by intraperitoneal injection with ketamine and infected intranasally with a dose of PR8 influenza virus[27] lethal to 50% of animals tested (LD_{50}) of 0.5.

ChIP-seq. ChIP-seq was performed as described previously[28]. The DNA fragments were sequenced with an Illumina 1G Analyzer at the Institute for System Biology. The output of the Solexa Analysis Pipeline was converted to browser-extensible data (BED) files, and the data were viewed in the UCSC genome browser.

Statistics. Unless specifically indicated, comparison between different groups was conducted with Student's *t*-tests. All *P* values below 0.05 were considered significant. Statistical analysis was performed with Graphpad Prism 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 9 April; accepted 26 November 2013.

Published online 19 January; corrected online 26 March 2014 (see full-text HTML version for details).

1. Liu, X., Nurieva, R. I. & Dong, C. Transcriptional regulation of follicular T-helper (Tfh) cells. *Immunol. Rev.* **252,** 139–145 (2013).
2. Vinuesa, C. G. & Cyster, J. G. How T cells earn the follicular rite of passage. *Immunity* **35,** 671–680 (2011).
3. Xu, H. *et al.* Follicular T-helper cell recruitment governed by bystander B cells and ICOS-driven motility. *Nature* **496,** 523–527 (2013).
4. Liu, X. *et al.* Bcl6 expression specifies the T follicular helper cell program *in vivo*. *J. Exp. Med.* **209,** 1841–1852 (2012).
5. van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136,** 903–912 (2009).
6. Crotty, S. Follicular helper CD4 T cells (T_{FH}). *Annu. Rev. Immunol.* **29,** 621–663 (2011).
7. Johnston, R. J. *et al.* Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325,** 1006–1010 (2009).
8. Nurieva, R. I. *et al.* Bcl6 mediates the development of T follicular helper cells. *Science* **325,** 1001–1005 (2009).
9. Yu, D. *et al.* The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity* **31,** 457–468 (2009).
10. Bauquet, A. T. *et al.* The costimulatory molecule ICOS regulates the expression of c-Maf and IL-21 in the development of follicular T helper cells and T_H-17 cells. *Nature Immunol.* **10,** 167–175 (2009).
11. Betz, B. C. *et al.* Batf coordinates multiple aspects of B and T cell function required for normal antibody responses. *J. Exp. Med.* **207,** 933–942 (2010).
12. Ise, W. *et al.* The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nature Immunol.* **12,** 536–543 (2011).
13. Kwon, H. *et al.* Analysis of interleukin-21-induced *Prdm1* gene regulation reveals functional cooperation of STAT3 and IRF4 transcription factors. *Immunity* **31,** 941–952 (2009).
14. Johnston, R. J., Choi, Y. S., Diamond, J. A., Yang, J. A. & Crotty, S. STAT5 is a potent negative regulator of T_{FH} cell differentiation. *J. Exp. Med.* **209,** 243–250 (2012).
15. Nurieva, R. I. *et al.* STAT5 protein negatively regulates T follicular helper (Tfh) cell generation and function. *J. Biol. Chem.* **287,** 11234–11239 (2012).
16. Gattinoni, L. *et al.* Wnt signaling arrests effector T cell differentiation and generates $CD8^+$ memory stem cells. *Nature Med.* **15,** 808–813 (2009).
17. Ballesteros-Tato, A. *et al.* Interleukin-2 inhibits germinal center formation by limiting T follicular helper cell differentiation. *Immunity* **36,** 847–856 (2012).
18. Lee, S. K. *et al.* B cell priming for extrafollicular antibody responses requires Bcl-6 expression by T cells. *J. Exp. Med.* **208,** 1377–1388 (2011).
19. Allen, C. D. *et al.* Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nature Immunol.* **5,** 943–952 (2004).
20. Chung, Y. *et al.* Follicular regulatory T cells expressing Foxp3 and Bcl-6 suppress germinal center reactions. *Nature Med.* **17,** 983–988 (2011).
21. Liang, H. E. *et al.* Divergent expression patterns of IL-4 and IL-13 define unique functions in allergic immunity. *Nature Immunol.* **13,** 58–66 (2012).
22. Miyazaki, M. *et al.* The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nature Immunol.* **12,** 992–1001 (2011).
23. Murre, C. *et al.* Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58,** 537–544 (1989).
24. Hiramatsu, Y. *et al.* c-Maf activates the promoter and enhancer of the IL-21 gene, and TGF-β inhibits c-Maf-induced IL-21 production in $CD4^+$ T cells. *J. Leukoc. Biol.* **87,** 703–712 (2010).
25. Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151,** 289–303 (2012).
26. Yoo, J. K., Fish, E. N. & Braciale, T. J. LAPCs promote follicular helper T cell differentiation of Ag-primed $CD4^+$ T cells during respiratory virus infection. *J. Exp. Med.* **209,** 1853–1867 (2012).
27. Zhang, Y. *et al.* MKP-1 is necessary for T cell activation and function. *J. Biol. Chem.* **284,** 30815–30824 (2009).
28. Wei, G. *et al.* Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating $CD4^+$ T cells. *Immunity* **30,** 155–167 (2009).

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

Acknowledgements We thank J. A. Whitsett and J. P. Bridges at the University of Cincinnati for their provision of *Ascl2* conditional knockout mice, D. Yi for help with ChIP-seq and microarray analysis, R. Dalla-Favera for *Bcl6*$^{-/-}$ mice, H. Hu for histochemistry staining, and the Dong laboratory members for their help. This work was supported in part by a grant from the National Institutes of Health (NIH; AI106654 to C.D.), an intramural research program (NIDCR to W.C. and H.N.), an NIH Lymphoma SPORE (to X.L.), an Odyssey fellowship from the MD Anderson Cancer Center (to X.L. and B.Z.), Chinese Ministry of Science and Technology '973' program grants (2014CB542501 and 2012CB910402), and a National Natural Science Foundation of China grant (81361120397 to H.Q.).

Author Contributions X.L. designed and performed the experiments, and wrote the manuscript; X.C. and H.Q. performed histochemistry in Fig. 2j, l and Extended Data Fig. 9a, b, and B.Z., A.W., X.W., R.I.N., F.C. and S.S.N. contributed to the performance of the other experiments; X.Y., P.C., Q.T. and L.W. performed microarray, ChIP product sequencing and data analysis; L.G.v.d.F., H.N., W.C. and H.C. provided important mouse strains; and C.D. designed and directed the research.

Author Information Microarray and ChIP-seq data have been deposited at GenBank under accession GSE52840. 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.D. (cdonglab@hotmail.com) or X.L. (xindongliu@hotmail.com).

LETTER

doi:10.1038/nature12978

Structure-based programming of lymph-node targeting in molecular vaccines

Haipeng Liu[1,2,3], Kelly D. Moynihan[2,3], Yiran Zheng[2,3], Gregory L. Szeto[2,3], Adrienne V. Li[2,3], Bonnie Huang[2,3], Debra S. Van Egeren[4], Clara Park[2] & Darrell J. Irvine[1,2,3,5,6]

In cancer patients, visual identification of sentinel lymph nodes (LNs) is achieved by the injection of dyes that bind avidly to endogenous albumin, targeting these compounds to LNs, where they are efficiently filtered by resident phagocytes[1,2]. Here we translate this 'albumin hitchhiking' approach to molecular vaccines, through the synthesis of amphiphiles (amph-vaccines) comprising an antigen or adjuvant cargo linked to a lipophilic albumin-binding tail by a solubility-promoting polar polymer chain. Administration of structurally optimized CpG-DNA/peptide amph-vaccines in mice resulted in marked increases in LN accumulation and decreased systemic dissemination relative to their parent compounds, leading to 30-fold increases in T-cell priming and enhanced anti-tumour efficacy while greatly reducing systemic toxicity. Amph-vaccines provide a simple, broadly applicable strategy to simultaneously increase the potency and safety of subunit vaccines.

