nature

EGYPTIAN
Genetic link
confirmed

CANCER
Control or cure?

FUSION ALTERS
ITER delayed, again

BIOMEDICAL SUPERMODEL

Germline transmission in a
transgenic non-human primate

SUPERLENSES

Abstractions

FIRST AUTHOR

Among the brain's many mysterious goings on are theta oscillations. This term describes a brain rhythm observed in the hippocampus — the area responsible for long-term memory and spatial navigation — in many mammals, including rodents and humans. Although the oscillations' function is not fully understood, they are thought to be the 'clock' that controls the timing of hippocampal activity. The prevailing view is that theta oscillations occur synchronously across the hippocampus. But on page 534, Evgueniy Lubenov and Athanassios Siapas of the California Institute of Technology in Pasadena challenge that assumption, demonstrating that theta oscillations do not occur in synchrony, and instead travel across the hippocampus. Lubenov explains why these experiments required considerable patience.

Why did you question the prevailing view?

While recording electrical activity in different regions of the rat hippocampus, we saw differences between electrodes in the phase and amplitude of theta oscillations. Such variations would usually be attributed to the fact that the electrodes were implanted at different depths, because theta phase and amplitude are known to vary according to depth in the hippocampal layers. We wondered whether phase might also depend on other factors, for example, anatomical location across the hippocampus.

How did you tackle the problem?

We started with a series of experiments involving long vertical probes with multiple recording sites, which allowed us to map the depth-dependence of theta oscillations with high resolution. We found a 400-micrometre-thick layer of the hippocampus where the phase of theta oscillations is insensitive to depth. Using custom arrays, we then positioned grids of 28 electrodes in this target region and compared the phases of theta oscillations recorded at different anatomical locations across the hippocampus in freely behaving rats.

What did you find?

That theta oscillations propagate in a consistent direction, and so travel through a series of 'time zones' along the length of the hippocampal axis. We plan to study the mechanisms responsible for these waves. This will help us to reveal how information flows and is processed in the hippocampus.

Was it hard to make these measurements?

Making simultaneous recordings over roughly one-third of the hippocampus in freely behaving animals is not easy. It can take several weeks to gradually position this many electrodes to their targets. ■

MAKING THE PAPER

Olga Fehér

Songbirds' melodies have roots in both genes and environment.

Just as humans teach their children to speak, juvenile birds learn to sing from their parents. In zebra finches, songs are the preserve of the male, and are passed from father to son. When young males are deprived of such training, their song is a poor rendition of that typical of the species. So what does that mean for generations to come if these untrained singers take on the role of tutor? For her doctoral dissertation at the City University of New York, Olga Fehér set out to find an answer.

Fehér's experiment involved having an 'isolate' male zebra finch — one reared in the absence of birdsong — raise a brood. One of the finch's progeny went on to become the main breeding male for the entire colony. Over three to four generations, each subsequent set of offspring came progressively closer to singing the species-specific song. From this finding, Fehér concludes that a culturally transmitted behaviour such as birdsong is determined in part by genetics and in part by environment.

The experiment's success relied on access to the latest sound-analysis technology. Previous studies of birdsong from untrained versus trained birds were able to merely eyeball the differences in 'syllables' as represented on a sonogram. "But that really doesn't cut it when you want to quantify subtle changes in birds' songs," Fehér says.

To obtain more precise measurements, Fehér took advantage of a long-standing collaboration between her PhD adviser, Ofer Tchernichovski, and Partha Mitra, a theoretical physicist at Cold Spring Harbor Laboratory in New York. Mitra helped Tchernichovski to develop sound-analysis software for studying birdsong. It took into account ten different variables of sound, including frequency variation and amplitude. As a result, Fehér was able to compare different birds' songs using objective measurements (see page 564).

Other experimental methods were less sophisticated, albeit equally effective. Fehér needed a low-budget, sound-attenuating chamber in which to keep her song-naive bird colony. After searching the streets of New York City for discarded refrigerators, she found one within her building. She laid the refrigerator on its side and had it equipped with ventilation, lighting and microphones, as well as perches, swings, nest boxes and food.

Mating the isolate male was not easy. At first, none of three females placed in the cage showed any interest in him, possibly because they were accustomed to normal zebra-finch song and were not attracted to his unusual vocalizations. "We don't know much about their preferences," Fehér says of her female finches, "but one of them eventually took to the male".

After that, Fehér was able to let the colony evolve naturally. "Who learned from whom wasn't really controlled. What fascinated me was that, looking at the sound, it seemed as if the male offspring were imitating their father. But they obviously also had input from their siblings, especially those from previous clutches." Such influences, Fehér believes, are what led to the gradual progression towards wild-type song.

Although the nature-versus-nurture debate was not the inspiration for the study, Fehér says, "it's an interesting example of this interaction, because it shows how strong the input is from the nature part. But it doesn't work without the nurture." ■

FROM THE BLOGOSPHERE

Where science is concerned, the media is often guilty of presenting only one side of a story. On the Climate Feedback blog, *Nature Reports Climate Change* editor Olive Heffernan addresses complaints that only 'doom and gloom' stories were reported from the recent International Scientific Congress on Climate Change in Copenhagen (http://tinyurl.com/r5c9bo). She gives readers a glimpse of how difficult it is for a reporter to glean the most newsworthy information from a large scientific conference.

Many news stories that emerged from the congress focused on estimates of sea-level rise — the only topic presented in a press conference. Some might argue that "reporters were simply picking the low-hanging fruit rather than seeking out diverse news stories amid the numerous (57) sessions," writes Heffernan. "But if you have to file on deadline, then getting a heads-up on new exciting research with sources in attendance isn't a bad strategy." She adds that the social-science presentations on avoiding climate catastrophes are more subtle stories that require long-term research and reporting beyond breaking news. ■

Visit Nautilus for regular news relevant to *Nature* authors ▸ http://blogs.nature.com/nautilus and see Peer-to-Peer for news for peer reviewers and about peer review ▸ http://blogs.nature.com/peer-to-peer.

nature

www.nature.com/nature

Vol 459 | Issue no. 7246 | 28 May 2009

Time to connect

Researchers who use genetically modified primates need to gear up for the inevitable public debate.

Japanese researchers announce a major accomplishment in this issue of *Nature*: the creation of the first transgenic primates able to pass on a foreign gene to their offspring (see pages 492, 515 and 523). Because the primates in question are marmoset monkeys that are distant from humans in an evolutionary sense, this experiment has little immediate bearing on the modification of human germ lines — a prospect that many people find unacceptable in any case. But the advance will lead to more sophisticated models for human disease, physiological development and neurogenetics. And in so doing, it will inevitably draw more attention from animal-rights activists.

Posters that feature an endearing marmoset face peering out of a cage and a caption denouncing experiments will make for an emotionally appealing campaign. Granted, the transgenic marmosets presented this week have been equipped with a simple green fluorescent protein; the only thing that they seem to do differently from normal monkeys is glow. But experiments such as these are headed in a new direction. The study of human diseases will require engineered marmosets that are born with and live their entire lives with a genetic defect that they pass on to their progeny. And this prospect — intentionally introducing a harmful gene into the primate gene pool — promises to raise the stakes in the long-standing controversy over animal rights.

Nature believes that such experiments are justified by the value of the research, as long as they are carried out in a responsible fashion. But researchers must be ready to deal with the broader ethical questions involved. If unprepared, they risk being caught wrong-footed when the debate inevitably takes off.

Although Japan is poised to become a leader in this field of primate transgenics, its researchers seem less than ready to take on the bioethical challenge. Japanese researchers and policy-makers have tried to answer concerns about animal rights by introducing animal-experimentation guidelines and by holding public symposia to expound the value of the research. But these initiatives, although worthwhile, are inadequate when public and political sympathy can so easily be influenced by opponents.

Animal research in Japan and its supporters should heed the lessons from the West, where researchers have won several political victories by addressing the issues openly. For example, during last year's debates over hybrid embryo research in the United Kingdom, scientists publicly demonstrated an acute awareness of the ethical aspects of research practice and of the potential applications, showing respect for opposing ethical positions yet rapidly rebutting misrepresentations about the research. This forward-looking and transparent approach won public and political support at the crucial moments at which regulation was being formulated.

"The lesson for the animal-rights debate is that more engagement, not less, is the best way to promote the research."

Following that model will not be easy for Japanese scientists, who freely admit their dislike for public confrontation. Although they have not yet had to face the kind of violence and intimidation that is sometimes practised by animal-rights activists in Europe and the United States, researchers in Japan generally shy away from presenting details of what they do. But if they do not actively seek a positive solution to promoting their research, they run the risk of an adverse one being reached — quite possibly in a bureaucratic back room in which flummoxed officials are under pressure from anti-research activists. In embryonic-stem-cell research, for example, Japanese researchers won an ostensibly lax legal framework, but critics were able to force the inclusion of laborious regulations that crippled the stem-cell community for years (see *Nature* **438,** 263; 2005). Primate researchers in continental Europe have also been caught off guard by behind-the-scenes negotiations (see *Nature* **444,** 812–813; 2006).

The lesson for the animal-rights debate is that more engagement, not less, is the best way to promote the research. Scientists everywhere must be ready to discuss animal welfare, the possibility of transgenic humans and other controversial topics, based on a thorough understanding of the ethical issues and a capacity to respond rapidly in public. ■

Transparency needed

Taxpayers deserve to know the cost of ITER, the international fusion project they are paying for.

The car park of ITER's headquarters in St Paul-lez-Durance, France, is neatly divided by a green metal fence. On one side is the temporary building that houses the ITER Organization, the body overseeing the international fusion-energy project. On the other lies a roughly one-kilometre by half-kilometre earthen platform on which the fusion reactor will be constructed. Yet members of ITER's central team cannot visit it without an escort. The site is under the tight control of the CEA, the French Atomic Energy Commission.

This is the realpolitik of the planet's most expensive science experiment. In 2006, when ITER was approved by the European Union, China, India, Japan, Russia, South Korea and the United States, the members were careful to ensure that their own agencies had strong influence on the project. Through ITER's governing council, domestic nuclear agencies have ultimate control over everything from the appointment of the central organization's senior staff to the project's scope and schedule. The same agencies will be the ones letting contracts to industry and overseeing production of most of the parts for the machine.

The original cost estimate for ITER was €5 billion (US$7 billion) for construction and the same amount again to operate it for 20 years. But costs are rising and the schedule is growing ever more drawn out (see page 488). Educated guesses now put the construction costs at roughly €10 billion.

These guesses might be the closest that anyone will get to knowing ITER's true cost. The national agencies running the project are under no obligation to tell the central organization how much they are paying their industrial contractors for each piece of the reactor. Some countries might disclose the cost of the components they build, but others might well wish to keep their budgets secret — both to protect their nation's industry and to shield themselves from potential embarrassment. As long as the pieces are technically satisfactory, the central team must accept them, no questions asked.

This set-up may be good for the nuclear agencies running ITER, but it is deeply unfair to those who will ultimately pay for the project — around half of the world's taxpayers. These citizens have a right to know how much their countries are paying in the quest for fusion energy. The seven members of ITER should explicitly commit themselves to providing cost estimates for their in-kind contributions to the project. The central organization should collect and review those estimates, and it should then make them available to the public, both individually and as a collective price tag.

If any members of ITER's council are unwilling to provide those data, then politicians should take up the issue. ITER is run by agencies that have to answer to political bodies such as parliaments. Where necessary, these groups should intercede to demand the publication of financial information about ITER.

A full financial disclosure could be painful. It is likely to show that the reactor is costing far more than originally promised and that some countries are paying more than others for the same components. Those revelations could lead the public to ask numerous questions — including whether fusion can deliver affordable electricity on a timescale that anyone could deem germane.

Such questions are difficult to answer, but they must be asked of every energy technology if humanity is to tackle the climate and energy challenges ahead. The only way to decide fusion's role in resolving these challenges is via an honest public debate. And that debate cannot take place without a transparent price tag for the world's first fusion reactor capable of producing net energy. ■

Media frenzy

A hyped-up fossil find highlights the potential dangers of publicity machines.

Last week's publication of paper describing a 47-million-year-old fossil primate with a remarkable degree of preservation (see http://tinyurl.com/oycvo8) prompted a trickle of news in *The Daily Mail* that quickly swelled to a flood of media coverage.

In normal circumstances, the interpretation of the specimen given in the paper (J. L. Franzen *et al. PLoS ONE* **4,** e5723; 2009) would have been no more contentious than that of any other fossil primate, and a good deal less so than some. The fossil, called *Darwinius masillae*, represents a new species that seems to be closely related to other (albeit fragmentary) primate fossils found at the same site, near Messel, Germany. These belong to a group of extinct primates called adapids, which are generally considered to be more closely related to the sub-order containing lemurs and bushbabies (strepsirrhines) than to the one that includes higher primates and humans (haplorrhines). That suggests that the new species has little to do with human ancestry.

Indeed, in the paper the authors explicitly state that *Darwinius masillae* "could represent a stem group from which later anthropoid primates evolved, but we are not advocating this here, nor do we consider either *Darwinius* or adapoids to be anthropoids". The authors also refrain from claiming that the fossil changes our understanding of primate evolution.

But the circumstances surrounding the paper's publication were anything but normal. Before the paper had even been submitted to the journal, Atlantic, a production company based in New York, had commissioned a television documentary and an accompanying book about the find. Just a week after the paper appeared, the book has been published and the documentary has been aired on the History Channel in the United States, as well as Britain's BBC and Norway's NRK.

Both book and documentary include the the suggestive words 'The Link' in their titles. A press release associated with the New York press conference at which the fossil was first officially described claimed that the fossil represents revolutionary changes in understanding. The History Channel website calls the find a "human ancestor", and the BBC website describes it as "our earliest ancestor".

To be fair, the authors' claims at the press conference were appropriately measured. Nonetheless, the researchers were fully involved in the documentaries and the media campaign, which associate them with a drastic misrepresentation of their research.

Another damaging aspect of the events was the unavailability of the paper ahead of the press conference and initial media coverage. This prevented scientists other than those in the team from assessing the work and thereby ensuring that journalists could give a balanced account of the research.

There is no reason to think that *PLoS ONE*'s editors and reviewers did less than their duty to the paper. Nonetheless, the clock was ticking at the time of submission. *Nature* has over the years received occasional offers of papers associated with television documentaries, and the offers usually come with broadcast dates attached. Where the refereeing process might have been compromised, we have always said no to the paper. When time is tight, there is a risk that the broadcast will go out even if any problems uncovered by peer review cause the paper to be delayed or rejected.

In principle, there is no reason why science should not be accompanied by highly proactive publicity machines. But in practice, such arrangements introduce conflicting incentives that can all too easily undermine the process of the assessment and communication of science. ■

RESEARCH HIGHLIGHTS

Big volcano, tiny troubles

Geology **37,** 435–438 (2009)

Potentially dangerous silica nanofibres have been identified in airbourne ash spewed across southern South America by a Chilean volcano. Martin Reich and his colleagues at the University of Chile in Santiago used high-resolution transmission electron microscopy to image the one-dimensional crystalline silica nanostructures, called cristobalites. They were formed during the eruption of Patagonia's Chaitén Volcano, which began on 2 May 2008 and is ongoing.

The researchers propose that amorphous silica was reduced by carbon monoxide and then oxidized to become breathable crystalline nanostructures. The formation of these structures was enhanced by micrometre- to nanometre-sized silica glass fragments in the volcanic column.

C. BROWN/EPA/CORBIS

NANOMEDICINE

Dude, where's my dot?

Nano Lett. doi:10.1021/nl900872r (2009)

Targeting diseases with nanotechnology-based therapies supposes precise knowledge of where nanodevices go when released in the human body. Yet current knowledge is anything but precise.