A major challenge in the development of subunit vaccines is the efficient delivery of antigen/adjuvant to secondary lymphoid organs, where immune responses are orchestrated[3,4]. Attempts to enhance vaccine delivery have included the use of depot-forming adjuvants[5] or nanoparticulate carriers that are preferentially internalized by antigen presenting cells (APCs)[4,6–12], but approaches that could use well-defined molecular conjugates would be attractive. Antigens conjugated to antibodies targeting dendritic cells reach these cells in the draining LNs but also drain into the systemic circulation and access APCs in distal tissues[13,14], which might promote tolerance unless inflammatory adjuvants are systemically co-administered. LN targeting is also required for cancer staging in sentinel LN mapping procedures, in which radioactive or coloured dyes are injected at tumour resection sites[1]. Compounds that bind avidly to serum albumin are particularly effective LN tracers[2]; albumin binding targets these molecules to lymphatics and draining LNs, where they accumulate in APCs[15,16]. Inspired by this strategy, we set out to create LN-targeting molecular vaccines designed to similarly 'hitchhike' on albumin to LNs. Exploiting albumin's role as a fatty acid transporter, we proposed that antigens/adjuvants modified with a lipophilic albumin-binding domain would accumulate in lymphoid organs after injection, through *in situ* complexation and transport with endogenous albumin. To develop this strategy, we studied model vaccines composed of peptide antigens and CpG DNAs, single-stranded oligonucleotides containing unmethylated cytosine-guanine motifs that bind Toll-like receptor 9 (TLR9) and serve as potent molecular adjuvants[17,18].

To identify an optimal albumin-binding domain that could be appended to either CpG or peptide antigens, we constructed a series of amphiphilic 20-base phosphorothioate (PS)-stabilized CpG oligonucleotides 5′-linked to lipophilic tails (amph-CpGs; Fig. 1a; for synthesis details see Supplementary Information). We first evaluated the interaction of fluorescein amidite (FAM)-labelled conjugates with serum proteins by size-exclusion chromatography (SEC; Fig. 1b). Fetal bovine serum (FBS) exhibited a major protein fraction eluting at 5.3 min in SEC that coincided with albumin (Extended Data Fig. 1a). The vast majority of mono-acyl-conjugated (C18-CpG) or cholesterol-conjugated (Cho-CpG) oligonucleotides eluted as monomers at 5.8 min in the presence or absence of serum, indicating a lack of interaction with albumin (Fig. 1b and Extended Data Fig. 1a). By contrast, diacyl lipid-conjugated CpGs (lipo-CpGs) in aqueous solution eluted as micelles (3.7 min), but, after incubation with serum, nearly 50% of the lipo-CpGs co-migrated with albumin (Fig. 1b). Biotinylated lipo-CpG (but not CpG) incubated with FBS and then captured with magnetic beads was found to pull down albumin, and lipo-CpG was efficiently captured by albumin-conjugated agarose (Extended Data Fig. 1b, c). Biolayer interferometry and spectroscopy measurements of fluorescence resonance energy transfer (FRET) between lipo-CpG and purified albumin further confirmed their molecular association in solution (Extended Data Fig. 1d, e).

We next characterized the *in vivo* trafficking of CpG conjugates. Amph-CpGs were injected subcutaneously in C57BL/6 mice, and 24 h later, draining LNs were excised intact for *in vivo* imaging system (IVIS) fluorescence imaging. C18- and Cho-CpG showed marginally increased uptake in LNs relative to unmodified CpG, reaching levels similar to CpG delivered by two prototypical vaccine vehicles known to enhance vaccine accumulation in LNs, incomplete Freund's adjuvant (IFA) or poly(ethylene glycol) (PEG)-coated liposomes[6] (Fig. 1c). By contrast, lipo-CpG accumulated eightfold more than soluble CpG (Fig. 1c). Over 7 days after injection, soluble CpG exhibited no LN accumulation above 0.3% of the injected dose at any time, whereas lipo-CpG accumulated for 3 days, giving a 12-fold greater area under the curve (AUC) for total LN exposure to CpG (Fig. 1d). LN accumulation was not dependent on TLR9-recognized CpG motifs, and was not due to increased nuclease resistance of lipid-modified PS-backbone CpGs (data not shown).

Our *in vitro* analysis indicated that lipo-CpG molecules equilibrated between micellar and albumin-bound forms in the presence of serum, making it unclear which state was responsible for LN targeting. However, injection of lipo-CpG pre-incubated for 5 h with freshly isolated 90 vol% mouse serum at 37 °C, conditions in which SEC showed the vast majority of the amphiphile co-migrated with albumin, led to essentially identical LN targeting compared with direct lipo-CpG injections (data not shown). To address this question further, we introduced poly-guanine repeats between the diacyl lipid and CpG sequence to lock the amphiphiles in the micellar state and block disassembly by albumin. G-quadruplex hydrogen bonding between adjacent oligonucleotide strands in lipo-G_n-CpG micelles containing four or more guanine repeats rendered the micelles stable in the presence of serum (Extended Data Fig. 2a–d). Labelled albumin did not co-migrate with stabilized micelles, as assessed by SEC (Extended Data Fig. 2e), suggesting that the amphiphiles do not interact with albumin as intact micelles. Despite forming micelles with similar sizes (Extended Data Fig. 2f), G-quartet-stabilized

[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard, Cambridge, Massachusetts 02139, USA. [6]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.

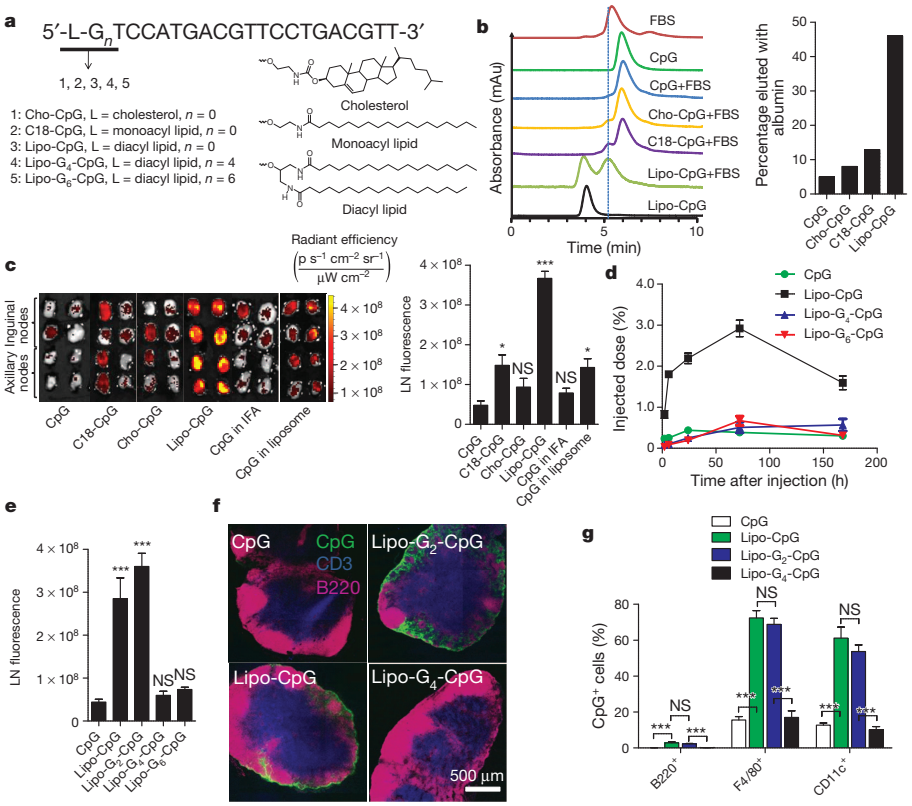


Figure 1 | Design of an LN-targeted molecular adjuvant. **a**, Structure of amph-CpGs. **b**, SEC of FAM-CpGs alone or after incubation with FBS for 2 h (left), and percentage CpG co-migrating with albumin peaks (right). Vertical dashed line provides a guide to the eye. **c**–**g**, IVIS fluorescence imaging of excised draining LNs from C57BL/6 mice ($n = 4$ LNs per group) injected with FAM-CpGs (3.3 nmol) in soluble form, emulsified in IFA, entrapped in liposomes, or as amphiphile conjugates. **c**, IVIS images and quantification from inguinal and axillary nodes at 24 h. **d**, CpG accumulation in draining LNs. **e**, IVIS quantification of CpG in LNs 24 h after injection of G-quadruplex-forming Lipo-G$_n$-CpGs. **f**, Immunohistochemistry of inguinal LNs 24 h after injection (CD3, blue; B220, pink; CpG, green). **g**, LN CpG$^+$ cells determined by flow cytometry at 24 h. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$ compared with soluble CpG by one-way analysis of variance (ANOVA) with Bonferroni post-test. Data represent mean ± standard error of the mean (s.e.m.) of 2–3 independent experiments. NS, not significant.

lipo-G$_4$-CpG or lipo-G$_6$-CpG micelles exhibited poor LN accumulation after injection compared with albumin-binding lipo-CpG and lipo-G$_2$-CpG (Fig. 1d, e). A possible explanation is that amplification of nonspecific matrix binding by the PS-DNA backbones[19,20] in the multivalent micellar form irreversibly trapped the stabilized micelles at the injection site. Histological sections of draining LNs showed little detectable CpG or lipo-G$_4$-CpG, whereas albumin-binding lipo-CpG and lipo-G$_2$-CpG accumulated in the subcapsular sinus and interfollicular areas (Fig. 1f). LN-accumulating amphiphiles were associated with F4/80$^+$ macrophages and CD11c$^+$ dendritic cells, with only a minor contribution from skin-derived CD207$^+$ dendritic cells (Fig. 1g and Extended Data Fig. 3). If albumin hitchhiking mediates LN targeting, then covalent conjugation of oligonucleotides to albumin should also enhance LN accumulation. We found that injection of CpG conjugated to mouse serum albumin (MSA) led to slightly lower oligonucleotide uptake in LNs than lipo-CpG, but much greater accumulation than soluble CpG (Extended Data Fig. 4). Altogether, these data suggest that efficient LN delivery of CpG oligonucleotides conjugated to lipophilic tails is enhanced by partitioning of oligonucleotides from micelles into a serum albumin-bound state.

To identify potential differences in the function of CpG versus optimally LN-targeted lipo-G$_2$-CpG beyond altered biodistribution, we assessed several aspects of CpG bioactivity *in vitro*: lipo-G$_2$-CpG was internalized by dendritic cells into endolysosomes in a pattern indistinguishable from CpG in confocal microscopy, albeit to twofold higher levels (Extended Data Fig. 5a, b). Lipo-G$_2$-CpG did not activate the lipid-binding receptor TLR2 in reporter cells, but both free CpG and amph-CpG activated RAW macrophages bearing a NF-κB reporter in a CpG sequence-specific manner (Extended Data Fig. 5c, d). When dendritic cells were activated with CpG or lipo-G$_2$-CpG and pulsed with OVA protein to test cross-presentation of antigen to OT-I (OVA-specific) T-cells, dendritic cell activation by soluble or amphiphile-CpG led to similar T-cell proliferation (Extended Data Fig. 5e). Thus, lipid modification of CpG increased uptake in the presence of serum but did not otherwise greatly alter the bioactivity of CpG.

To determine the impact of LN targeting on the immune response, we immunized mice with ovalbumin protein (OVA) mixed with unmodified CpG, Cho-CpG, CpG emulsified in IFA or lipo-G$_n$-CpGs ($n = 0$, 2, 4, 6). OVA (which has only 13% sequence identity with albumin) showed minimal association with lipo-CpGs (data not shown) and thus these vaccinations assessed the impact of LN targeting of the adjuvant only, relying on normal lymphatic drainage of the OVA antigen. Lipo-G$_n$-CpG-adjuvanted vaccines primed significantly increased frequencies of antigen-specific, cytokine-producing CD8$^+$ T cells compared with unmodified CpG, Cho-CpG or CpG in IFA, but the strongest responses (up to 32-fold greater than unmodified CpG) were elicited by lipo-CpG and lipo-G$_2$-CpG (Fig. 2a, b and Extended Data Fig. 6a, b). Notably, the magnitude of the T-cell response was strongly correlated with LN accumulation of CpG (Fig. 2c). Lipo-CpG also modestly increased antibody responses by approximately threefold and enhanced CD8$^+$ T-cell responses to the model HIV antigen simian immunodeficiency virus (SIV) Gag (Extended Data Fig. 6c, d). Importantly, no antibodies against albumin were detected for any of the amph-CpG vaccines (Extended Data Fig. 6e, f and data not shown). Control immunizations with non-TLR agonist lipo-GpC or diacyl-PEG conjugates (lipo-PEG) mixed with OVA were ineffective and amph-CpG responses were identical in *Tlr2*$^{-/-}$ mice (Extended Data Fig. 6g, h), ruling out a direct adjuvant effect of the diacyl lipid tail. CpG that is not captured in local LNs drains to the systemic circulation, leading to systemic inflammatory toxicities[21]. Despite lymphadenopathy of draining LNs, indicating local activity, repeated injections of lipo-CpGs showed greatly reduced systemic inflammation relative to free CpG (Fig. 2d, e and Extended Data Fig. 7a, b). Although further work will be needed to determine any potential autoimmune toxicities related to LN targeting of CpG[18], these results suggest that the LN targeting achieved by amph-CpGs greatly enhances the potency of this molecular adjuvant while simultaneously lowering acute systemic side effects.

Synthesis of lipo-CpG is straightforward owing to the solubility imparted by the long polar oligonucleotide block, but, depending on the amino acid sequence, lipidated polypeptides can be essentially insoluble[22].