John Frangioni of Beth Israel Deaconess Medical Center in Boston, Massachusetts, Moungi Bawendi of the Massachusetts Institute of Technology in Cambridge and their colleagues went looking for answers in rats. They used near-infrared-emitting semiconductor nanoparticles coated with a polymer of varying lengths to determine how size and hydrophilicity affect where particles end up. Very small particles (around 5.5 nanometres wide) can be excreted by the kidneys, but even smaller ones get trapped in the liver; larger particles seem to target the pancreas, usually difficult to get to because it has few molecular targets. Meanwhile, the largest particles remain in the vasculature for long periods of time.

ANIMAL BEHAVIOUR

Simian support networks

Biol. Lett. doi:10.1098/rsbl.2009.0204 (2009)

Game theory can explain the evolution of cooperation in large, randomly structured groups through the balance between costs and benefits. However, groups of primates that exhibit cooperative behaviour are often neither large nor randomly structured.

Bernhard Voelkl and Claudia Kasper at the Strasbourg arm of the CNRS, France's basic-research agency, examined data on 70 primate groups ranging in size from 4 to 35 individuals, taking in a total of 30 species. They modelled the groups to see whether, over time, individuals would be better off cooperating or not.

They found that 61 of the groups were more likely to cooperate than they would have been in a less structured group of the same size, suggesting that the structure assists in the fixation of cooperative behaviour.

ATOM OPTICS

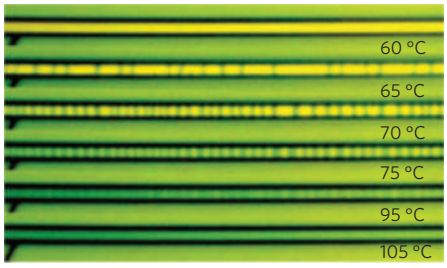
Seeing spots

Phys. Rev. A doi:10.1103/PhysRevA.79.053823 (2009)

The original 'Poisson spot' experiment had a crucial role in proving the wave nature of light: because of the way that waves diffract, light shining at a circular object casts a bright spot in the centre of the object's shadow.

Thomas Reisinger of the University of Bergen in Norway and his colleagues now demonstrate that a beam of deuterium molecules can also create a Poisson spot — replicating the classical experiment using neutral matter waves.

The authors suggest several potential applications, ranging from 'printing' larger molecules precisely on a surface to the study of quantum decoherence and other effects at the boundary of quantum and classical physics. Unlike other diffraction methods, Poisson's spot is wavelength independent.



MATERIALS SCIENCE

A material turnabout

Nature Nanotechnol. doi:10.1038/NNANO.2009.141 (2009)

Vanadium dioxide (VO_2) has long been a puzzle to condensed-matter physicists — above 68 °C it acts as a conductor, but below that temperature it is an insulator. Some say that collective motion of the electrons is mainly responsible for this transition, whereas others suggest that interactions between the electrons and the vibrations in the material's crystal lattice contribute to the odd behaviour.

David Cobden and his colleagues at the University of Washington in Seattle approached the mystery using nanobeams of VO_2, which are devoid of the defects and variations that obscure the results of experiments on larger samples. (Pictured below left, a 40-micrometre-long nanobeam imaged during warming shows dark metallic domains appearing and widening.) Their findings suggest that electron–electron interactions are behind the transition.

The same approach may prove useful in studying other materials. As for VO_2, it may find uses in electronics or mirrors that can be switched on and off at will.

GENETICS

A tippling gene

Cell doi:10.1016/j.cell.2009.03.020 (2009)

Mutating a gene in the fruitfly *Drosophila* increases the insect's resistance to the sedating effects of alcohol, report Ulrike Heberlein at the University of California, San

Francisco, and her colleagues. They call the gene *happyhour* and have found, through a series of genetic manipulations in flies, that the normal happyhour protein seems to work by inhibiting the activity of the extracellular growth factor receptor pathway. This is conserved in flies and mammals, and is a target of some cancer drugs.

One such drug, erlotinib, made lightweights of flies and mice, enhancing their sensitivity to ethanol, and reduced alcohol consumption in rats that had become accustomed to a tipple. The authors suggest the results might point to therapeutic avenues for people with drinking problems.

OCEANOGRAPHY

Arctic freshening

Geophys. Res. Lett. doi:10.1029/2009GL037525 (2009)

An inventory carried out during the International Polar Year 2007–2008 reveals substantial increases in fresh water in parts of the Arctic Ocean.

Matthew Alkire of Oregon State University in Corvallis and his colleagues found that in March and April 2008 the freshwater content in the Canada and Makarov basins had increased by around 8,500 cubic kilometres, or 26%, compared with average freshwater content in winters past. Other areas saw small decreases.

Chemical analysis suggests that precipitation and increased river runoff, as well as Pacific water, are dominant freshwater sources; accelerated sea-ice melting also has a role. Prevailing anti-cyclonic wind patterns seem to have favoured freshwater redistribution towards the western Arctic basin.

ANIMAL BEHAVIOUR

Singing in the rain

Curr. Biol. doi:10.1016/j.cub.2009.04.061 (2009)

Mockingbird songs are more elaborate in species that live in more variable climates.

Carlos Botero of Cornell University in Ithaca, New York, and his colleagues analysed almost 100 separate recordings spanning 29 species of mockingbird collected from across the New World. Places where precipitation and temperature were more variable and less predictable had birds that sang more consistently, had a broader repertoire and were better at copying the calls of other species.

The researchers suspect that the correlation relates to a sexual display of fitness, either spurred by increased competition for limited resources and females in variable climes, or tracking with neurological adaptations needed to survive in an unpredictable world.

MICROBIOLOGY

Bacterial break-in

J. Clin. Invest. doi:10.1172/JCI36759 (2009)

Many have wondered how pathogens that cause bacterial meningitis slip through the tight defences of the blood–brain barrier.

Dlawer Ala'Aldeen at the University of Nottingham, UK, Elaine Tuomanen at St Jude Children's Research Hospital in Memphis, Tennessee, and their collaborators used murine and human cells and live mice to show that *Streptococcus pneumoniae*, *Neisseria meningitidis* and *Haemophilus influenzae* all use the same means of entry. These bacteria recognize and attach to a specific target on laminin receptors on the inner surface of the brain's many blood vessels.

The team isolated the proteins used by the pathogens to bind to this receptor. These could be used in the design of a broadly protective meningitis vaccine.

ELSEVIER

DEVELOPMENTAL BIOLOGY

Use it or lose it

Dev. Cell **16,** 734–743 (2009)

Embryos need to flex their growing muscles if developing cells are to give rise to joints, says a team led by Elazar Zelzer at the Weizmann Institute of Science in Rehovot, Israel.

They found that mutant mouse embryos with defective muscles fail to form various joints, including elbows, shoulders and hips (a normal embryo is pictured above). Without muscle contraction, the cells that generate joint tissues do not activate a key regulatory pathway controlled by the protein β-catenin, and the progenitors switch fate to form cartilage instead.

One idea put forward by the authors is that the mechanical stress created by developing muscles might inform the cells of where they are and what cell type they should generate. The study could be relevant to rare human cases in which babies whose movement is restricted *in utero* develop abnormal joints.

JOURNAL CLUB

Maite Huarte
Broad Institute, Cambridge, Massachusetts

A biologist looks at new functions for non-coding RNAs.

The increasing study of small and large RNA molecules that do not encode protein — non-coding RNAs — is widening our view of their relevance, and of their roles in important developmental mechanisms such as gene silencing and X-chromosome inactivation. Nevertheless, our knowledge covers only a fraction of the non-coding transcripts produced from the mammalian genome.

Much of the non-coding RNA transcribed is associated with protein-coding genes: for example, the transcripts that are complementary or 'antisense' to the gene sequence. These can be created by 'bidirectional' transcription from either DNA strand. Kevin Morris of the Scripps Research Institute in La Jolla, California, and his colleagues have now shed light on the function of this type of transcription (K. V. Morris *et al. PLoS Genet.* **4,** e1000258; 2008).

They focused on the gene encoding the tumour suppressor p21, transcription of which must be finely tuned, and show that an endogenous antisense transcript of p21 controls the amount of p21 mRNA made by silencing its promoter. This transcriptional suppression is dependent on Argonaute-1, a protein implicated in RNA-mediated gene silencing. Suppression correlates with bidirectional transcription within p21's promoter.

This observation is not limited to p21: a similar regulatory mechanism controls gene expression of the protein E-cadherin, suggesting that this balancing of sense and antisense transcription might be a common mechanism of transcriptional regulation.

The next challenge is to understand how RNAs can induce transcriptional gene silencing; information that will probably reveal just how much power RNA wields in the control of gene expression.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

SPECIAL REPORT

Fusion dreams delayed

International partners are likely to scale back the first version of the ITER reactor. **Geoff Brumfiel** reports.

ST PAUL-LEZ-DURANCE, FRANCE

ITER — a multi-billion-euro international experiment boldly aiming to prove atomic fusion as a power source — will initially be far less ambitious than physicists had hoped, *Nature* has learned.

Faced with ballooning costs and growing delays, ITER's seven partners are likely to build only a skeletal version of the device at first. The project's governing council said last June that the machine should turn on in 2018; the stripped-down version could allow that to happen (see *Nature* **453,** 829; 2008). But the first experiments capable of validating fusion for power would not come until the end of 2025, five years later than the date set when the ITER agreement was signed in 2006.

The new scheme, known as 'Scenario 1' to ITER insiders, will be discussed on 17–18 June in Mito, Japan, at a council meeting that will include representatives from all seven members: the European Union (EU), Japan, South Korea, Russia, the United States, China and India. It is expected to be approved at a council meeting in November.

Indeed, the plan is perhaps the only way forward. Construction costs are likely to double from the €5-billion (US$7-billion) estimate provided by the project in 2006, as a result of rises in the price of raw materials, gaps in the original design, and an unanticipated increase in staffing to manage procurement. The cost of ITER's operations phase, another €5 billion over 20 years, may also rise.

"Building ITER is like building the space station, but having to set up NASA and ESA in the process."

In fact, the ultimate cost of ITER may never be known. Because 90% of the project will be managed directly by individual member states, the central organization has no way of gauging how much is being spent, says Norbert Holtkamp, ITER's principal deputy director-general. "They won't even tell us," he says. "And that's OK with me."

Holtkamp says that the only way to get ITER built is to do the skeletal version first. Before scaling up to do energy-producing experiments, he says, "you really need to know whether the major components work. It's absolutely clear that this is the right approach." As to why Scenario 1 is being touted only now, Holtkamp says it took him time after joining the project to review the original schedule.

Fusion researchers say that Scenario 1 is preferable to the alternative: a permanent smaller machine that would never produce significant amounts of power. "You can't build a half ITER because then you'll just go on and on not quite knowing what the answer is," says Steven Cowley, director of the UK Atomic Energy Authority's fusion laboratory at Culham.

The project's rising price and lengthening schedule have angered some of ITER's members, who plan to finalize the schedule and budget by the end of this year. "People are pissed off," says a source close to the negotiations who requested anonymity because of political sensitivity.

ITER is the most ambitious fusion experiment ever proposed. At its heart is a doughnut-shaped device known as a tokamak (see graphic), which uses magnetic fields to squeeze and heat hydrogen isotopes to hundreds of millions of kelvins, until they fuse. The consequent fusion reactions release high-energy neutrons that can, in principle, be harnessed to generate electricity. Normal hydrogen will not generate enough fusion events to produce large amounts of power, but when scientists inject deuterium and radioactive tritium into the machine, it should generate roughly 500 megawatts of thermal power — around 10 times the amount of power needed to run it. Such an achievement would be long-sought proof that fusion power can work, although a commercial reactor would still be decades away.

A 180-hectare stretch of land has already been cleared for ITER.

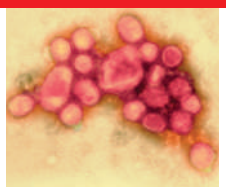
ITER

The machine's costs have doubled once before. Budgeted at US$6 billion in 1989, costs climbed to nearly twice that during the decade that followed. Angered, the US Congress withdrew the country from the project in 1999. The remaining partners — Europe, Russia and Japan — pressed ahead to redesign the machine to keep it within $6 billion.

Design scaled back

The redesign fell to Robert Aymar, a prominent French physicist who had overseen the French tokamak Tore Supra, in Cadarache. Aymar redesigned the entire device in just three years with a staff of 70 — half the size of the original design team. In addition to shrinking ITER to its present-day size, the team made other money-saving assumptions. The new design did not include the cost for auxiliary equipment and some spare parts, and it assumed that some buildings could be reused multiple times. The design was accepted by ITER's partners in 2001 as a way to keep within the $6-billion price tag, and used as a baseline when the United States rejoined and India, South Korea and China joined the project.

Aymar maintains that his redesign was never meant as a final blueprint for the machine, but to act as a guideline for ITER's members to decide who would contribute which portion. It did not



AMI IMAGES

UNEXPECTED ACTIVITY
Old antibodies may protect elderly from swine flu.
www.nature.com/news

assign costs to the values of various components, but used 'ITER units of account' to allow nations to negotiate which parts they would provide as their share. "What we provided was not a cost," he says. "We provided a value."

Some ITER partners say Aymar's design contained an appropriate level of detail. "It was a conceptual design," says Octavi Quintana Trias, director for energy at the European nuclear research organization Euratom, and an EU representative on the ITER council. "You will not develop and refine the design until you have a real commitment to go ahead."

But some US officials see things differently. "We thought that it was 80% designed and that you only had 20% to fill in, and it turned out to be more like 40%," says Raymond Orbach, former head of science at the US Department of Energy and until this year a US representative on the ITER council.

When the project was approved in 2006, the newly formed ITER organization set to work reviewing and completing Aymar's 2001 design, adding parts that included a set of superconducting magnets to control a type of instability not anticipated in the earlier design. The adjustments have cost time and money.

Agencies in the seven member governments are also struggling to set up the complex network that will eventually supply ITER with parts. Although the specifications for each piece can be set by the central lab, it is individual governments that will award contracts to industry and oversee production.

Who provides what is detailed through a series of procurement agreements, which have been slow in coming: 26 were originally scheduled to be ready by the end of 2008, but as of late May this year, only 17 had been signed. The delay is in part due to setting up domestic organizations that will award the contracts, says Niek Lopes Cardozo, vice-chairman of Fusion for Energy, the Barcelona-based body overseeing contracts to European industry. "Building ITER is like building the space station, but having to set up NASA and ESA [the European Space Agency] in the process."

During negotiations, each nation bargained for a stake in the most technically sophisticated parts of the machine, and, as a result, single components will be built with parts from several nations. For example, the 150,000 kilometres of superconducting wire for the magnetic coils will be produced in China, Japan, Russia, Korea, the EU and the United States.

Awarding contracts, setting up production lines and ensuring quality control across the seven partners will also require significantly more manpower than anticipated in 2001, when there were only three member states. Holtkamp says that the central project office will need up to 750 staff to do the job — a 25% increase over the original plan.

Nations could bring down the cost of the machine by swapping procurement agreements and consolidating production, but that's not politically realistic, says Kaname Ikeda, ITER's director-general. "The first objective is [for members] to get the experience of designing, fabricating and operating this machine," he says. Akeda adds that most seem willing to live with the cost increase, so long as they gain the knowledge needed to build a power-producing fusion reactor.

Member governments are preparing to shoulder the increase. Last year, the United States upped the estimated cost of its contribution to $2.2 billion — double the preliminary price. The EU, meanwhile, is grappling with increased spending to complete the buildings for the project. We are "trying to get rid of everything that is not indispensable," says Trias. "Each country has its own internal problems," adds Evgeny Velikhov, a Russian representative on the ITER council, although he adds that "Russia does not see any problem in fulfilling its obligation".