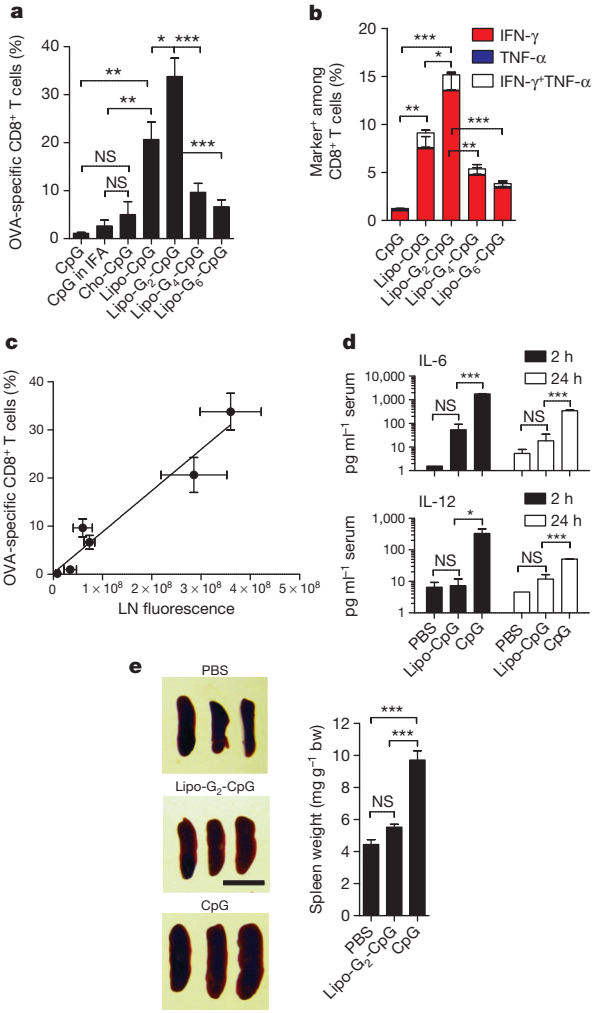


Figure 2 | LN targeting enhances potency while reducing systemic toxicity of CpG. **a**–**c**, C57BL/6 mice ($n = 4$–8 per group) were immunized with OVA (10 μg) plus CpG (1.24 nmol) on day 0 and day 14; shown are SIINFEKL tetramer (**a**) and intracellular cytokine staining (**b**) on peripheral blood at day 20. **c**, LN CpG fluorescence correlation with T-cell response. **d**, Serum cytokines after injection ($n = 3$ per group) of 6.2 nmol CpG. **e**, Splenomegaly ($n = 3$ per group) assessed on day 6 after three injections of CpG. bw, body weight. Scale bar, 1 cm. $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$ by one-way ANOVA with Bonferroni post-test. Data show mean ± s.e.m. of 2–4 independent experiments. NS, not significant.

To generalize this LN targeting strategy to antigens and other potential vaccine components, we synthesized peptides linked to a diacyl lipid tail via a PEG block chosen to promote conjugate solubility (amph-peptides; Fig. 3a). Amph-peptides and lipo-PEGs in water form micelles, but these amphiphiles can also insert their diacyl tails into cell membranes. We found that lipo-PEG amphiphiles with short PEG blocks exhibited preferential plasma membrane insertion when incubated with cells in the presence of albumin *in vitro* (Fig. 3b and Extended Data Fig. 8a, b), which might limit transit to LNs on albumin *in vivo*. However, increasing the polar block to 48 ethylene glycol units yielded amphiphiles that partitioned preferentially into solution while retaining albumin binding (Fig. 3b and Extended Data Fig. 8c), consistent with previous studies[23]. This *in vitro* partitioning directly predicted *in vivo* draining patterns, as lipo-PEG-FAM amphiphiles injected subcutaneously showed increasing LN accumulation with increasing PEG block length (Fig. 3c). Although optimal immunostimulatory CpGs are ~20 bases, an analogous trend was observed for DNA amphiphiles as a function of oligonucleotide length (Extended Data Fig. 8d, e). Like amph-CpGs, the structure of the hydrophobic block was also important; whereas lipo-PEG amphiphiles with long diacyl tails (≥16 carbons, which exhibit a high affinity for albumin[24]) showed intense LN accumulation, shorter lipid tails with low affinity for albumin showed low LN accumulation (Fig. 3d). On the basis of these findings establishing design rules for efficient targeting of lipo-PEG amphiphiles to LNs, we conjugated peptide antigens to 1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-*N*-PEG (DSPE-PEG 2 kDa) to generate amph-peptides for vaccination studies (Extended Data Fig. 9a).

To test the potency of combined antigen and adjuvant targeting to LNs, we prepared amph-peptide (DSPE-PEG-peptide) conjugates of a model HIV antigen (AL11 epitope from SIV Gag)[25], the tumour-associated self-antigen Trp2 from melanoma[26], and a peptide derived from the human papillomavirus (HPV)-derived cervical cancer antigen E7 (ref. 27). Amph-peptides accumulated efficiently in LNs (Extended Data Fig. 9b, c). C57BL/6 mice immunized with amph-peptides and amph-CpG (lipo-G_2-CpG) showed markedly increased expansion of antigen-specific, cytokine-producing $CD8^+$ T-cells and enhanced cytolytic activity relative to unmodified peptide/CpG immunizations (Fig. 4a–c). To test whether amph-vaccine delivery enhances the protective efficacy of peptide vaccines, animals bearing established TC-1 tumours (expressing the E7 oncoprotein from HPV) or B16F10 melanomas were vaccinated. Amph-vaccines triggered sustained regression of large TC-1 tumours that were only modestly affected by soluble vaccines (Fig. 4d) and slowed the growth of melanoma tumours, in which a traditional soluble vaccine had no effect (Fig. 4e). In addition to enhancing the effectiveness of optimal T-cell epitopes, synthetic 'long peptide' antigens[28] also exhibited enhanced immunogenicity when delivered as amph-peptides (Extended Data Fig. 10a–e). This result is of particular interest because a finite pool of long sequences permits peptide vaccines to provide effective coverage of epitopes across the diverse haplotypes of a given target patient population. Amphiphile-long peptides were also approximately tenfold more potent than soluble peptides when combined with non-CpG, non-LN-targeted alternative TLR adjuvants (Extended Data Fig. 10f), showing that CpG is not required for an enhanced response to LN-targeted peptides.

Altogether, the results presented here define design rules for amphiphile conjugates as a general strategy to enhance the potency and safety

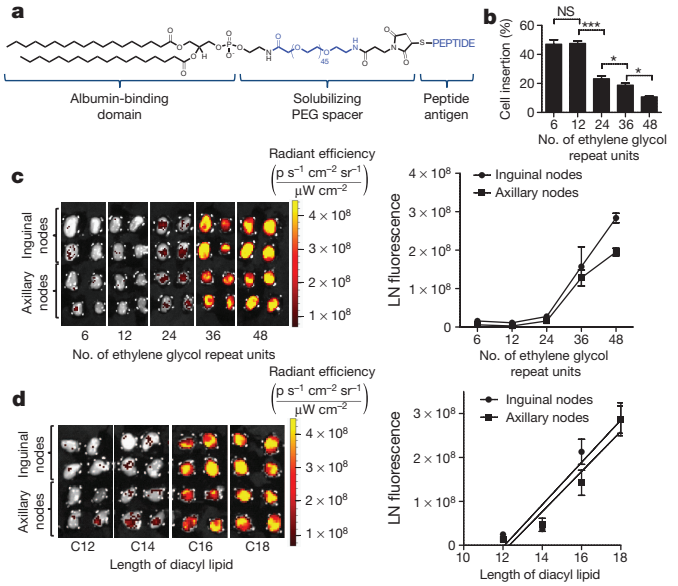


Figure 3 | Design of LN-targeted amph-peptides. **a**, Structure of amph-peptides. **b**, Amph-PEG-fluorescein insertion into cell membranes quantified by fluorescence spectroscopy after 1 h incubation with splenocytes in the presence of 100 μM albumin. **c**, **d**, C57BL/6 mice ($n = 4$ LNs per group) were injected with fluorescent amph-PEGs with varying PEG lengths (fixed C18 diacyl lipid tails; **c**) or lipid tail lengths (fixed PEG length 48 EG units; **d**); draining LNs were excised and imaged by IVIS after 24 h. $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$ by one-way ANOVA with Bonferroni post-test. Data represent mean ± s.e.m. of two independent experiments. NS, not significant.

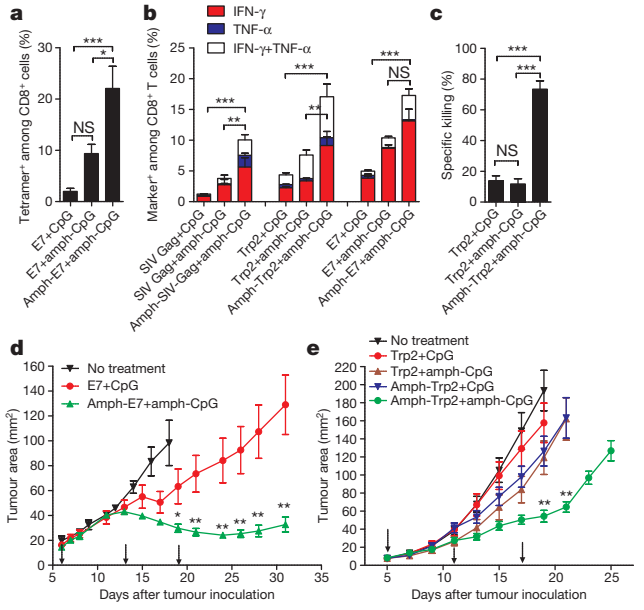


Figure 4 | Amph-vaccines maximize the immunogenicity and therapeutic efficacy of polypeptide vaccines. a–c, C57BL/6 mice ($n = 3$–4 per group) were immunized with SIV Gag, Trp2, or E7 peptides (10 μg) plus CpG (1.24 nmol) on day 0 and day 14; shown are tetramer-positive $CD8^+$ T-cells (**a**) and intracellular cytokine production (**b**) in peripheral blood on day 20. **c**, Trp2-specific cytotoxicity measured using an *in vivo* killing assay on day 21. **d**, **e**, Tumour growth in C57BL/6 mice ($n = 8$ per group) inoculated with 3×10^5 TC-1 (**d**) or B16F10 (**e**) tumour cells and vaccinated with CpG plus E7 peptide or Trp2 peptide (10 μg prime, 20 μg boost), respectively, on days indicated by arrows. Statistically significant differences between soluble and amph-vaccines are indicated by asterisks: $***P < 0.001$, $**P < 0.01$, $*P < 0.05$ by one-way ANOVA with Bonferroni post-test. Data show mean ± s.e.m. of 2–4 independent experiments. NS, not significant.

of LN-active compounds, an approach that may be applicable to a broad range of immunomodulatory therapeutics and imaging agents. These findings also have implications for how the immune system may survey lipophilic antigens. Further work will be needed to determine whether albumin binding is functionally critical, or alternatively whether other rare serum components may have a role in the observed LN targeting.

METHODS SUMMARY

Synthesis of vaccine amphiphiles. Oligonucleotide amphiphiles were synthesized using an ABI 394 synthesizer on a 1.0 μmol scale. All lipophilic phosphoramidites were conjugated as a final 'base' on the 5′ end of oligonucleotides[19]. Antigen amphiphiles were synthesized by reacting N-terminal cysteine-modified peptides with maleimide-PEG_{2000}-DSPE in dimethyl formamide.

Immunizations. Six-to-eight week old C57BL/6 mice (female, Jackson Laboratory) were immunized with 10 μg of antigen mixed with 1.24 nmol CpG adjuvant in 100 μl of PBS subcutaneously at the base of the tail. All procedures were performed under an IUCAC-approved animal protocol in accordance with the guidelines for animal care in a Massachusetts Institute of Technology animal facility inspected by the US Department of Agriculture.

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 May; accepted 30 December 2013.
Published online 16 February 2014.

1. Salhab, M., Patani, N. & Mokbel, K. Sentinel lymph node micrometastasis in human breast cancer: an update. *Surg. Oncol.* **20,** e195–e206 (2011).
2. Tsopelas, C. & Sutton, R. Why certain dyes are useful for localizing the sentinel lymph node. *J. Nucl. Med.* **43,** 1377–1382 (2002).
3. Johansen, P., Mohanan, D., Martínez-Gómez, J. M., Kündig, T. M. & Gander, B. Lymphogeographical concepts in vaccine delivery. *J. Control. Release* **148,** 56–62 (2010).
4. Moon, J. J., Huang, B. & Irvine, D. J. Engineering nano- and microparticles to tune immunity. *Adv. Mater.* **24,** 3724–3746 (2012).
5. Gupta, R. K. Aluminum compounds as vaccine adjuvants. *Adv. Drug Deliv. Rev.* **32,** 155–172 (1998).
6. Bachmann, M. F. & Jennings, G. T. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nature Rev. Immunol.* **10,** 787–796 (2010).
7. Hubbell, J. A., Thomas, S. N. & Swartz, M. A. Materials engineering for immunomodulation. *Nature* **462,** 449–460 (2009).
8. De Temmerman, M. L. *et al.* Particulate vaccines: on the quest for optimal delivery and immune response. *Drug Discov. Today* **16,** 569–582 (2011).
9. Pal, I. & Ramsey, J. D. The role of the lymphatic system in vaccine trafficking and immune response. *Adv. Drug Deliv. Rev.* **63,** 909–922 (2011).
10. Reddy, S. T. *et al.* Exploiting lymphatic transport and complement activation in nanoparticle vaccines. *Nature Biotechnol.* **25,** 1159–1164 (2007).
11. St John, A. L., Chan, C. Y., Staats, H. F., Leong, K. W. & Abraham, N. A. Synthetic mast-cell granules as adjuvants to promote and polarize immunity in lymph nodes. *Nature Mater.* **11,** 250–257 (2012).
12. Trubetskoy, V. S. & Torchilin, V. P. Use of polyoxyethylene-lipid conjugates as long-circulating carriers for delivery of therapeutic and diagnostic agents. *Adv. Drug Deliv. Rev.* **16,** 311–320 (1995).
13. Keler, T., He, L., Ramakrishna, V. & Champion, B. Antibody-targeted vaccines. *Oncogene* **26,** 3758–3767 (2007).
14. Tenbusch, M. *et al.* Immunogenicity of DNA vaccines encoding simian immunodeficiency virus antigen targeted to dendritic cells in rhesus macaques. *PLoS ONE* **7,** e39038 (2012).
15. Faries, M. B. *et al.* Active macromolecule uptake by lymph node antigen-presenting cells: a novel mechanism in determining sentinel lymph node status. *Ann. Surg. Oncol.* **7,** 98–105 (2000).
16. Schaafsma, B. E. *et al.* The clinical use of indocyanine green as a near-infrared fluorescent contrast agent for image-guided oncologic surgery. *J. Surg. Oncol.* **104,** 323–332 (2011).
17. Vollmer, J. & Krieg, A. M. Immunotherapeutic applications of CpG oligodeoxynucleotide TLR9 agonists. *Adv. Drug Deliv. Rev.* **61,** 195–204 (2009).
18. Bode, C., Zhao, G., Steinhagen, F., Kinjo, T. & Klinman, D. M. CpG DNA as a vaccine adjuvant. *Expert Rev. Vaccines* **10,** 499–511 (2011).
19. Liu, H., Kwong, B. & Irvine, D. J. Membrane anchored immunostimulatory oligonucleotides for *in vivo* cell modification and localized immunotherapy. *Angew. Chem. Int. Ed. Engl.* **50,** 7052–7055 (2011).
20. Krieg, A. M. & Stein, C. A. Phosphorothioate oligodeoxynucleotides: antisense or anti-protein? *Antisense Res. Dev.* **5,** 241 (1995).
21. Bourquin, C. *et al.* Targeting CpG oligonucleotides to the lymph node by nanoparticles elicits efficient antitumoral immunity. *J. Immunol.* **181,** 2990–2998 (2008).
22. Zeng, W., Ghosh, S., Lau, Y. F., Brown, L. E. & Jackson, D. C. Highly immunogenic and totally synthetic lipopeptides as self-adjuvanting immunocontraceptive vaccines. *J. Immunol.* **169,** 4905–4912 (2002).
23. Kastantin, M. *et al.* Thermodynamic and kinetic stability of DSPE-PEG(2000) micelles in the presence of bovine serum albumin. *J. Phys. Chem. B* **114,** 12632–12640 (2010).
24. Peters, T. *All About Albumin: Biochemistry, Genetics, and Medical Applications* (Academic, 1995).
25. Barouch, D. H. *et al.* Immunogenicity of recombinant adenovirus serotype 35 vaccine in the presence of pre-existing anti-Ad5 immunity. *J. Immunol.* **172,** 6290–6297 (2004).
26. Schreurs, M. W. *et al.* Dendritic cells break tolerance and induce protective immunity against a melanocyte differentiation antigen in an autologous melanoma model. *Cancer Res.* **60,** 6995–7001 (2000).
27. Feltkamp, M. C. *et al.* Vaccination with cytotoxic T lymphocyte epitope-containing peptide protects against a tumor induced by human papillomavirus type 16-transformed cells. *Eur. J. Immunol.* **23,** 2242–2249 (1993).
28. Kenter, G. G. *et al.* Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia. *N. Engl. J. Med.* **361,** 1838–1847 (2009).