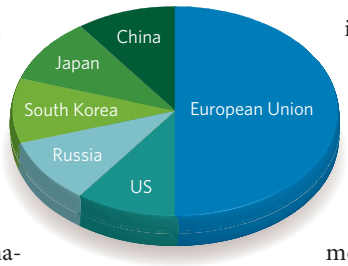
Those close to the project now see Scenario 1 as the only practical way forward. Under the plan, the reactor would initially be built without several crucial and expensive components, including an inner shielding wall and test bed for new materials such as lithium blankets that generate tritium for the machine, along with the diverter, a series of tiles at the bottom of the tokamak that shunts heat safely out of the device. Also gone will be expensive accelerators to pump neutral beams of fuel into the machine, and some radio-frequency devices designed to further heat the plasma. Without these components, ITER can handle only plasmas of hydrogen, not deuterium or tritium.

The plan would allow scientists and engineers to ensure that ITER works long before it is injected with tritium — a process that will make large sections of the machine inaccessible. "They want to see success at the end, and the way to success is this road," says Holtkamp. "There really is no other one." ■

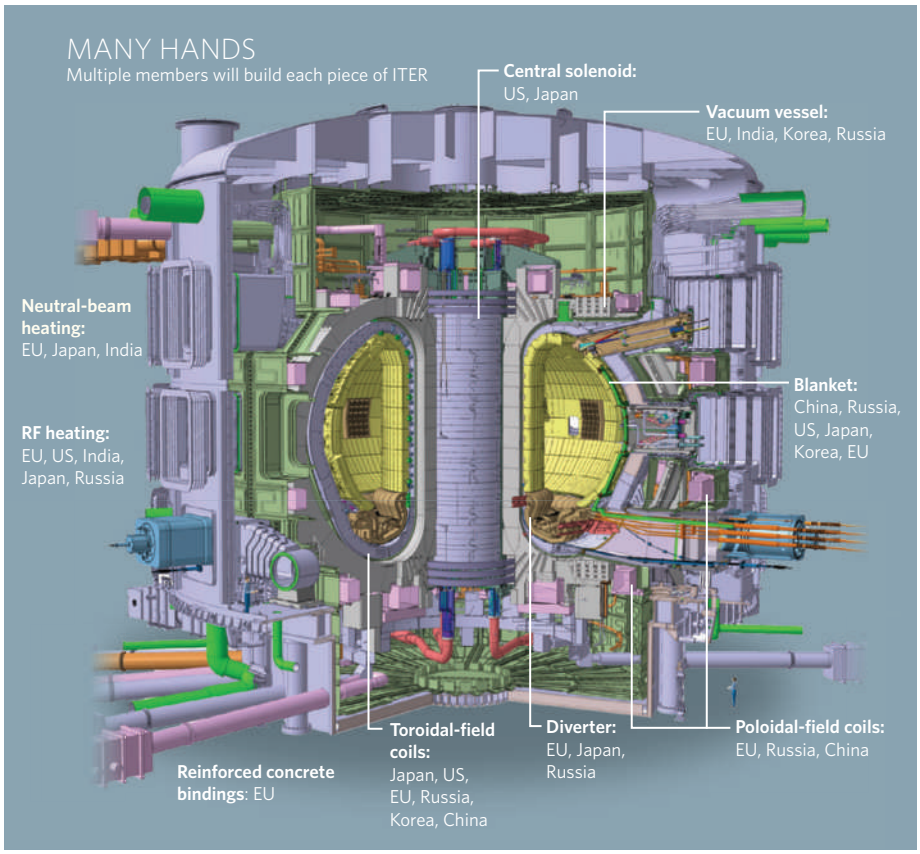
Additional reporting by David Cyranoski

See Editorial, page 483.

INTERNATIONAL CONTRIBUTORS TO ITER



ITER



NEWS

Swine flu attention turns to the tropics

New flu strains are more likely to arise in equatorial countries, where influenza is present the year round and surveillance is poor.

With the influenza season over in the temperate Northern Hemisphere, and just getting under way on the other side of the world, scientists are watching the A(H1N1) swine flu virus to see where it goes next and whether it will reassort with other flu viruses, or mutate, to cause more severe disease or acquire resistance to antiviral drugs.

Some researchers are warning, however, that such changes might be more likely to occur not in the northern or southern temperate zones where flu is seasonal, but in the narrow, often-overlooked belt of tropical countries where flu circulates all year round.

"We should be getting the message across that it is probably in the tropical countries, more than in the Southern Hemisphere, that this virus will be going through some reassortment contortions in the coming months," says Ken Shortridge, a veteran of flu research in China and southeast Asia, now retired in New Zealand.

Outbreaks of the new virus have so far been largely confined to the Northern Hemisphere, and public-health officials are crossing their fingers that as summer approaches outbreaks may wane — although that is far from a given. That would buy time for a new vaccine to become available just before the expected wave of new cases next winter.

European Union countries have so far succeeded in holding the virus at bay by intensively tracking the first cases imported from abroad, and treating those patients and their contacts with antiviral drugs. Many low-income countries in the south will not have the resources for such a strategy, says Jeremy Farrar, director of the Oxford University Clinical Research Unit in Ho Chi Minh City, Vietnam. "This inevitably means that when this highly transmissible virus arrives in a densely populated country with more limited resources, it will spread rapidly and many, many thousands of people will be infected," he says.

Scientists have been frustrated by their inability to predict the behaviour of flu viruses, although they are more prepared now for a pandemic than ever before. When the last flu pandemic occurred, in 1968, researchers did not have the web or the molecular-biological and genomic tools to track the disease. "It's the first time in history that we have been able to watch a potentially pandemic virus unfold in advance of a pandemic," says Shortridge. "This is a magnificent advance, but we are still feeling our way."

Jimmy Smith, a member of the World Bank's global H1N1 and H5N1 flu task force in Washington DC, says there are three immediate priorities: "Surveillance, surveillance and surveillance." That includes not only detailed epidemiological analysis of human cases and their clinical spectrum and age profiles, but also more detailed monitoring of animal flu viruses, in particular in pigs and poultry. Intensive surveillance of cases will need to be sustained for at least two years before any conclusions can be drawn as to the severity of the virus, says Keiji Fukuda, acting assistant director-general of the World Health Organization's (WHO's) Health Security and Environment division. "Pandemics do not occur in a couple of days," he says.

In the tropics and the Southern Hemisphere, the infrastructure for surveillance capacity in low-income countries has improved greatly over the past five years because of the need to deal with the threat of H5N1 avian influenza, including expanding the WHO's Global Influenza Surveillance Network. But progress has been slow, and there is still insufficient capacity in many countries, particularly in rural Asia and Africa. The World Bank is in the process of ramping up funding for flu surveillance for Mexico and any other countries that request it, says Smith.

> "This is a magnificent advance, but we are still feeling our way."

Surveillance is not just academic; it is key to getting early warnings of events that call for swift adaptations of control strategies. Swine flu is currently sensitive to the antiviral drug oseltamivir (Tamiflu), for example, but seasonal H1N1 is resistant. Were the new virus to acquire resistance, that would render redundant the Tamiflu stockpiled by many nations as part of their pandemic plans.

Tropical countries may be pivotal in such changes. "Tropical zones are the black box of influenza," says Antoine Flahault, dean of the French School of Public Health in Rennes and Paris. Tropical southeast Asian countries in particular are a hotbed of flu viruses, ranging from H9N2 in children in Hong Kong and China, to H6N1 in birds and, by far the most prevalent, H4N6 in ducks and other waterbirds. The threat of reassortments occurring in poor tropical countries is compounded by the fact that humans tend to live in close proximity with livestock. "If it reassorts, who knows what could result," says Shortridge.

Influenza outside the developed countries of the Northern Hemisphere is often overlooked.

A. BIRAJ/REUTERS

Change in focus

In a 1988 workshop, Shortridge, along with Alan Kendal, then director of influenza at the US Centers for Disease Control and Prevention in Atlanta, Georgia, first drew attention to the tropics as likely to be the main reservoir for the mixing of flu viruses. Nonetheless, the tropics are largely relegated to an afterthought in textbooks and papers about flu, with most attention being given to seasonal flu in the temperate regions, where most flu researchers work. "Articles about flu in the tropics tend to get glossed over," says Shortridge.

Two flu genome papers[1,2] published last year, however, provide strong genetic evidence for mixing and mutation in the tropics, to yield new future strains. "All the interesting stuff happens in these tropical reservoirs, with the temperate epidemics in both the Northern and Southern Hemispheres being spillovers from that," says Oliver Pybus, an evolutionary geneticist at the University of Oxford, UK, and co-author of one of the papers.

The logical conclusion, he says, is that although surveillance is needed in temperate countries in the Southern Hemisphere during the upcoming flu season, deep sampling of tropical regions would be even more helpful. "But unsurprisingly, surveillance has historically been best in rich countries" that tend to be in temperate latitudes, he says.

On 22 May, Margaret Chan, director-general of the WHO, told the annual meeting of the World Health Assembly — the agency's supreme decision-making body — that nations must unite to help poor countries in the Southern Hemisphere implement control measures. "I strongly urge you to look closely at anything and everything we can do, collectively," she said, " to protect developing countries from, once again, bearing the brunt of a global contagion." ■

Declan Butler

1. Rambaut, A. *et al. Nature* **453,** 615–619 (2008).
2. Russel, C. A. *et al. Science* **320,** 340–346 (2008).

For *Nature*'s swine flu coverage, see nature.com/swineflu

CDC

H1N1 update

As confirmed by the World Health Organization, as of 26 May:
Cases: 12,954, in 46 countries
Deaths: 92, in 4 countries

AUSTRALIA: Raised the pandemic level to 'Contain' on 22 May, giving authorities the power to close schools.
JAPAN: Urged people not to panic as case numbers reached 350 and spread to Tokyo.

S. DORONENKO

FORESTS AT RISK
International preservation targets will not be met by 2010.
www.nature.com/news

North Korea's bigger blast

On the morning of 25 May, when North Korea announced that it had tested a second nuclear device, supporting evidence was already available in the form of waves picked up by seismic stations around the world within minutes. But piecing together what the blast means in terms of the nation's nuclear capacity will take a lot longer.

At 09:55 local time on 25 May, monitors in the Shimane and Tottori prefectures on the Japanese coast picked up seismic waves emanating from the region where North Korea tested a nuclear weapon in 2006 — an area with little natural seismic activity. Movements along natural fault lines transmit most of their energy through 's-waves', whereas explosions at a single point release a greater proportion through compressional p-waves. In the waves detected in Japan, the s-wave component was just one-fifth that of the p-wave. "You can't say it's impossible for a natural earthquake, but it would be very rare," says Gen Aoki of the Japan Meteorological Agency in Tokyo.

Japanese officials determined the tremors to be of magnitude 5.3. The US Geological Survey put the strength at 4.7, and the Comprehensive Test Ban Treaty Organization (CTBTO) recorded a magnitude of around 4.5. All three estimates are consistent with a blast equivalent to a few thousand tonnes of TNT.

The yield of North Korea's 2006 test was estimated to be under a kilotonne, and most experts think that it was a 'fizzle' — a detonation in which the nuclear device fails to create a complete chain reaction in its fuel. North Korea's weapons programme uses plutonium, which needs to be compressed with carefully synchronized explosive charges to detonate properly.

Linking the magnitude readings to the explosive power depends on knowledge of the surrounding earth, but the latest test seems to have been around five times the size of the 2006 test, according to Paul Richards, a seismologist at Columbia University in New York. Richards says that the blast was too large to have come from conventional explosives. "To me, a chemical explosion on the order of a few thousand tonnes, with all the stuff detonated at all the same instant, is not a credible scenario," he says.

Martin Kalinowski, a nuclear physicist at the University of Hamburg in Germany, agrees, but says that the test may not have been a complete success. He was expecting a yield of roughly 16 kilotonnes, in line with the first nuclear tests of other nations. Assuming that was the case, then Monday's test "was probably not the yield that North Korea expected", he says.

Others, however, think that North Korea wanted a small blast. In 2006, the nation told the Chinese that its bomb would have a yield of around 4 kilotonnes, according to Andreas Persbo, a senior researcher at

the Verification Research Training and Information Centre (VERTIC) in London. Given the limited amount of nuclear material available to the North Koreans, a small yield might be a sensible choice. "It certainly looks successful to me," he says.

Further confirmation of the nuclear nature of the test is likely to come in the next few weeks. If the underground blast was indeed nuclear, then it will have released radioactive xenon isotopes in ratios that are distinct from those released by other sources, such as nuclear power stations. A representative at Japan's Ministry of Foreign Affairs says that its two radionuclide monitoring stations will not have any results for a couple of days; ground-based CTBTO testing stations in Russia could see the first xenon within the next few days, although it will take weeks to complete the measurements, Kalinowski says. Specially equipped US military aircraft are also likely to be used in the follow up. ■

Geoff Brumfiel and David Cyranoski

Marmoset model takes centre stage

Newly created transgenic primate may become an alternative disease model to rhesus macaques.

Japanese researchers this week report the passing of a transgene from a primate to its offspring[1,2] (see pages 515 and 523). The work could establish marmosets as a model research organism to rival the more commonly used rhesus macaque, and usher in a new era of primates as human-disease models.

Erika Sasaki and her colleagues at the Central Institute for Experimental Animals in Kawasaki injected viral vectors with green fluorescent protein (GFP) into 91 marmoset embryos, then transferred the 80 healthy transgenic embryos to surrogate mothers. Five offspring were born — including twins Kei and Kou; keikou means fluorescence in Japanese — all of which expressed the glowing transgene in some features at some point during development. Most exciting, says Sasaki, was the birth in April of a male produced by conventional *in vitro* fertilization using Kou's sperm. Since then, two more glowing second-generation marmosets have been born, although one died after being bitten by his mother.

"This is a great advancement, and it will bring more attention to primate models from people who don't normally think about primates," says Anthony Chan, a geneticist at the Yerkes National Primate Research Center of Emory University in Atlanta, Georgia.

Biomedical reseachers have long wanted primate models that can inherit and express introduced genes, as mice can. In 2001, a group led by Chan and Gerald Schatten, then at the Oregon Regional Primate Research Center in Portland and now at the University of Pittsburgh, Pennsylvania, introduced GFP into a rhesus macaque oocyte[3] and produced the animal named ANDi. In 2008, Chan reported rhesus macaques with the Huntington's disease gene[4].

"Marmosets will be better as models than rodents, but are they good enough?"

But passing transgenes to the next generation — the key to making a useful research model — had never been done. ANDi has not had any luck with natural breeding — "perhaps owing to his short stature and relatively gentle demeanour", says Schatten, who is now working instead on intracytoplasmic sperm injection and cloning techniques.

With Chan's Huntington's monkeys, all but one of the first generation died from an early and extreme manifestation of the disease. Since then, three more have been born; Chan must wait another year or two to breed them.

Sasaki's marmosets could move more quickly into disease modelling. Her first target is Parkinson's disease, but she is considering amyotrophic lateral sclerosis and Huntington's.

Other Japanese researchers are looking to expand on the work through the country's project in brain sciences, a 5-year programme for which 8 universities and institutes get an annual ¥600 million (US$6 million) to develop and use a "highly original model research animal". Tetsuo Yamamori, at the National Institute for Basic Biology in Okazaki, plans to introduce genes into the marmosets that control the expression of *OCC-1*, which he thinks is involved in vision.

CENTRAL INSTITUTE FOR EXPERIMENTAL ANIMALS

Erika Sasaki (centre) with transgenic marmosets Hisui, Banko, Wakaba, Kei and Kou.

Monkey magic?

Marmosets are small and relatively easy to handle. They reach sexual maturity in just over a year and a female can have up to 80 babies, compared with about 10 for the rhesus macaque. But the bottom line is whether marmosets can usefully model human disease. "They'll be better than rodents," says Chan. "But are they good enough?"