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

Acknowledgements This work was supported in part by the Koch Institute Support (core) grant P30-CA14051 from the National Cancer Institute, the National Institutes of Health (grants AI091693, AI104715 and AI095109), the Department of Defense (W911NF-13-D-0001 and W911NF-07-D-0004, T.O. 8) and the Ragon Institute of Massachusetts General Hospital, the Massachusetts Institute of Technology and Harvard. D.J.I. is an investigator of the Howard Hughes Medical Institute. We thank T. C. Wu for kindly providing the TC-1 tumour cells. We thank the Koch Institute Swanson Biotechnology Center for technical support, specifically the applied therapeutics and whole animal imaging core facility, histology and flow cytometry core facility. The authors acknowledge the service to the MIT community of the late Sean Collier.

Author Contributions H.L. designed and performed most experiments and analysed the data, and wrote the manuscript; Y.Z. carried out tumour therapy experiments and analysed the data. K.D.M. carried out *in vitro* bioactivity studies of CpG, biolayer interferometry binding studies and *in vivo* immunizations of SIV Gag and analysed the data. A.V.L. and B.H. assisted in tetramer/*in vivo* cytotoxicity assays and contributed experimental suggestions. G.L.S. assisted with optimization of proinflamatory cytokine assays and helped *in vitro* bioactivity studies of CpG. G.L.S., C.P. and D.S.V.E. contributed to *in vitro* T-cell proliferation assays. D.J.I. supervised all experiments and wrote the manuscript.

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

CAREERS

GRANTS Success rate for UK government funding is lower for women **p.525**

SALARIES Faculty pay up more at US public universities than at private **p.525**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

SERGEV NIVENS/SHUTTERSTOCK

RESEARCH TOOLS

Jump off the page

Researchers are learning to embrace online lab notebooks, but not without growing pains.

BY AMANDA MASCARELLI

Two years into his PhD, Carl Boettiger needed a better way to organize his data and synthesize his ideas. Fishing around online, he stumbled across chemist Cameron Neylon's open electronic lab notebook. Boettiger, who was studying mathematical ecology, liked what he saw. Neylon, now advocacy director at the Public Library of Science in San Francisco, California, had pulled back the curtain on the steps and thought processes behind his protocols and research. His data collection, protocols and results were linked together and available online, making the concepts easy to reference and explore.

Inspired, Boettiger created his own electronic notebook, reporting online about his day-to-day research in a publicly available wiki that is followed by the open-science community. Viewers can find the notebook through the wiki's RSS feed and links on social media as well as in Google searches, and can post comments. He soon started to receive suggestions about, and valuable feedback on, his research and methods from other scientists — mostly followers of open science from outside his field — and some even led to collaborations.

Four years on, Boettiger, now a theoretical ecologist at the University of California, Santa Cruz, is a leader in the use of the open notebook. He is convinced of its value, and he is not alone — the idea is steadily gaining traction in some (but not all) scientific circles.

Whether on paper or in digital form, lab notebooks are meant to document exactly what, when and why experiments were done (see 'Record of achievement') and usually contain much more information than will ever be published in an academic paper. They can be used as evidence for securing patents, to settle legal issues or to pass a project from one researcher to another. Industry labs almost always require their researchers to maintain such records, as do many academic principal investigators, and until very recently the information would be kept securely under wraps in the lab until it was published.

Open electronic notebooks are a radical departure from this ethos. Data and methods are no longer cloistered in books or tucked away on private hard drives. Instead, they can be shared online for all to see. Some scientists might shudder at the thought of anyone but lab mates and close collaborators knowing the detailed logic and steps behind their research projects before publication. But open ▶

science is becoming more widely accepted as technologies change and as younger generations of researchers discover alternative tools and approaches.

EMBRACE THE OPENNESS

Boettiger readily admits that having his entire scientific process exposed online at each step of the way carries a risk. But "you have to take risks to be successful", he points out. "The idea that there's a risk-free way of getting your science out there and understood and engaging collaborators is a myth." He posts updates to his work throughout the day, as if scribbling notes in a conventional notebook, but publishes synthesized analyses and summaries several times a week or month. Boettiger still has some concerns about being scooped (see *Nature* **493,** 711; 2013), but says that those reservations are mostly offset by the many benefits of making his science open and accessible — such as providing another way to attract collaborators and to gain recognition in his field.

The gradual acceptance of open science in fields such as chemistry, mathematics, neuroscience and ecology is highlighting the information-management challenges that scientists face. "Now that science has become more complicated and not all of the details and data fit in the papers we publish, we're at a loss," says Boettiger. "How do we communicate exactly what we're doing? How do we keep science replicable?"

Scientists collect, store and analyse data in different ways and with a growing array of tools. That makes it difficult to compare one researcher's results with another's, and even harder to know how to reproduce their results, says Carly Strasser, a data-curation and open-science specialist at the California Digital Library in Oakland. "We still have paper notebooks, we still have Post-it notes, we still have phone calls that we don't transcribe and a ton of e-mails going back and forth," she says. "There are just a lot of moving parts. Scientists don't ever really learn how to capture the process."

The problem is mainly down to a lack of training. Undergraduates turn in detailed notebooks for relatively simple experiments, but as the complexity of their work increases when they become graduate students, they may struggle to document these larger experiments. Many labs do not have formalized notebook-writing conventions in place. Strasser says that her approach as a PhD student in marine biology was makeshift — cutting and pasting computer printouts into a paper notebook, storing data in "chaotic heaps" on her computer and using the comment feature in Excel to describe results.

CARLY STRASSER

"There are just a lot of moving parts. Scientists don't ever really learn how to capture the process."
Carly Strasser

GO WITH THE FLOW

As researchers move beyond simply documenting their daily tasks, they are turning to electronic lab notebooks that have myriad bells and whistles and offer 'workflow' capabilities. These software packages aim to capture the entire scientific process — from study design to data analysis and visualization — in close to real time, and enable colleagues to review and replicate the work. Users can share data and methods as they develop, or choose to wait until publication.

A piece of software called R, for example, is widely used in the ecology community. With it, researchers can weave together text, programming code and data analysis into a narrative. The steps leading to a published paper no longer need to be impenetrable to other researchers — anyone can access any step along the way and reproduce it, or take the work in other directions. "These tools allow me to explore alternative ideas in a structured and documented way without disrupting my existing work," says Karthik Ram, a computational ecologist at the University of California, Berkeley.

Between 2009 and 2011, Ram was a postdoc at the University of California, Santa Cruz, studying the impacts of climate change on large mammals in Yellowstone National Park in Wyoming. He needed to incorporate decades of data from multiple sources — on the natural history of herbivores and on how migratory behaviour has changed over time, for example — and link it all to long-term climate data and changes in snowpack and vegetation. Because he uses existing data and models to answer questions, he often needs to understand intermediate steps along the way, such as the statistical methods used. Ram found it time-consuming and difficult to tease apart other researchers' data sets so that he could build on them. Workflow platforms such as R and IPython (see 'Taming the workflow') are now making this type of work much more manageable, he says. Other platforms include Projects (developed by Digital Science, a sister company of Nature Publishing Group).

Once a workflow system is populated with the nitty-gritty of the project, scientists can take advantage of Internet tools, such as wikis and social media, to post and share their work. Jean-Claude Bradley, an organic chemist at Drexel University in Philadelphia, Pennsylvania, has several projects in his lab that rely on wikis. Any raw data a project obtains, including pictures, videos and spectroscopy results, are quickly shared among team members, usually within a day. Lab members also use the wikis to share their findings. One wiki page represents one page in a lab notebook, with sections, objectives, methods and other components.

All these tools help to create what open-notebook aficionados call a living paper, which contains all components of the research, from e-mails to conference chatter

RECORD OF ACHIEVEMENT

A glossary of electronic notebooks

There are many ways to build a twenty-first-century electronic notebook — and to make work accessible to others.

- **Electronic notebook (open or closed)** A digital version of the conventional paper lab notebook in which the entire scientific process can be captured. Unlike conventional notebooks, which are difficult for others to access, electronic notebooks make it easier for researchers to organize, manage and share the many components of their work.
- **Workflow software** Platforms such as R and IPython integrate all the pieces of research into a single system. They capture data collection and analyses in near real time and these can then be shared with colleagues for review or collaboration simply by clicking a button.
- **Wiki** Web tools that serve as platforms for electronic notebooks. Wikipedia, the online encyclopaedia, is one example of a wiki. A wiki can be made accessible and modifiable by others. Many researchers who practice open science use wikis to record their progress and then share it with others. For example, a site called OpenWetWare allows researchers to build a notebook wiki.
- **Living paper** A term coined by researchers to describe the concept of archiving the entire scientific process, from phone calls and Twitter conversations to methods and analysis, in a way that is openly accessible. It allows other researchers to view the steps leading up to the final products and even to build on or derive ideas from the original work. A.M.

TAMING THE WORKFLOW

The open-access platform IPython

In 2001, as a final-year PhD student in theoretical physics, Fernando Perez started on a side project. For his research on quantum field theory, he needed to bring together his hotchpotch of computer code and data-analysis tools. Using a programming language called Python, he created IPython (ipython.org), an open-source, integrated platform that allowed him to type code, run his analyses, plot and visualize his data and add rich graphics within a single system.

"When I started working on IPython I told myself and my adviser that this was just going to be an afternoon hack and that I would get back to 'real work' very soon," he says. Some 13 years later, Perez is a computational scientist at the Brain Imaging Center at the University of California, Berkeley, and he develops IPython for computational science, publication and education across domain disciplines as his full-time job.

In 2011, he, along with collaborator Brian Granger at California Polytechnic State University in San Luis Obispo and their colleagues, added in a web-enabled notebook, which has been rapidly adopted by computational scientists working in the fields of biology, physics and neuroscience. It functions like a word processor, with normal text and formatting, but also enables users to insert programming language and rich graphics and data analyses, and easily go back and forth between them. "It's like having a very powerful calculator in the middle of your word processor that can do anything programming language can do," says Perez.

Researchers have even begun to publish papers directly from IPython, says Perez. The University of California, Berkeley, offers courses in IPython, and Harvard University and the Massachusetts Institute of Technology in Cambridge, and Columbia University in New York, among others, have adopted it.

The National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara, runs workshops on collaborative synthesis and data-sharing for graduate students and postdocs. And short, on-site workshops by a worldwide volunteer group called Software Carpentry offer hands-on training in Python, R, GitHub and other data-synthesis programs at various locations. **A.M.**

on Twitter, and links them together in an openly accessible, easily updatable, digital workflow. "A living paper is alive in that it gets updated and it reflects the ongoing process," says Boettiger. It allows scientists to document their entire scientific process, which they — and whomever they choose to share it with — can add to and derive new ideas from, says Matthew Jones, an expert in informatics at the National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara.

PRESERVED FOR POSTERITY

All these records must be stored, and archiving not only data and computations, but also intellectual discussions, presents challenges. Boettiger links his Twitter feed to his open notebook, and generates a running online tab of his reading and notes through the Mendeley reference manager. "Our literature has expanded so rapidly that other people are often our best sources of being able to figure out what to read," he says. Peter Andras, a computer scientist at Newcastle University, UK, also logs his reading through Mendeley and encourages his students to do the same.

But web-based resources often decline in popularity and if that happens, the record can languish unseen, or potentially even be lost. FriendFeed, for example, a hybrid between Twitter and Facebook, had a healthy researcher following in 2010 but now has substantially fewer users. Phillip Lord, a bioinformatician and a collaborator of Andras, says that for security he archives digital files through both archive.org, a general web repository, and an initiative at the British Library in London. Archiving services act as networks of libraries, storing data in multiple places to ensure that if one copy were lost or destroyed, it could be retrieved elsewhere.

Such challenges make it difficult to maintain electronic and open notebooks — but they are unlikely to stop their increasing adoption. "We're moving away from the science that we can document with a pen and paper to the science that we do in six different venues on instruments and with GPSs and laptops and paper and pen," says Strasser. "How do we capture that? It's super hard." Even so, the advantages are hard to ignore. "The days of paper lab books," Lord says, "are well past their best." ■

Amanda Mascarelli *is a freelance writer based in Denver, Colorado.*

GRANTS

Gender differences

Women are less successful than men in getting grants from Research Councils UK (RCUK), according to an 11 March report from the umbrella body for the UK government's research councils. RCUK's first such analysis compared the annual success rate during 2010–13 of grant applications from women and men in eight age groups. Across all age and grant categories, women averaged a 25% success rate, compared with men's 29%. Iain Cameron, head of research careers and diversity at RCUK, says that the reason for women's lower success is not clear. One council is launching unconscious-bias training for peer reviewers, and RCUK may do a follow-up analysis. "The point of publishing these data is to expose them," Cameron says.