Marmosets, a New World monkey, are genetically further from humans than macaques. Nicole Déglon, of the French Atomic Energy Commission's Department of Medical Research in Gif-sur-Yvette, says that normal marmosets fail many cognitive ability tests that are used to judge the effects of conditions such as Alzheimer's disease. She says she stopped using marmosets because their brains are too small for positron emission tomography scans.

Tadashi Isa, a developmental physiologist at the National Institute for Physiological Sciences in Okazaki who is also a member of the brain-science programme, says marmoset research is likely to boom, but only as a complement to macaque research. Marmosets, he says, will be useful for studying cognitive and behavioural disorders because they share some social characteristics with humans, such as maintaining familial relationships. "But a higher brain function research paradigm has been established for things like attention and decision-making in the macaque," he says. "With marmosets, there is no history and no accumulation of knowledge."

The transgenic marmoset project is also likely to face challenges from bioethicists and animal-rights groups over creating a colony of animals born with a disease. "You must think of the cost–benefit," Sasaki responds. "If you can do it in rodents or *in vitro*, we should. But for many diseases, like Parkinson's, there is not a good model now." ■

David Cyranoski

1. Sasaki, E. *et al. Nature* **459,** 523–527 (2009).
2. Schatten, G. & Mitalipov, S. *Nature* **459,** 515–516 (2009).
3. Chan, A. W. S., Chong, K. Y., Martinovich, C., Simerly, C. & Schatten, G. *Science* **291,** 309–312 (2001).
4. Yang, S.-H. *et al. Nature* **453,** 921–924 (2008).

See Editorial, page 483, News & Views, page 515, and Article, page 523.

US climate legislation advances

Landmark legislation to reduce US greenhouse-gas emissions advanced through a key committee on 21 May in the face of staunch Republican opposition.

The House of Representatives Energy and Commerce Committee approved the legislation on a 33–25 vote, with 4 Democrats voting against and 1 Republican, Mary Bono Mack (California), voting aye. The bill would employ a cap-and-trade system to reduce US greenhouse-gas emissions to 17% below 2005 levels by 2020 and upwards of 80% below by mid-century.

Committee chairman Henry Waxman (California) crafted the bill with fellow Democrat Edward Markey (Massachusetts). With President Barack Obama advocating aggressive action on global warming, this is the first time lawmakers have struck a deal that could actually become law.

"This is much bigger than a committee vote," says Tony Kreindler, a spokesman for the advocacy group Environmental Defense Fund in Washington DC. "Waxman and Markey managed to broker a deal between some very diverse industrial interests. They threaded the needle."

The deal comes amid an economic crisis that has damaged businesses and left millions of people jobless. Last week the Energy Information Administration reported that because of declining industrial activity and high energy prices, US carbon dioxide emissions from fossil fuels last year fell by 2.8%, to 5.8 billion tonnes.

Democrats Edward Markey and Henry Waxman.

R. KELLY/CONGRESSIONAL QUARTERLY/GETTY IMAGES

Opponents say that the US economy is too fragile for a regulatory shake-up, but Democratic leaders in the House of Representatives could bring the bill up for a full vote as early as this summer.

Building on a compromise endorsed by the US Climate Action Partnership — which represents numerous environmental groups, utilities and other business interests — Waxman and Markey spent weeks brokering deals with moderate Democrats from conservative energy-producing states such as Illinois, Texas and Virginia. They scaled back the 2020 target from 20% to 17% and agreed to give away roughly 85% of the emission allowances in the early years of the programme.

Obama had pushed for 100% auctioning, an approach that raises revenue and forces companies to pay to be able to pollute. But giving the allowances away in the early years is an easy way to ease the pain of new regulations; the deal would allocate 35% of the allowances directly to local electricity providers to try to protect consumers from a potential price spike.

"The House energy and commerce committee really is a microcosm of the whole Congress," says Scott Segal, an industry lobbyist with law firm Bracewell and Giuliani in Washington DC. "To produce a bill, there had to be substantial discussions with a number of moderate Democrats who have unique regional or industrial concerns. It needed that kind of seasoning."

The compromise held over four days as Democrats shot down dozens of hostile amendments from Republicans blasting the legislation as a 'cap-and-tax' programme. Waxman even hired a speed-reader, in case the Republicans forced the Democrats to read the entire 932-page bill aloud. (They didn't.)

The sweeping legislation would require utilities to produce 20% of their electricity from renewable sources by 2020, although a quarter of that could be met through improvements in energy efficiency. It would also advance clean-energy research and development, pump billions of dollars into international efforts to curb deforestation and create separate regulations for the powerful greenhouse gases hydrofluorocarbons.

The vote came on the heels of another historic deal on vehicle regulations announced by the Obama administration on 20 May. With many US companies already dependent on federal financial support, the automobile industry agreed to back a new federal regulation requiring a roughly 40% boost in fuel efficiency compared with today — to more than 35.5 miles per gallon (15 kilometres per litre) by 2016, four years earlier than previously scheduled. ■

Jeff Tollefson

K. CAMPBELL/GETTY

Former shuttle pilot nominated as NASA head

US President Barack Obama announced on 23 May that he is nominating Charles Bolden, a retired Marine general and former space-shuttle pilot, as NASA administrator.

The announcement came as the space shuttle *Atlantis* was finishing the last ever mission to the Hubble Space Telescope — fitting, in its way, as Bolden was the pilot on the mission that launched the telescope in 1990. Speaking to the *Atlantis* astronauts on 20 May, after meeting with Bolden, Obama said: "It's a high priority of mine to restore that sense of wonder that space can provide."

Filling the administrator's post — the position requires confirmation by the Senate — will end months of speculation and delay, during which Bolden's name has been bandied around along with those of other candidates. He has been a favourite of senator Bill Nelson (Democrat, Florida), whom Bolden piloted into space in 1986 on another of his four shuttle flights.

After graduating from the US Naval Academy in 1968 with a degree in electrical science,

USMC

Charles Bolden is in line to take over at NASA.

Bolden began a career as an aviator in the US Marine Corps, becoming an astronaut in 1981. In early 2002, then-administrator Sean O'Keefe picked him to be his deputy, but Bolden wanted to keep his commission while serving at NASA, and there were concerns in the Pentagon and Congress over a military officer serving in a civilian position. Bolden's name was withdrawn, and he retired six months later and moved to Houston, Texas, where he became president of a freshwater aquifer development company and served on the boards of oil and aerospace companies.

His views on science are relatively unknown, but Bolden served on a National Academies panel that endorsed a human rather than robotic final servicing of the Hubble Space Telescope. "He was very enthusiastic about the Hubble science," says Louis Lanzerotti, a physicist at the New Jersey Institute of Technology in Newark and the panel's chair. "He seemed to have followed it since he launched the Hubble. He was very alert and inquisitive about the results that were presented to our committee."

Lori Garver, who was a space-policy adviser to the Obama campaign, has been nominated as Bolden's deputy. ■

Eric Hand

Russia shifts stance on climate-change policy

Russia's government has quietly made a dramatic change to its policy on climate change, accepting that anthropogenic global warming poses severe risks and requires immediate action to limit carbon emissions.

"Russia's diplomatic approach to [December's scheduled climate talks in] Copenhagen was until now just one big silence," says Kristin Jørgensen, a climate-policy expert with Bellona, an environmental watchdog based in Norway that has a network of activists in Russia. "This is a totally surprising move. There were no hearings, no stakeholder discussion, no public debate — just nothing."

Policy analysts believe that the new climate 'doctrine', adopted in late April, marks a historic turning point. Principally a position statement, the doctrine also outlines a checklist of key climate actions. It follows a February assessment report compiled by leading Russian climate scientists, which was presented to the Russian cabinet on 23 April.

For a longer version of this story, see http://tinyurl.com/q2tyjn.

United States urged to boost global health funds

An expert committee convened by the US Institute of Medicine has called for the country to increase investment in global health initiatives to $15 billion per year by 2012.

Earlier in May President Barack Obama asked Congress to dedicate an average of $10.5 billion per year over the next six years to a global health initiative, with $8.5 billion of that funding dedicated to the HIV/AIDS programme PEPFAR (see *Nature* 457, 254–256; 2009).

But the committee, which included former National Institutes of Health head Harold Varmus, recommended that $13 billion be invested in fulfilling health-related Millennium Development Goals put forward by the United Nations, with another $2 billion for combating injuries and non-communicable conditions, such as heart disease. In addition, it advised that an inter-agency global health committee be created and located in the White House to coordinate such activities.

JOHN MADDOX

A memorial meeting to celebrate the life of Sir John Maddox, the editor of *Nature* for many years, will be held on Friday 5 June 2009 at 6 p.m., followed by a reception, at The Royal Institution, 21 Albemarle Street, London W1S 4BS. Those wishing to attend should notify Diane Kempinski at *Nature*, d.kempinski@nature.com. Places are limited.

THAT FOSSIL FRENZY IN FULL

"This specimen is like finding the lost ark for archaeologists. It is the scientific equivalent of the Holy Grail."

Jørn Hurum, University of Oslo, at the New York press unveiling of a 47-million-year-old fossil, christened Ida (pictured), last week.

"[It is] like the eighth wonder of the world."

Jens Franzen, Natural History Museum of Basel, Switzerland, who was involved in analysing the fossil.

"Any pop band is doing the same thing. We have to start thinking the same way in science."

Hurum justifies the attention.

"There was a TV company involved and time pressure. We've been pushed to finish the study; it's not how I like to do science."

Hurum's co-worker Philip Gingerich of the University of Michigan, Ann Arbor, reflects.

J. L. FRANZEN *ET AL. PLOS ONE* **4,** E5723 (2009)

For more on Ida see http://tinyurl.com/idafossil and Editorial, page 484.

Sources: *The Wall Street Journal*, *The New York Times*, BBC, Fox News, *The Guardian*

UK geographers vote against large expeditions

Members of London's Royal Geographical Society have thrown out a resolution to resume large exploratory expeditions. Campaigners say that the vote highlights growing division within the world's biggest scholarly geographical society.

The society stopped organizing large expeditions such as open-ended explorations of rainforests, after reviews of research practices in 2001 and 2004 suggested that smaller, more focused projects were the best ways to tackle global problems such as climate change and the security of food and water. "Twenty-first-century geography, not nineteenth-century geography, is what we're talking about," says Gordon Conway, the society's president.

The resolution was defeated by 2,590 votes to 1,607 on 18 May. Supporters of the 'Beagle Campaign' that put forward the resolution say they will continue to advocate their cause.

For a longer version of this story, see http://tinyurl.com/pbx929.

Delays to satellite launches put GPS at risk

Concern over the future performance of the US Global Positioning System (GPS) went up a notch last week as a government watchdog official warned that the US Department of Defense faced substantial challenges meeting its space-programme commitments.

A report from the Government Accountability Office (GAO) on 30 April had cautioned that new GPS satellites might not be launched in time to replace the ageing constellation that is currently in orbit. And on 20 May, Cristina Chaplain, GAO director of acquisition and sourcing, told the Senate committee on armed services that cost overruns of space programmes are part of the problem.

Dave Buckman, of the US Air Force Space Command, quickly replied on a Twitter feed that "GPS isn't falling out of the sky". Still, a temporary decline in performance might cause a problem for scientists who rely on GPS-positioned equipment that cannot be easily upgraded, such as low Earth-orbiting satellites, says Marek Ziebart, a space geodesy researcher at University College London.

For a longer version of this story, see http://tinyurl.com/gpsthreat.

Corrections

The News Feature 'The sleeping dragon' (*Nature* **459,** 153–157; 2009) misstated the number of landslides thought to result from the 2008 Sichuan earthquake. It should have said that scientists have identified at least 15,000 resulting landslides and rock avalanches, and perhaps as many as 50,000 or more.

The News story 'Even big societies feel the pinch' (*Nature* **459,** 17; 2009) cited incorrect information, provided by the American Chemical Society, that the society had already posted a $36.5-million bond in a lawsuit it recently lost. The society expects the bond will be posted by the first week of June.

THE MAN WHO SMELLS FORESTS

Chemist Allen Goldstein has spent his career tracking elusive compounds emitted by trees. **Erik Vance** joined him for a tour of the woods.

R. REID/NATIONAL GEOGRAPHIC/GETTY

Standing almost 20 metres above the forest floor on a scaffolding tower in the Sierra Nevada, Allen Goldstein looks over the spiky tops of a young ponderosa pine forest. At ground level, the air is warm, still and rich with the sweet smell of pine and cedar. But above the forest, a stiff breeze from the west causes the tower to sway disconcertingly.

Goldstein inhales deeply before explaining the forest's daily chemical rhythm. At sunrise, the trees start pumping out a complex mix of volatile organic compounds (VOCs), such as pine-scented terpenes. By mid-morning, the westerly breeze adds a dose of the VOC isoprene from oak woodlands about 30 kilometres away. Then, as the sun reaches its peak in the sky, pollution from California's Central Valley makes its way up into the mountains.

A chemist at the University of California, Berkeley, Goldstein specializes in interpreting the scents of the forest. He has built his career on finding and characterizing some of the more elusive airborne chemicals in nature. For 10 years at this site near the University of California's Blodgett Forest Research Station (see map), he and his team have described more than a dozen plant-released compounds that no one had previously measured or, in some cases, even known existed in the atmosphere.

Working at the tops of the trees with ever more sensitive detectors, he has found that forests play a crucial part in the chemistry of aerosol particles and with pollutants such as ozone. His discoveries may help to fill in confounding gaps in atmospheric science, such as how VOCs from plants affect air quality and how they influence climate.

"Where Allen has really made his mark is in new ways of making measurements — basically seeing things that you have never seen before," says atmospheric scientist Inez Fung, a colleague at Berkeley.

Goldstein, now in his mid-forties, has broad shoulders, a quick smile, and an unpretentious demeanour. He is popular with students and easily becomes animated when talking chemistry. He is also popular with faculty members, although this might have something to do with his hobby as a winemaker. Fung says that recent vintages are "getting better" — other colleagues rave about them.

Goldstein came to Berkeley in 1996 from Harvard University, where he developed one of the first devices that could track VOCs from plants continuously throughout the day. Testing his device in Berkeley, it was sensitive enough to pick up the morning spike in airborne caffeine from nearby cafes, as well as traces of marijuana.

But Goldstein was not interested in tracking people's vices. His focus was the many types of VOCs emitted by forests. Estimates suggest that the vast majority, perhaps up to 90%, of Earth's VOC output comes from vegetation[1], a fact infamously alluded to by US president Ronald Reagan, when he said that forests pollute more than cars.

Goldstein scowls when asked about that statement, in part because it conflates natural emissions with car exhaust and also because journalists always ask about it. But in a way, Reagan's muddled formulation captures Goldstein's primary interest: the relationship between human pollution and plant emissions.

Goldstein often says he went into chemistry not purely for science, but to make a tangible difference to society (he originally planned on being a lawyer). "I've always tried to pick scientific problems that were interesting and challenging," he says, "but also had relevance to how we manage the world around us."

Spare the pores

The term VOC is a bit of a chemical catch-all, encompassing everything from exhaust fumes and the stench of solvents to the bouquet of Goldstein's wines. Some VOCs are highly reactive, existing in nature for just moments. Others float around for years.

At the Sierra Nevada research site, they were a solution to a mystery. When ozone pollution from more populous regions to the west blows into the forest, a significant fraction of it seems to disappear. For years, scientists thought almost all of this missing ozone went into trees through pores in their leaves called stomata, where it could cause damage. But in his first Blodgett readings in the late 1990s Goldstein quickly discovered that in the summer the trees absorbed at most only a third of the missing ozone[2].

High and mighty: Allen Goldstein works in the treetops of California's Sierra Nevada.

E. VANCE

A. GOLDSTEIN

The key, he learned, was that when temperatures increased, more ozone went missing. To Goldstein, this implicated VOCs because warmer weather causes plants to release more VOCs (which is why forests smell more pungent on hot days). He found that some of these VOCs react in seconds, quickly stripping oxygen atoms from ozone[3]. The now oxygen-rich molecules get heavier and stickier, forming aerosols that can create haze over a forest.

Joost de Gouw, an atmospheric chemist with the National Oceanic and Atmospheric Administration in Boulder, Colorado, says that Goldstein's work linking VOCs and these aerosols was a crucial breakthrough[4]. "He was one of the first to start thinking about VOCs and organic aerosols as a whole," says de Gouw, leading to what he calls the provocative idea "that the sources of organic aerosol are much [larger] than we think".

"Goldstein's more recent wine vintages are getting better."

Of the 1.3 billion tonnes of VOCs released by humans and plants every year, conservative estimates say as little as 12 million become aerosol particles[4]. But Goldstein estimates that 150–200 million tonnes end up converted into aerosol. "That creates quite a chemical soup that we are not keeping track of," he says.

Furthermore, many of these compounds are good at hiding. For example, estragole (methyl chavicol), a liquorice-like aromatic found in aniseed, is a semi-volatile compound — at room temperature it could be either a gas or a particle. This unpredictability means that Goldstein has to make one detector able to detect the compound in either state. It also means that the stuff is prone to sticking to the inside of tubes before it can be measured. Once Goldstein and his students worked out how to detect it, they found that it is fairly common in the air around ponderosa pines. Near his site, Goldstein picks a few needles from a ponderosa, crushes them between his fingers, and sniffs. Sure enough, there is an unmistakable hint of aniseed.

Planetary power

Now Goldstein is starting to look at the global picture. Atmospheric modellers say that, on the planetary scale, the climatic impacts of aerosols from plants are unknown. Such aerosols may exert a cooling effect, but scientists know little about the strength of that cooling and how it might change in the future.

Understanding what is happening is crucial, says de Gouw. "If models don't have the right emissions going into them, then they don't have the right chemistry and they don't have the right products coming out of them."

It is difficult to tell how much of the aerosol blanket over a given region comes from plants, and estimates vary widely[5]. Fung, Goldstein and their colleagues looked at aerosols that spike in the southeastern United States during hot summer days. The aerosols could not be explained by man-made sources alone and indicate, just as in Blodgett Forest, that plants are responsible for a large fraction. These aerosols in the southeast block enough sunlight that they cool the region, the researchers reported this month[6].

Although much of his work during the past decade has focused on the Blodgett site, Goldstein is getting ready to say goodbye. This summer, he will conduct his last field campaign there and then will box up his towers and shut down the site. When he first came there, Goldstein was a young professor. Now he is head of his department and well respected in his field. Looking around at the well-used equipment shacks, he seems a little sad.

"It's time to move on," he says. "I am moving out of the mountains and into the valley." By this he means he is moving from crisp mountain air to some of the worst air quality in the country. Arvin, a small Central Valley town near which Goldstein might soon set up a tower, exceeds federal ozone levels for a record one-fifth of the year. Goldstein will be looking at the emissions in the Central Valley, hoping to learn where they come from and how to minimize them.

It is one of several projects he is starting that might ask uncomfortable questions of the state's farmers. In many regions, VOCs from plants react with nitrogen pollution to form ozone and Goldstein wonders how much of the ozone in the Central Valley is due to the almonds, pistachios, oranges, lemons and grapes grown in the area.

These are the questions Goldstein ponders as he inspects cables in preparation for the summer experiments at Blodgett. He says the chemistry of trees and the chemistry of man are linked in ways that scientists are only beginning to understand. But there's no time to dwell on this topic today. He has to get out of the forest, down the hill, and back through the Central Valley before the traffic hits. ■

Erik Vance is freelance science writer in Berkeley, California.

1. Guenther, A. *et al. J. Geophys. Res.* **100,** 8873–8892 (1995).
2. Kurpius, M. R. & Goldstein, A. H. *Geophys. Res. Lett.* **30,** 1371 (2003).
3. Lee, A. *et al. J. Geophys. Res.* **111,** D07302 (2006).
4. Goldstein, A. H. & Galbally, I. E. *Environ. Sci. Technol.* **41,** 1514–1521 (2007).
5. Kanakidou, M. *et al. Atmos. Chem. Phys.* **5,** 1053–1123 (2005).
6. Goldstein, A. H., Koven, C. D., Heald, C. L. & Fung, I. Y. *Proc. Natl Acad. Sci. USA* advance online publication doi: 10.1073/pnas.0904128106 (2009).

Emissions control

Could hydrogen sulphide be the new nitric oxide? **Katharine Sanderson** reports on the rotten-egg gas that is earning a reputation in human physiology.

When Rui Wang saw the painted Easter eggs his ten-year old daughter had brought home from school in 1998, he was unaware of the stink they were about to create. The eggs were proudly displayed as objects of art in the family's glass cabinet. Then one day, says Wang, "I came home and the whole house was filled with a stinky smell". One of the eggs had fractured, and the distinctive whiff of hydrogen sulphide filled the air.

What might be garbage for one man was inspiration for Wang, then a cardiovascular researcher at the University of Saskatchewan in Saskatoon, Canada. He was already studying nitric oxide and carbon monoxide, gases produced in tiny quantities by the body that have potent physiological effects. The stench emanating from the cracked egg coincided with Wang's desire to seek more gases that are important in human biology. "Hydrogen sulphide jumped out from both my home and my head," says Wang, now working at Lakehead University in Thunder Bay, Ontario. He quickly bought a canister of the gas and applied it to his cultures of rat vascular tissue. The work culminated in the discovery that hydrogen sulphide is made by the tissue, and can lower blood pressure[1].

> **"Somewhere in us the hydrogen sulphide is still knocking about."**
> — Phil Moore

Since then, many researchers have turned their attention to hydrogen sulphide and its role in human physiology. In the past year alone, Wang and others have demonstrated how the body manufactures the gas, along with a plausible way that it could be modifying a whole range of metabolic proteins. On the clinical side, the gas has been implicated in inflammation and shown to protect the heart. And next month will see the first conference dedicated entirely to hydrogen sulphide, organized by Wang and held in Shanghai.

Some researchers now think that hydrogen sulphide is poised to have as big an impact on biology as nitric oxide did — and that impact was very large indeed. The discovery of nitric oxide's role as a signalling molecule in the cardiovascular system won three scientists the Nobel Prize in Physiology or Medicine in 1998, opened up a huge new field of physiology and, as it turned out, helped to boost the profits of pharmaceutical companies: the erectile dysfunction drug Viagra (sildenafil) is now known to work by enhancing the release of nitric oxide.

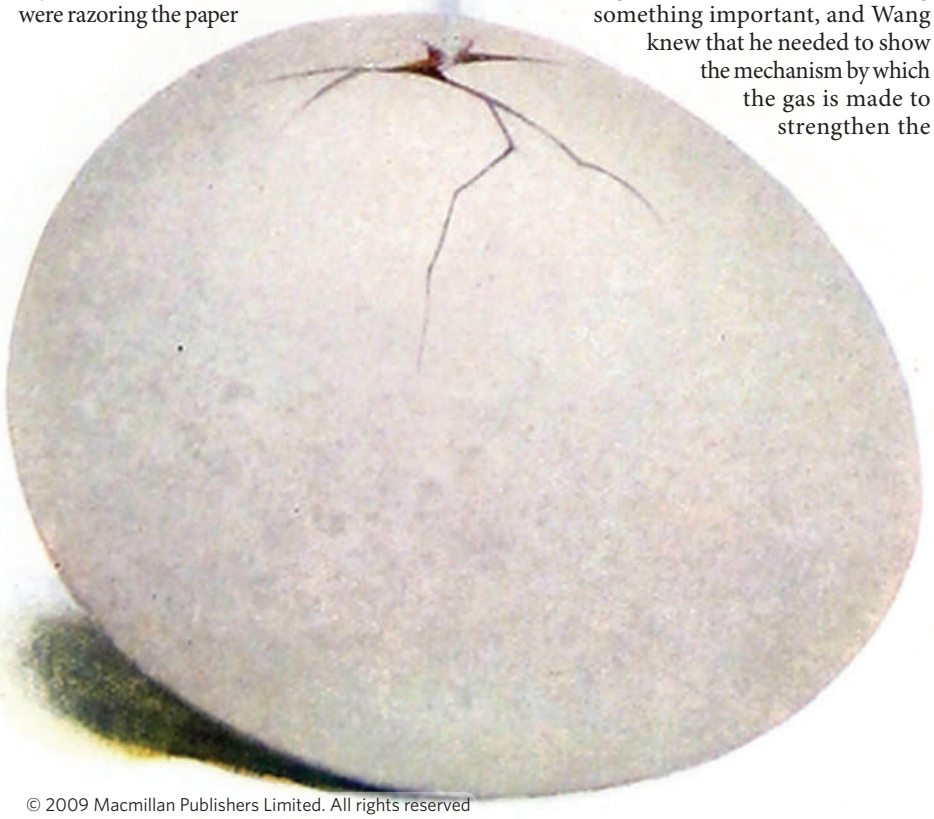
Salvador Moncada at the Wellcome Research Laboratories in Beckenham, UK, started nitric oxide's rise to fame when he showed[2] in 1987 that it accounted for the activity of a sought-after and mysterious agent known as endothelium-derived relaxing factor (EDRF), which relaxes blood vessels. The idea that cells could be making a gas in the body — particularly one known mainly as an air pollutant — was "a very strange concept", says pharmacologist Phil Moore of Kings College, London.

But once it sank in, the concept generated a tremendous thrill. Moore recalls being asked by university librarians to stop lecturing about Moncada's discovery because voracious students were razoring the paper out of library copies of *Nature*. Nitric oxide has since been established as a key signalling molecule that is involved in processes ranging from neurotransmission to immune-system regulation. And in the mid 1990s it became apparent that it wasn't the only 'gasotransmitter', when researchers showed that carbon monoxide also acted in cell signalling.

Pungent potential

Hydrogen sulphide wasn't an obvious candidate to join the list of gasotransmitters, being known mainly for its smell and toxicity (a dose of more than 700 parts per million can kill a human). Wang's discovery that hydrogen sulphide could have a beneficial role in blood vessels was a "second coming", says Moore, who immediately started working on the gas. Moore and others say that the human body's ability to use hydrogen sulphide could be an evolutionary legacy from microbes that used it as a source of nutrients: "Somewhere in us the hydrogen sulphide is still knocking about."

Knocking about isn't the same as doing something important, and Wang knew that he needed to show the mechanism by which the gas is made to strengthen the



ILLUSTRATED LONDON NEWS LTD/M. EVANS

case that it is vital in mammalian physiology. In 2004, he contacted neuroscientist Solomon Snyder at Johns Hopkins University in Baltimore, Maryland. Back in 1990, Snyder had discovered an enzyme responsible for producing nitric oxide[3], and Wang wondered whether Snyder could help him do the same for hydrogen sulphide.

Wang arrived in Snyder's lab already armed with mice that he had engineered to remove the enzyme cystathionine-γ-lyase (CSE). This enzyme is involved in hydrogen sulphide production in bacteria, and Wang hoped that the mutant mice would show it could do the same in mammals. They did. In October last year, Wang and Snyder reported in *Science* that their knockout mice had low concentrations of the gas in the heart and suffered from hypertension[4]. The paper catapulted hydrogen sulphide into the scientific limelight, says Matt Whiteman from the Peninsula Medical School in Exeter, UK, who is looking at hydrogen sulphide concentrations in patients with arthritis and diabetes: "It has given the field more credibility."

Snyder says that nitric oxide can no longer be viewed as the only EDRF. The Nobel winners, he says, "showed that it was an EDRF, they didn't show it was the only EDRF". Snyder has evidence that hydrogen sulphide accounts for more of the vessel-relaxing activity of EDRF than nitric oxide does. The two gases seem to act in very different ways though. Nitric oxide activates an enzyme called guanylyl cyclase, initiating a chain of events that relaxes the muscles in blood-vessel walls. Hydrogen sulphide reaches the same end by activating ATP-sensitive potassium channels[1].

The work on CSE-knockout mice and subsequent unpublished findings are enough to convince Snyder that "hydrogen sulphide is potentially more exciting", than nitric oxide. Others in the field remain cautious about the risk of overstating its biological status. "It would be like when The Beatles said they were bigger than Jesus," says Whiteman.

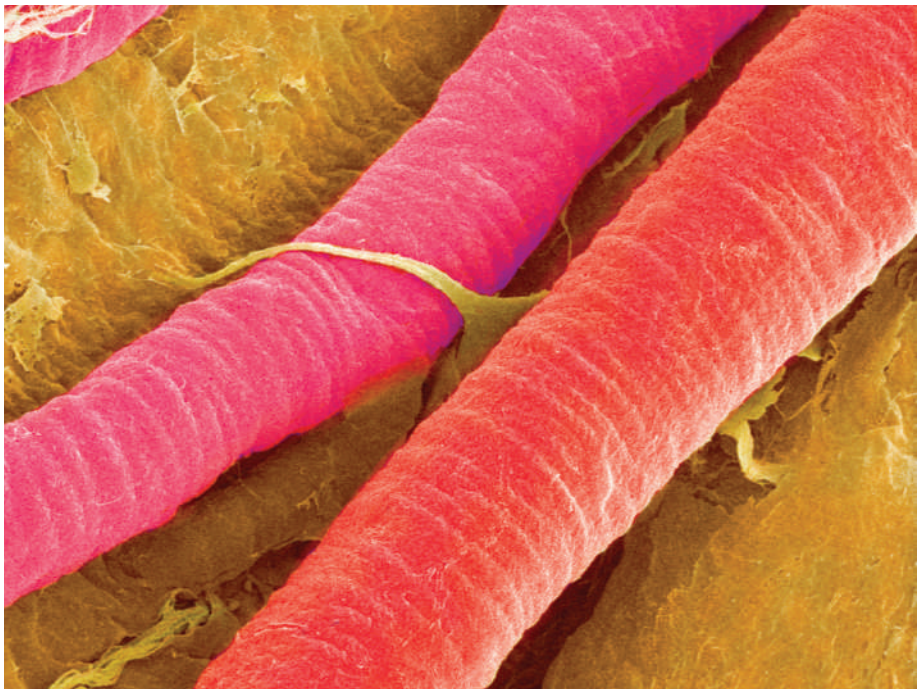
> **"It is surprising that there would be any system like nitric oxide without a backup."**
> — Giuseppe Cirino

Snyder's unpublished work could explain his faith: he thinks he has found an important mechanism by which hydrogen sulphide acts. Snyder drew on parallels with nitric oxide to work this out. Nitric oxide often acts through a process called nitrosylation, whereby a part of the nitric oxide molecule attaches itself to a specific amino acid in a protein, often inhibiting the protein's activity. That amino acid is a sulphur-containing one, called cysteine.

Snyder says that hydrogen sulphide is working by a parallel process he calls sulphydration, and that this could be an entirely new way to modify proteins and influence their activity. Sulphydration works by furnishing cysteine with an additional sulphur atom. The atom reacts with a protruding, reactive part of the cysteine that contains a sulphur–hydrogen (S–H) bond, converting it into a sulphur–sulphur–hydrogen bond that pokes even further out from the body of the protein. "It is more exposed," explains Snyder, making it accessible for further chemical reactions.

The consequences of sulphydration are more profound than those of nitrosylation, Snyder says. His tests so far have shown that nitrosylation affects roughly 1 in every 100 protein molecules in a given sample, whereas sulphydration seems to affect one in every 10–20 proteins. This could be because hydrogen sulphide is less selective than nitric oxide about which cysteines in a protein it reacts with.

Snyder says he has good evidence that sulphydration opens the potassium channels that lead to blood vessel relaxation, and that it alters the activity of at least 40 proteins in the liver. One protein he has examined closely is the enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH) which is involved in glycolysis, the process by which glucose is broken down to supply fuel to cells. When GAPDH is nitrosylated, its ability to catalyse the reaction is lowered; with hydrogen sulphide the picture is dramatically reversed. "If you sulphydrate GAPDH you increase its activity 700% *in vivo*," Snyder says. "The bottom line is sulphydration is a new mode of post-translational modification of proteins, probably comparable in prevalence and influence to phosphorylation," he says, referring to another widespread protein modification that is vital for all manner of cell functions. Snyder says that sulphydration could be important to explain the control of many metabolic processes besides glycolysis. "The regulation of metabolic pathways was largely worked out 50 years ago. The question is what turns them on and off? How do hormones and other signals regulate metabolism? That hasn't been adequately addressed."



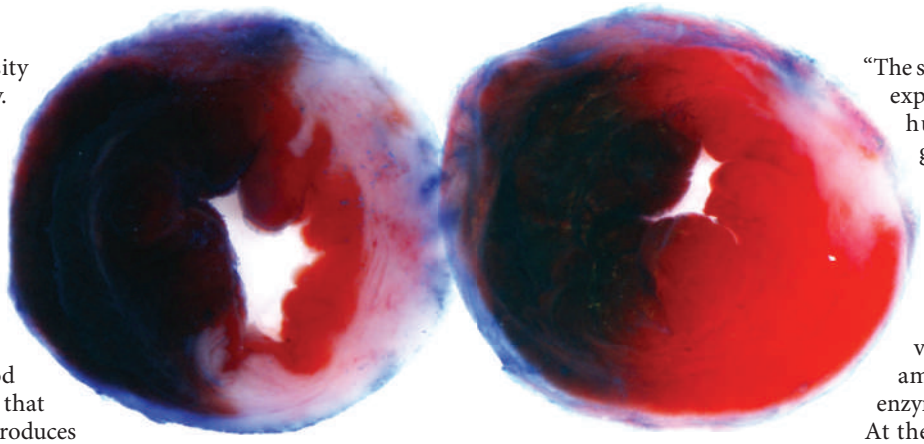
S. NISHINAGA/SPL

Like nitric oxide, hydrogen sulphide acts to relax blood vessels.

Wait and see

So far, Snyder's work has been published only in outline[5], and researchers say that they are waiting to see it in full before judging it for themselves. Whiteman agrees that Snyder's hypothesis is plausible: "What he's seeing makes sense with my observations as well," he says. Sulphydration "would be a major breakthrough," Wang adds.

But whereas Snyder puts hydrogen sulphide centre stage in cellular signalling, others in the field think that the gas could be the body's second fiddle — there to offer support when something goes wrong with nitric oxide. "It is surprising that there would be any system like nitric oxide without a backup," says Giuseppe



D. LEFER

A dose of hydrogen sulphide before a heart attack reduces the area of cardiac injury (white region, right).

Cirino from the University of Naples Federico II, Italy. "We are talking about a system that is important for survival."

When it comes to controlling vasculature, for example, Cirino has found that hydrogen sulphide becomes particularly important in the body if the endothelium of blood vessels — the thin lining that regulates blood flow and produces nitric oxide — is damaged so that nitric oxide production drops[6]. He and his collaborators studied men's corpus cavernosum, the penile tissue that fills with blood during an erection when stimulation causes the nervous system to release nitric oxide and relax the blood vessels. (Viagra mimics this by releasing nitric oxide into the corpus cavernosum.) Cirino and his colleagues took strips of smooth muscle from this tissue that lacked blood-vessel endothelium and hence weren't making nitric oxide. "Hydrogen sulphide can relax tissue in the absence of the endothelium," says Cirino. "The next question is why, and the question after that is how."

Delivery details

The why and how would be easier to answer if researchers had a way to control the delivery of hydrogen sulphide to tissues, and Moore has developed a molecule able to do so. It is an organic, water-soluble molecule that slowly releases hydrogen sulphide in a way that can generate physiological concentrations better than a direct shot of the gas itself. Moore has applied for a patent on this molecule, which he hopes to use to study hydrogen sulphide's effects on the cardiovascular system, and perhaps to deliver it therapeutically[7].

Other groups are also seeking ways to use hydrogen sulphide in the clinic. "We're most excited by the idea that low levels of hydrogen sulphide therapies have very, very significant effects in reducing cell damage," says cardiac researcher David Lefer at Emory University School of Medicine in Atlanta, Georgia. In work that has yet to be published, Lefer has used single shots of hydrogen sulphide to try to shield mice from the effects of a simulated heart attack. The gas seems to promote the production of powerful antioxidants and enzymes that protect against cell damage, he explains. "It looks like there's a programme of protection we can develop with a single exposure." Lefer says the work might lead to a therapy that could be taken by patients before major heart surgery.

Whiteman has been examining the role of the gas in inflammation. He looked at the synovial fluid from the joints of patients with rheumatoid arthritis — a chronic inflammatory disease — as well as osteoarthritis, a degenerative disease of cartilage. Those with rheumatoid arthritis had high concentrations of hydrogen sulphide in their synovial fluid, but those with osteoarthritis did not. Wang, though, says that researchers remain divided about the molecule's importance in the inflammatory process. "Some have shown pro-inflammation, some have shown anti-inflammation."

There is another intriguing side to hydrogen sulphide. In 2005, Mark Roth at the Fred Hutchinson Cancer Research Center in Seattle, Washington, showed that hydrogen sulphide could put mice into a state of suspended animation. Their metabolism slowed drastically, and this is how Roth explained the deeper-than-deep sleep they were in[8]. Snyder, however, says that this experiment has little bearing on what he has been observing.

Q. LI

Mice that are unable to produce the enzyme cystathionine-γ-lyase are hypertensive.

"The suspended-animation experiments used doses hundreds of times greater than normally exist and so likely involved mechanisms unrelated to the normal hydrogen sulphide physiology," he says. "It's very different to a tiny amount formed by an enzyme."

At the Shanghai meeting next month, which some 300 participants are expected to attend, one of the major topics for discussion will be how to translate results in animals into humans — and, perhaps, who will capitalize first on the gas's therapeutic effects. "I think that there are drugs to be had here," says Moore. He and Wang are collaborating with a company called CTG Pharma in Milan, Italy, where researchers are trying to insert hydrogen-sulphide-donating groups into known drugs, including sildenafil and the painkillers aspirin and diclofenac. Researchers at the biotechnology firm Ikaria in Seattle, Washington, are developing therapies based on all three known gasotransmitters. They are testing in phase I trials whether injections of sodium sulphide, an agent that is used to deliver hydrogen sulphide to the body, can treat conditions such as blocked arteries.

As such studies move forwards, researchers hope to learn much more about hydrogen sulphide — and more about gasotransmitters as a whole. Moore and Wang think that the human body may have the ability to use other simple gases that microbes use, and one example that Moore mentions is ammonia — another gas known for its pungent, rather than physiological, properties. This mean that Wang may be able to follow his nose to the next gasotransmitter, just as he did to the last. "If we learn anything from nitric oxide, carbon monoxide and hydrogen sulphide that would be these three gases are not alone," he says. "There must be more." ■

Katharine Sanderson is a reporter for *Nature* in London.

1. Zhao, W., Zhang, J., Lu, Y. & Wang, R. *EMBO J.* **20,** 6008–6016 (2001).
2. Palmer, R. M. J., Ferrige, A. G. & Moncada, S. *Nature* **327,** 524–526 (1987).
3. Bredt, D. S. & Snyder, S. H. *Proc. Natl Acad. Sci. USA* **87,** 682–685 (1990).
4. Yang, G. *et al. Science* **322,** 587–590 (2008).
5. Mustafa, A. K., Gadalla, M. M. & Snyder, S. H. *Sci. Signal.* **2,** re2 (2009).
6. d'Emmanuele di Villa Bianca, R. *et al. Proc. Natl Acad. Sci. USA* **106,** 4513–4518 (2009).
7. Li, L. *et al. Circulation* **117,** 2351–2360 (2008).
8. Blackstone, E., Morrison, M. & Roth, M. B. *Science* **308,** 518 (2005).

NATURE|Vol 459|28 May 2009

IDEAL FOCUS

Before they were touted as invisibility cloaks, metamaterials promised a perfect lens. **Geoff Brumfiel** reports on the struggle for superior vision.

An eye for detail: John Pendry.

M. FINN-KELCEY/IMPERIAL COLLEGE

As far as John Pendry is concerned, we are as good as blind. Sitting in his office at Imperial College London, the theoretical physicist gestures at the table in front of him. Assume, he says, that the fundamental limit of detail in the table is about the size of an atom. Then consider that the smallest feature the human eye can see, even using a high-quality optical microscope, is a few tenths of a micrometre across — roughly a thousand times bigger than an atom. That means that, even with 20–20 vision and the best optics, humans can access only 0.0001% of the information right before their eyes. "That's not much," he observes.

Nearly a decade ago, Pendry proposed a way to do thousands of times better using a film of silver just tens of nanometres thick as a 'superlens'[1]. The film would behave as a simple metamaterial, a substance with a small-scale structure that allows it to manipulate light in a way that no bulk composition could. This means that the film could capture details that elude conventional optical microscopes. In theory, the superlens could also etch nanometre-scale patterns on to a surface, which would make it very useful in the manufacture of microchips.

> "It's a very, very simple technology. It's just a question of getting it right."
> — John Pendry

Pendry's idea brought many researchers into the field, and for a while expectations were high. But nine years on, those expectations have yet to be fully realized. Although there have been proof-of-concept devices for both the superlens and the related 'hyperlens', industry has switched its focus elsewhere. And the field of metamaterials has itself been diverted by another high-profile potential application: cloaking devices.

Making objects 'invisible' may have stolen the limelight, but one of the researchers behind that work believes that superlenses remain the most alluring idea yet proposed for metamaterials. "I think that the superlens will probably find more applications [than cloaking]," says Xiang Zhang, based at the University of California, Berkeley, who grabbed headlines in April for using metamaterials to create a primitive cloak[2].

The key to all this comes down to the structures of metamaterials, which directly affect their optical properties. The composition of conventional substances, such as glass, is uniform at optical wavelengths, but metamaterials feature regularly patterned structures at those scales. When light of a certain wavelength interacts within the pattern, it sets up a resonance, similar to the way a musical tone triggers vibrations in a tuning fork. The resonance can cause light beams to be deflected in the 'wrong' direction and it can also enhance certain properties.

Pendry's initial work on metamaterials was aimed at beating a problem known as the diffraction limit, which says that optical instruments such as telescopes and microscopes are subject to a fundamental fuzziness. No matter how well they are made, their ability to resolve fine detail is constrained by the wavelength of the light being used. For optical microscopes, this means that any attempt to resolve features below around 200 nanometres will fail.

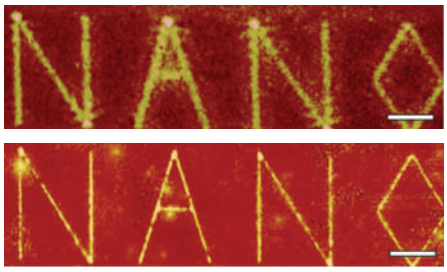
Pendry recognized a loophole: the diffraction limit does not necessarily hold for metamaterials. Although silver is not a perfect metamaterial, he showed that it would work well enough as a lens. If a silver film were placed close enough to an illuminated object, it could catch the 'evanescent waves' — short-range electromagnetic fields that carry the detailed, subdiffraction-scale information about the object, but that cannot be picked up beyond a few tens of nanometres from the object's surface. These evanescent waves would be amplified by resonances in the silver film, and their information would be carried through the film to create an exquisitely detailed image on the far side.

Window of opportunity

Richard Blaikie, an electrical engineer at the University of Canterbury in Christchurch, New Zealand, was excited by this concept — not as a way to see, but as a way to write. He thought that the superlens could help chip-makers in their struggle to create subdiffraction-scale circuit elements. But, as Blaikie soon learned, what looked elegant on paper was messy in the lab. "It's very easy to do some modelling," he says, "but the experiments are very hard."

Blaikie did manage to make a prototype superlens in 2005 that captured subdiffraction information[3]. But because the evanescent waves fade away over such a short distance, the silver

film had to be fixed to the object being imaged. Similarly, whatever recorded the image had to be clamped tightly to the other side of the film. And even then, says Blaikie, "the images that we got, although good in terms of their resolution, were poor in terms of their fidelity". Even a few nanometres of variation in the silver's surface created hotspots that made straight lines appear jagged. Blaikie says his group has struggled to follow up that 2005 demonstration. "You haven't seen anything because we haven't had any success," he says bluntly.



Superlens in action: a thin film of silver delivers a more detailed image (bottom) of nanometre-scale lines.

N. FANG *ET AL. SCIENCE* **308,** 534–537 (2005)

Hyper activity

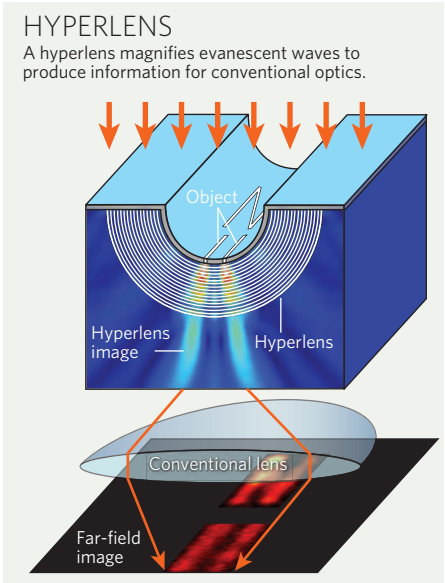
In 2006, partly in reaction to the practical limitations of superlenses, two theoretical groups independently developed the hyperlens[4,5]. Instead of a single film, the hyperlens features alternating layers of a metal and an insulator. Experimentally, the layers are arranged to form a half-cylinder (see diagram). The object is placed in the centre of the cylinder and evanescent waves from it are caught by the lens and magnified as they pass through the various layers, emerging as light that shines freely. The hyperlens is more complex than the superlens, but has a major advantage in that it should be able to feed the light emerging from its surface into conventional optics. "I'm hopeful that this hyperlens could be a revolutionary instrument," says Nader Engheta, a theorist at the University of Pennsylvania in Philadelphia and one of the originators of the idea.

> **"You haven't seen anything because we haven't had any success."**
> — Richard Blaikie

Again, experimentalists rushed into the lab to build hyperlenses, and again they met with initial success. Less than a year after the idea was proposed, two groups had made proof-of-concept hyperlenses that beat the resolution of conventional optics. But, as with superlenses, the devil is in the detail. "The principle is working; on the other hand, you are limited by engineering problems," says Igor Smolyaninov of the University of Maryland in College Park, who built one of the demonstration hyperlenses. Smolyaninov's lens, for example, works in only one dimension. A two-dimensional version would be more complex to manufacture, he says. And then there is the problem of loss: imperfections at the interfaces between each metal–insulator layer can cause light to scatter randomly.

Progress on the perfect lens has now slowed considerably — although this is partly a result of the excitement over cloaking devices. Pendry was again the source for the idea when, in 2006, he showed that metamaterials could potentially bend light around an object, effectively shielding it from view[6]. Many researchers are now working on developing the metamaterial nanostructures required for the job. As a result, some scientists believe industry needs to start investing in super- and hyperlenses to give that technology a boost. But for both imaging and circuit etching, the lenses would be entering markets in which competing subdiffraction technologies are already further along in development. Semiconductor researchers can use nonlinear optics to write circuits below the diffraction limit. And for biological imaging, other groups are developing ways to beat the diffraction limit using a combination of fluorescent proteins and clever optics.

Biologists have shown only tepid interest in superlenses, admits Nicholas Fang, a researcher at the University of Illinois at Urbana-Champaign. Still, Fang and many others remain fiercely devoted to the superlens idea. If the lenses can be made to work, they would be able to image living biological specimens with unprecedented detail. Granted, he says, "there are many open issues that we did not realize at the very beginning". But he points out that this is hardly unusual with new technologies.



Superlenses do work in theory, he contends, and many of the problems are "more of an engineering issue". For his part, he believes that the field may have to draw a little more from industrial expertise. For example, he has used industrial-grade germanium to deposit atomically flat films of silver[7] that can reduce the distortion seen by Blaikie.

Other groups are working to make hyperlenses more practical and, again, techniques from the semiconductor industry seem to be helping. In April, Stefan Mendach and his colleagues at the University of Hamburg, Germany, created a hyperlens by rolling up alternating layers of semiconductor materials[8]. The technique seems to be an easy way to make a hyperlens, although Mendach needs to get more layers in his system before it can work. And Vladimir Shalaev at Purdue University in West Lafayette, Indiana, is trying to build 'flat' hyperlenses. Although they are harder to make — the individual layers can no longer be uniform — they could prove easier to use, as the specimen would not have to be precisely positioned along the cylinder axis.

Zhang, meanwhile, is pursuing a simpler but related concept that he believes could revolutionize data storage. Rather than using a single superlens, he is using several 'plasmonic' lenses, which focus evanescent waves. When normal light strikes the lens it interacts with electrons on the lens surface. This concentrates evanescent waves into a single point, effectively creating a subdiffraction hotspot that could be used to write a piece of information on to an optical disk. Zhang has developed a prototype array of plasmonic lenses that does indeed write information far below the diffraction limit.

Back at his desk, Pendry is willing to wait for his revolution with a patience that is perhaps unique to theoretical physicists. Before he proposed superlenses, he spent years calculating the quantum mechanical forces arising between two blocks of glass flying past each other at near the speed of light, simply because the problem was theoretically interesting. Compared with that earlier work, realizing a perfect lens seems eminently practical. "It's really a very, very simple technology," he says. "It's just a question of being able to get it right." ■

X. ZHANG/UNIV. CALIFORNIA, BERKELEY

Geoff Brumfiel is a senior reporter for *Nature* in London.

1. Pendry, J. B. *Phys. Rev. Lett.* **85,** 3966–3969 (2000).
2. Valentine, J., Li, J., Zentgraf, T., Bartal, G. & Zhang, X. *Nature Mater.* doi:10.1038/nmat2461 (2009).
3. Melville, D. & Blaikie, R. *Opt. Express* **13,** 2127–2134 (2005).
4. Salandrino, A. & Engheta, N. *Phys. Rev. B* **74,** 075103 (2006).
5. Jacob, Z. Alekseyev, L. V. & Narimanov, E. *Opt. Express* **14,** 8247–8256 (2006).
6. Pendry, J. B., Schurig, D. & Smith, D. R. *Science* **312,** 1780–1782 (2006).
7. Logeeswaran, V. J. *et al. Nano Lett.* **9,** 178–182 (2009).
8. Schwaiger, S. *et al. Phys. Rev. Lett.* **102,** 163903 (2009).

CORRESPONDENCE

Cognition: evolution does help to explain how minds work

SIR — In their Essay 'Can evolution explain how minds work?' (*Nature* **458,** 832–833; 2009), Johan Bolhuis and Clive Wynne use Darwin's claim that there is "no fundamental difference between man and the higher mammals in their mental faculties" to explain how people have gone down the wrong path in studying cognition. But in homing in on examples of convergent evolution, in which humans and some distantly related species such as songbirds seem to have come up with similar solutions to the same problem, Bolhuis and Wynne neglect one key feature that distinguishes humans from all other animals.

The feature that is peculiar to humans is their understanding about the causal interactions between physical objects (see, for example, L. Wolpert *Six Impossible Things Before Breakfast*; Faber, 2006). For example, children realize from an early age that one moving object can make another move on impact. It is this primitive concept of mechanics that is a crucial feature of causal belief, and that conferred an advantage in tool-making and the use of tools — which, in turn, drove human evolution.

Animals, by contrast, have very limited causal beliefs, although they can learn to carry out complex tasks. According to Michael Tomasello (*The Cultural Origins of Human Cognition*; Harvard Univ. Press, 1999), only human primates understand the causal and intentional relations that hold among external entities. Tomasello illustrates this point for non-human primates with the claim that even though they might watch the wind shaking a branch until its fruit falls, they would never shake the branch themselves to obtain the fruit. Some primates are, nevertheless, at the edge of having causal understanding.

Once causal belief evolved in relation to tools and language, it was inevitable that people would want to understand the causes of all the events that might affect their lives — such as illness, changes in climate and death itself. Once there was a concept of cause and effect, ignorance was no longer bliss, and this could have led to the development of religious beliefs.

Lewis Wolpert Department of Cell and Developmental Biology, University College London, London WC1E 6BT, UK
e-mail: l.wolpert@ucl.ac.uk

Cognition: theories of mind in animals and humans

SIR — I believe that Johan Bolhuis and Clive Wynne, in their Essay 'Can evolution explain how minds work?' (*Nature* **458,** 832–833; 2009), profoundly misrepresent the burgeoning interdisciplinary field of comparative cognition research as being rife with anthropomorphism and ignorance of evolutionary principles.

Darwinism in animal cognition is represented, not by the naive anthropomorphism practised by Darwin and his contemporaries, but by Darwin's claim that humans differ mentally from other species "in degree but not in kind". There is evidence from behavioural studies that many of humans' mental powers are shared by other animals, including simple forms of learning, memory and categorization, and the elements of social, spatial and numerical cognition.

Only against this background does it make sense to propose, as some have, that there is a distinct small set of mental powers that is unique to humans, including theory of mind (see, for example, D. C. Penn *et al. Behav. Brain Sci.* **31,** 109–178; 2008). Contrary to what Bolhuis and Wynne are suggesting, careful analysis of the behaviours taken as evidence for theory of mind in species from chimpanzees to dogs to birds has led to a rethink of claims that were initially anthropomorphic.

The authors suggest that relatedness among species is emphasized at the expense of convergently evolved cognitive similarities. On the contrary: apparently similar performances of distantly related species — as in tool-using, social cognition and teaching-like behaviour — are now increasingly being studied precisely because convergence is a recognized test-bed for functional hypotheses. For example, the proposal that monkeys and apes have evolved exceptional social skills to navigate a particular kind of social group is tested with non-primate species that have complex social organization, such as hyenas, some birds and even fish. Comparison of the cognitive mechanisms underlying such functionally similar behaviours is an active area of research.

Bolhuis and Wynne end with the proposal that we should "study animal and human minds empirically" — as if that is not already being done (for an overview, see S. J. Shettleworth *Behav. Processes* **80,** 210–217; 2009). As they point out, evolution cannot explain how minds work (any more than it can explain precisely how nerves or genes work), and our own psychology can be an obstacle to understanding that of other species.

Nevertheless, studying minds in their phylogenetic and functional context can provide an indication of what they were selected to do. And, as in the other biological sciences, it can offer a theoretical framework for meaningful comparative research into mechanisms.

Sara J. Shettleworth Department of Psychology, University of Toronto, 100 St George Street, Toronto, Ontario M5S 3G3, Canada
e-mail: shettle@psych.utoronto.ca

John Maddox and the Medical Research Council

SIR — In his Obituary 'Maddox by his successor' (*Nature* **458,** 985–986; 2009), Philip Campbell describes an account given by the former editor of *Nature*, John Maddox, of how he walked out of a meeting of the UK Medical Research Council (MRC) when he realized he had been summoned to defend his public criticisms of that institution. In fact, Maddox was never asked to attend a meeting of the MRC. In my ten years as MRC secretary, and my time as a member of the council during Harold Himsworth's tenure as secretary, the council has never "summoned" anyone to a meeting to "defend" their criticisms, and indeed they never would.

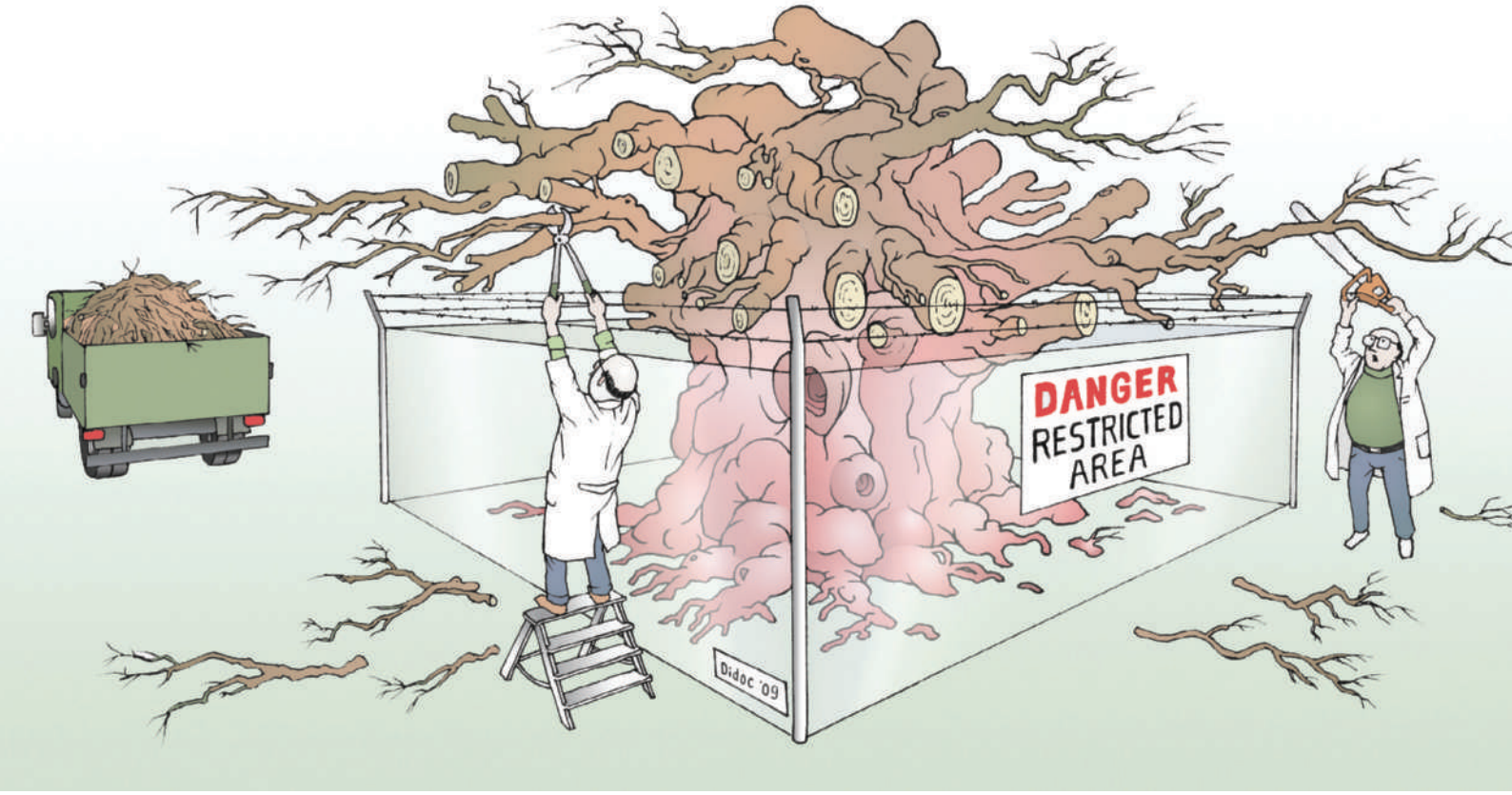
Maddox must have been referring to the occasion (about which I never informed the council, before or after the event) when I asked him if he would like to drop into the MRC office to hear why I had asked two individuals to consider withdrawing letters that they had sent to *Nature* for publication. The letters, of which the authors had sent me copies, were about a clinical trial we planned to carry out in women who had had a child with the birth defect spina bifida. I had thought, perhaps mistakenly, that the two letters might further increase the growing public and political opposition to the trial, which was threatening to prevent its launch.

When Maddox came to the MRC office, he simply handed me a copy of the next week's *Nature*, pointing out a blank column where the letters might have appeared (*Nature* **300,** 310; 1982). I handed it back and he left with hardly a word (certainly not a discourteous word) passing between us. In that sense, he did "walk out".

James Gowans Cumnor Hill, Oxford OX2 9HX, UK
e-mail: jamesgowans@btinternet.com

The Nature *John Maddox special is at www.nature.com/jm*

ESSAY



J. H. VAN DIERENDONCK

A change of strategy in the war on cancer

Patients and politicians anxiously await and increasingly demand a 'cure' for cancer. But trying to control the disease may prove a better plan than striving to cure it, says **Robert A. Gatenby**.

The German Nobel laureate Paul Ehrlich introduced the concept of 'magic bullets' more than 100 years ago: compounds that could be engineered to selectively target and kill tumour cells or disease-causing organisms without affecting the normal cells in the body. The success of antibiotics 50 years later seemed to be a strong validation of Ehrlich's idea. Indeed, so influential and enduring was medicine's triumph over bacteria that the 'war on cancer' continues to be driven by the implicit assumption that magic bullets will one day be found for the disease.

Yet lessons learned in dealing with exotic species, combined with recent mathematical models of the evolutionary dynamics of tumours, indicate that eradicating most disseminated cancers may be impossible. And, more importantly, trying to do so could make the problem worse.

"The principles for successful cancer therapy might lie in the evolutionary dynamics of applied ecology."

In 1854, the year Ehrlich was born, the diamondback moth, *Plutella xylostella*, was first observed in Illinois. Within five decades, the moth, whose larvae feed on vegetables such as cabbage and Brussels sprouts, had spread throughout North America. It now infests the Americas, Europe, Asia and Australia. Attempts to eradicate it using various chemicals suppressed populations only fleetingly and, in the late 1980s, biologists found strains resistant to all known insecticides. Over the past couple of decades, agriculturalists have abandoned efforts to eliminate the diamondback moth. Instead, most now apply insecticides only when infestation exceeds some threshold level with the goal of producing a sustainable and satisfactory crop.

Under the banner of 'integrated pest management', hundreds of invasive species are now successfully controlled with strategies that restrict population growth. By contrast, very few such species have been eradicated. An infestation of the giant African snail, *Achatina fulica*, was eliminated in Miami, Florida, in the 1960s, for instance. But the snail is easy to catch and, in this case, it had spread to only a few city blocks. Two centuries of experience have shown that the vast majority of introduced species are simply too heterogeneous, too dispersed and too adaptive to be eliminated.

Adapt and conquer

The dynamics of exotic species and invasive cancers differ in many obvious and subtle ways, yet there are important similarities. The invasion of pests involves dispersal, proliferation, migration and evolution — all of which are analogous to the processes that allow cancer cells to spread from a primary tumour into adjacent tissues or to new locations in the body via the lymphatic system or blood

vessels. Furthermore, the ability of tumour cells to adapt to a wide range of environmental conditions, including to toxic chemicals, is very similar to the evolutionary capacities shown by invasive species.

As with invasive species, for disseminated cancers, successful eradication is rare. Hodgkin's lymphoma, testicular cancer and acute myeloid leukaemia can be consistently cured using aggressive chemotherapy. But, like the African snail, these malignant cells seem to have characteristics that make them particularly responsive to 'treatment'. Some are unusually homogeneous, for example, so have limited capacity to adapt.

Eradicating the large, diverse and adaptive populations found in most cancers presents a formidable challenge. One centimetre cubed of cancer contains about 10^9 transformed cells and weighs about 1 gram, which means there are more cancer cells in 10 grams of tumour than there are people on Earth. Unequal cell division and differences in genetic lineages and microenvironmental selection pressures mean that the cells within a tumour are diverse both in genetic make-up and observable characteristics. Additionally, tumours are complex ecosystems: they include normal cells as well as regions of low blood flow and oxygen content where cancer cells are relatively protected from the effects of chemotherapy. If blood flow is poor, for example, so is the delivery of the toxic drugs.

Outwitted by evolution

The typical goal in cancer therapy, similar to that of antimicrobial treatments, is killing as many tumour cells as possible under the assumption that this will, at best, cure the disease and, at worst, keep the patient alive for as long as possible. Indeed, for more than 50 years, oncologists have tried to find ways to administer ever-larger doses of ever-more cytotoxic therapy. But, just as invasive species consistently adapt to pesticides, regardless of concentration or cleverness of design, so too do cancerous cells adapt to therapies. Indeed, the parallels between cancerous cells and invasive species suggest that the principles for successful cancer therapy might lie not in the magic bullets of microbiology but in the evolutionary dynamics of applied ecology.

Support for this idea comes from *in vivo* experiments, computer simulations and recently developed mathematical models of tumour evolutionary dynamics. These suggest that efforts to eliminate cancers may actually hasten the emergence of resistance and tumour recurrence, thus reducing a patient's chances of survival[1].

The reason for this arises from a component of tumour biology not ordinarily investigated: the cost of resistance to treatment. Cancer cells pay a price when they evolve resistance to a particular treatment. For instance, to cope with chemotherapy, a cancer cell may increase its rate of DNA repair, or actively pump the drug out across the cell membrane. In targeted therapies, in which drugs interfere with the molecular signalling needed for proliferation and survival, a cell might adapt by activating or upregulating alternative pathways. All these strategies use up energy that would otherwise be available for invasion into non-cancerous tissues or proliferation, and so reduce the fitness of the cell. And the more complex and costly the mechanisms used, the less fit the resistant population will be.

> "Efforts to eliminate cancers may actually hasten the emergence of resistance and tumour recurrence."

That cancer cells pay a cost for resistance is supported by several observations. Cells in laboratory cultures that are resistant to chemotherapies and to tyrosine kinase inhibitors, a form of targeted therapy, typically lose their resistance when the chemicals are removed[2]. In cell lines identical except for their sensitivity to tyrosine kinase inhibitors, resistant populations typically grow more slowly than sensitive ones. Lung cancer cells resistant to the chemotherapy gemcitabine are less proliferative, invasive and motile than their drug-sensitive counterparts[3,4]. Also, although resistant forms are commonly found in tumours that haven't yet been exposed to treatment, they generally occur in small numbers[5]. This suggests that the resistant cells are not so unfit as to be completely out-competed by the drug-sensitive ones, but that they struggle to proliferate when both types are present.

Our models show that in the absence of therapy, cancer cells that haven't evolved resistance will proliferate at the expense of the less-fit resistant ones. And, when a large number of the sensitive cells are killed, for instance by aggressive therapies, the resistant types are able to proliferate unconstrained. This means that high doses of chemotherapy might actually increase the likelihood of a tumour becoming unresponsive to further therapy.

So, just as the judicious use of pesticides can be used to successfully control invasive species, a therapeutic strategy explicitly designed to maintain a stable, tolerable tumour volume could increase a patient's survival by allowing sensitive cells to suppress the growth of resistant ones. To test this idea, we treated a human ovarian cancer, grown in mice, with conventional high-dose chemotherapy[1]. The cancer rapidly regressed but then recurred and killed the mice. Yet when we treated the mice with a drug dose continuously adjusted to maintain a stable tumour volume, the animals, although not cured, survived.

The body's predators

Designing therapies to sustain a stable tumour mass rather than eradicate all cancer cells will require a long-term, multilayered strategy that looks beyond the immediate cytotoxic effects of any one treatment. Researchers will need to establish the mechanisms by which cancer cells achieve resistance and what it costs them. They will also need to understand the evolutionary dynamics of resistant populations, and design strategies to suppress or exploit the adapted characteristics.

An obvious problem is the accumulation of toxicity in patients exposed to prolonged, albeit lower, levels of drugs. But in parallel with the diverse array of predators and pathogens currently used to control invasive species, the immune system offers a rich potential source of 'predators' such as T lymphocytes that could sustain a stable cancer population — both by killing tumour cells and selecting for fitness-lowering adaptations.

I am not suggesting that cancer researchers should abandon their search for ever-more-effective cancer therapies, or even for cures. However, instead of focusing exclusively on a glorious victory, they should address the possible benefits of an uneasy stalemate in appropriate situations. Even now, many oncologists agree in principle that therapeutic strategies aimed at controlling cancer could prove more effective than trying to cure it. But the idea of killing not the maximum number of tumour cells possible but the fewest necessary will be difficult for both physicians and patients to accept in practice.

Certainly in a war that is steeped in the tradition of magic bullets and all-out attacks with high-dose chemotherapy, such an approach may seem defeatist. However, in battles against cancer, magic bullets may not exist and evolution dictates the rules of engagement. ■

Robert A. Gatenby is in the departments of radiology and integrated mathematical oncology, Moffitt Cancer Center, Tampa, Florida 33612, USA.
e-mail: robert.gatenby@moffitt.org

1. Gatenby, R. A., Silva, A., Gillies, R. & Frieden, B. R. *Cancer Res.* (in the press).
2. Scappini, B. *et al. Cancer* **100,** 1459–1471 (2004).
3. Gautam, A. & Bepler, G. *Cancer Res.* **66,** 6497–6502 (2006).
4. Davidson, J. D. *et al. Cancer Res.* **64,** 3761–3766 (2004).
5. Izzo, J. G. *et al. Mol. Cancer Ther.* **5,** 2844–2849 (2006).

BOOKS & ARTS

An Italian vision of a scientific Utopia

A century ago, artists and writers from Italy imagined a world governed by science and technology. But their vision of modernity also glorified violence and misogyny, finds **Ziauddin Sardar**.

Inventing Futurism: The Art and Politics of Artificial Optimism
by Christine Poggi
Princeton University Press: 2009. 392 pp. £32.50, $45

Futurist Umberto Boccioni conveys the dynamic forces behind city construction in *The City Rises* (1910).

MRS SIMON GUGGENHEIM FUND © 2009 DIGITAL IMAGE MOMA, NEW YORK/PHOTO SCALA, FLORENCE

A hundred years ago, a group of Italian visionary artists declared war on civilization. Rejecting the artwork, poetry, music and architecture of the period, these 'Futurists' wanted to create the world anew. Science and technology formed the building blocks of their brave new world, which they expressed not just in art but in violence and naked nationalism. In *Inventing Futurism*, art historian Christine Poggi describes how the Futurist movement's raw passion for technology was moulded by the atmosphere of political foreboding of the times.

Like futurists working now in the fields of futures studies and foresight, these artists wanted to shape the future. But the goals of the two groups could not be more different. Futurists today forecast how science and technology will change our lives, and predict alternative paths. By applying the best lessons of history to build on what exists now, they aim to find policies to ensure futures that are more equitable and just. By contrast, the Italian Futurists rejected everything that was old. They were determined to destroy the existing order and desired a future in which speed and technology represented the absolute triumph of man over nature. They glorified electricity, the car, the aeroplane and the industrial city. They despised women, the human body and the idea of a peaceful coexistence.

The godfather of Futurism was writer and poet Filippo Tommaso Marinetti (1876–1944), who published his 'The Founding and Manifesto of Futurism' on the front page of the French newspaper *Le Figaro* on 20 February 1909. "We want to free this land," Marinetti wrote, "from its smelly gangrene of professors, archaeologists, tourist guides and antiquarians … the numberless museums that cover her like so many graveyards." He urged his readers to set fire to library shelves and to flood museums. "Take up your pickaxes … and wreck, wreck the venerable cities, pitilessly! Art, in fact, can be nothing but violence, cruelty, and injustice." Marinetti saw science as a modern, virile enterprise to be pursued at all costs, and technology as the instrument that would usher the world into sunlight with velocity and violence. A mythical struggle had to be waged between the masculine forces of science and technology, represented by the sea, and the seductive feminine power of the stars that prevented civilization from advancing forwards.

Within a year of Marinetti's sermon, the Italian Futurist movement was born. Artists such as Umberto Boccioni, Antonio Sant'Elia and Luigi Fillia emphasized speed, energy, flight, industrial landscapes and destructive violence in their paintings. The original manifesto was followed by a host of others on almost everything from clothing, food and smells to wars and lust. "We hurl our defiance to the stars," they declared; "We will glorify war."

The Futurists imagined a world governed by electricity. Their electrical fantasies, writes Poggi, take a Utopian turn in their vision and evolve into an orgy of violence. They saw Italy as being "fertilized" by electricity, banishing hunger, poverty, disease and work. Air temperature and ventilation would be controlled automatically, telephones would be wireless, and crops and forests would spring up at speed. But in this world of ease and plenty, fierce competition would arise over superabundant industrial production. War would break out, fought by "small mechanics" whose flesh resembled steel. Deploying "steel elephants" and battery-powered trains from afar, they would wage a thrilling interplanetary war.

The union of man and machine is central to Futurist thought, and is best defined in Marinetti's 1909 novel *Mafarka the Futurist*. Mafarka is an Arabian king with imperialist ambitions who creates a mechanical son, Gazurmah, to be his immortal substitute. Carved out of oak and modelled on an aeroplane, Mafarka finds his coarse skin, squared jaw, ribs of iron and formidable metallic member alluring, and breathes life into his son with a lingering homoerotic kiss. But his creation devours him — a fate Mafarka has foreseen and desired so that he might be reborn in the immortal son. Gazurmah proceeds to rape and obliterate Earth.

Gazurmah is not far removed from the Terminator cyborg in the 1984 film of the same name. But whereas *The Terminator* is a dystopia, Marinetti's vision, with its aspirations of autogenesis, immortality and demonization of women's bodies, is presented as a distinctive Utopia.

Shades of the liquid-metal terminator from *Terminator 2: Judgment Day* (1991) are evident

in Fillia's 1929 painting *Spirituality of Aviator*. Fillia, who published a manifesto on 'The mechanical idol', presents his aviator as a fluid biomorphic shape embedded in a semi-transparent, tilted plane. Man and machine become a single permeable body with fluid boundaries. The aviator's mystical body seems to give birth to an industrial city, indicated by smoke gushing through circular openings, carrying within their stream three small buildings. Poggi explains how Fillia painted a number of other notable pictures in which landscapes and bodies merged with technology to depict a "religion of velocity".

To exemplify their message, the Futurists equated science with modernity and technology — speeding trains, giant electric power plants, bridges and abstractions from photographs. Boccioni's paintings, for example, compare and contrast the industrial radiance of science with the obscurity of the past and the oppression of natural forces, often depicted as trees, the Moon and stars. In his *Study for Homage to the Mother* (1907–08), he depicts two children. One is working and questioning science, with a window on to modern life. The other works by lamplight, the window showing a cloudy evening sky and a glimmering Moon. A central panel shows a tired mother with a figure on either side to symbolize the feelings of the children: one sweet and feminine, the other angry and defiant.

The final destination of this quest is illustrated in Sant'Elia's 1914 work *The New City*. Poggi suggests that Sant'Elia created this series of lyrical drawings of visionary hydroelectric plants to exemplify the human annexation of the energy inherent in matter. These are bold, spectacular images of geometric masses, symbolic of the dawning of a new age. The perspective in *The New City* is always from below, dwarfing and making the imaginary viewer insignificant.

The visions and concerns of the Futurists, Poggi tells us in this difficult, sometimes frightening but always illuminating study, emerged out of the uncertainty and confusion produced by modernity. Their artificial optimism sought to produce a philosophy for a new life, not just new art or architecture. It is not surprising that the Futurists saw an echo of excessive nationalism in their notion of modernist violence and war. Misplaced faith in science, as rational dogma, as the enemy of pessimism, as a theory of salvation, often serves as the glue that binds modernity and fascism together. ■

Ziauddin Sardar is the editor of *Futures*, the journal of policy, planning and futures studies. e-mail: futures@ziasardar.com

***Futurism*, an exhibition at the Tate Modern in London, opens on 12 June. For details see http://tinyurl.com/tatefuturism.**

A limited view of the future

What's Next? Dispatches On the Future of Science
Edited by Max Brockman
Vintage Press: 2009. 256 pp. $15

Now is the time for a broad and critical look at the future of science. Just as US President Barack Obama has vowed to "restore science to its rightful place", Max Brockman has published this collection of future-oriented essays.

Brockman, a literary agent whose firm represents popular-science authors such as Jared Diamond, Richard Dawkins and Steven Pinker, has access to some of the most established thinkers. But in compiling *What's Next?,* Brockman chose instead to work with "the coming generation of scientists". Readers might wonder whether these young people have enough experience to provide the long view, but their biographies and publication records are impressive. For instance, climatologist Laurence C. Smith, 15 years out of a PhD, has more than 50 papers to his credit and briefs Congress on global climate change.

What's Next? is slim and readable, but lacks detailed references. Technical language is kept to a minimum, so it is accessible to a general audience. But it did not strike me as an arresting view of the future — its scope is unimaginative, covering more about what we already know than what we don't.

A pervasive theme in the book, which is heavily slanted toward psychology, is the scientific basis for ethical behaviour. Neuroscientist Christian Keysers explains that mirror neurons are activated when we perform certain activities, and when we watch others do those same activities: "The emotions of others are contagious because our brain activates our own emotions at the sight of them." This facilitates both learning and empathy. "Our brain," he concludes, "is ethical by design."

Psychologist Joshua D. Greene sees morality as a tension between the opposing forces of intuitive emotion and cognition. "If I'm right," he says, "this tension between competing neural systems underlies not only centuries-old disagreements … but also contemporary tussles over … stem-cell research and the torturing of suspected terrorists." When issues are personal, the parts of the brain associated with emotional response tend to dominate, whereas in less personal situations, those regions associated with cognition are stimulated.

Philosopher Nick Bostrom tackles human enhancement. He is concerned because humans "are a marvel of evolved complexity", something with which we tamper at our own risk. So he proposes a "rule of thumb, for identifying promising human enhancements". Bostrom sees some of our limitations as resulting from selection pressures that no longer exist for most humans. Today, for instance, we can feed the higher metabolic demands of a larger brain, whereas in our recent past, we could not. We might also overcome evolutionary restrictions. Bostrom suggests that genetic 'medication' could be administered to confer an advantage, such as the protection a mutant haemoglobin gene offers against malaria in people with the sickle-cell trait. Alternatively, embryo screening could promote favourable genetic profiles. Thus, Bostrom sees the morality of human enhancement as an issue of what is achievable rather than what is acceptable. His heuristic is useful from the scientific point of view, offering us a test for whether we should even consider a particular kind of enhancement, but it probably won't be accepted by the ethics community.

C. PENN/PANOS PICTURES

People who inherit the sickle-cell trait are protected from malaria — could other traits hold promise as genetic treatments?

In a global society, we must all live alongside people who hold different beliefs to our own. Psychologist Matthew D. Lieberman believes that some ideas are more 'sticky' than others, and that the ideas that persist differ from one cultural group to another. He argues that "Big Ideas sometimes match the structure and function of the human brain such that the brain causes us

to see the world in ways that make it virtually impossible not to believe them." Lieberman thinks that East Asian cultures stress interconnectedness among individuals, whereas Western