MOBILITY

Swiss frozen out

Early-career researchers in Switzerland will have limited access to the Erasmus Plus student-exchange programme and other European research funds for the next academic year. Switzerland lost its status as an Erasmus Plus programme country after it voted to impose quotas on immigrants from European Union (EU) countries that restrict the free movement of people from Croatia, the newest EU member. The change means that Swiss nationals who wish to study in the EU will no longer be eligible for Erasmus grants and that students from eligible countries will not be able to obtain an Erasmus grant to study in Switzerland. Erasmus funds around 7,000 exchanges a year between Switzerland and other countries.

SALARIES

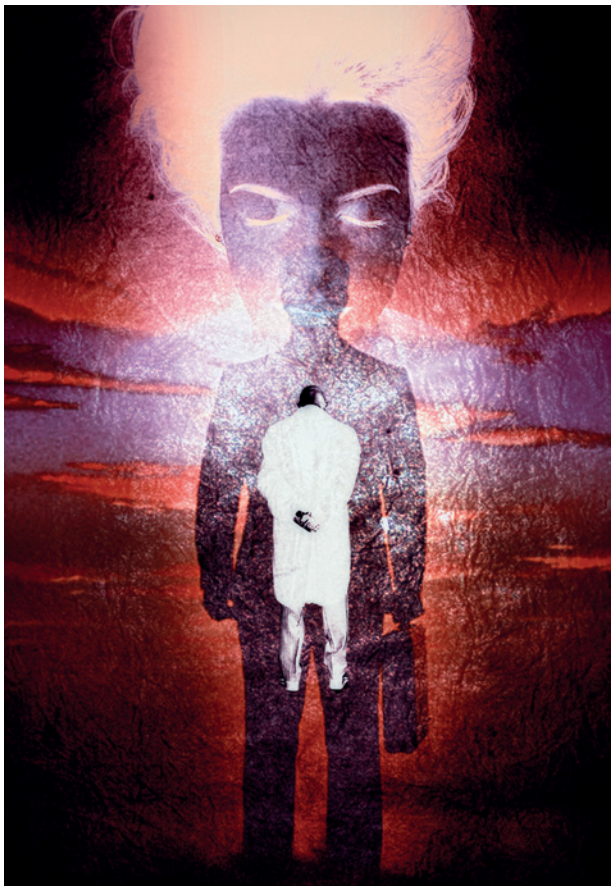
Public boost

Pay rises were slightly higher for faculty members at US public institutions (2.2%) than at private institutions (2.0%) for the first time since 2009, finds the College and University Professional Association for Human Resources (CUPA-HR) in Knoxville, Tennessee. The change suggests an overall strengthening of state economies, says Andy Brantley, CUPA-HR president. CUPA-HR's poll of 792 institutions and 180,000 faculty members also showed that compensation for tenure-track and tenured faculty members across both university categories rose by an average of 2.1% this academic year, the same increase as the year before.

LOOKBACK

Time well spent.

BY GEORGE ZEBROWSKI



JACEY

As usual, I would have to leave before I arrived. Memory threw me there, exiling me from all my other times with her, no matter how much I concentrated. She was alive in all our pasts, but only these earlier presents were open to me.

"Me again?" she asked in our present, jealous and prideful.

"No one else," I said.

I kissed her before I bridged, fighting off Maxim Gorky's claim that "love is the failure of mind to understand nature". If so, then love was an opposed way, an uphill fight at best.

She was asleep back there in our off-campus apartment, as I came up by train from New York City. I would have just enough time to get there and spend some time with her before the train arrived.

I always prepared by losing a pound or two, colouring my hair a bit and exercising, even using some make-up to look younger than my late 60s, so that she would not notice in the dim light of the apartment at night. Nearsighted and in bed, it helped that she would not be wearing her glasses.

I grasped my key from decades past, and summoned the vision of the pale-skinned young woman who had dyed her hair black after a silly blonde experiment, and then cut it short when I was away. I would again compliment the change.

My appearance at my old door shook the back porch for a moment. I stood before the curtained glass, but no light went on. I was fearful that this might be my last chance to regain this time, so I had to make it count. Other times with her might open to me if this one closed, but that was far from certain.

The theory of jumps was not perfect. There might not even be any real time displacement at all, but instead a reality-like recreation of significant memories that suddenly occupied a mind with a quantum flood of insistence to the point at which it made no difference to the experience; it might just as well be happening in the naive sense. Time probably did not exist outside the biology of human perception except as a timeless persistence, a stubborn duration, inexpressible endurance beyond time-like words.

I turned my key in the lock, pushed the door open and went in, closing it behind me.

"Who's there?" asked her voice, from somewhere inside me, it seemed.

"It's me," I said, hoping to sound younger.

"Oh," she said uncertainly.

I crossed the small living room to the double bed in the alcove and sat down. Her head came out from beneath the covers, hair cut short and dyed black. She looked up at me like a queen on a divan.

"Beautiful!" I said, and she giggled as I lay down beside her.

"You must be tired," she said as I sighed. "It's okay, we can just sleep," she whispered. "We have all tomorrow."

She dozed. I lay there, afraid to disappoint her.

After a while I looked at my watch and saw that I would soon arrive, and it would not matter. Were anachronisms real or only apparent? You can have all the anachronisms you want in your mind, where they happen all the time. Was I asleep up ahead? I felt a rocking sense of loss as I heard the train whistle.

"What is it?" she asked softly as my moments with her fled into some deep abyss where I could not follow.

"I'll be right back," I said and got up. She turned over and closed her eyes, expecting that I was only undressing. I stood there, looking at the glory of her bare back. To kiss her now might be fatal if I arrived early.

I went out the door and breathed the night air, knowing that I was coming down the street, and that I could leave it all to him.

The starry night was blue. I went across the yard and stood by the brick wall of the garage. I would come through the narrow alleyway from the street beyond, less than a block from the train station.

My memory moved within him. I was waiting here only to see him walk by in the dark. The unreality of time seized me, and seemed about to inflict a wretched pain, but relented.

Time could show mercy if you remembered well enough, a collective delusion, the setting of human psychology, biologically based, to conceal the fact that everything was in one place and happened all at once.

I would return, inexactly, to this very time and place as often as my force of memory struck out against loss. I looked around the dark yard. I was present here in these shadows more than once, similarly aware, only moments apart.

He came through the dark alleyway, and I felt the flow of love for his waiting beloved; love then, not my love now. His step was sure, the past his own, his youth holding back the incoming future. The same key I still held was in his pocket.

He would not disappoint the dark-haired goddess in her bed.

I was gone before I got to the door. ■

George Zebrowski *is best known for his Campbell Award winning novel* Brute Orbits, *the classic* Macrolife, *and* Cave of Stars, *recently chosen as one of the best SF novels since 1985. A collection of stories with Jack Dann,* Decimated, *is from Borgo/Wildside Press. All his fiction is available from Gollancz's SF Gateway (www.sfgateway.com) and E-Reads (ereads.com).*

➲ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm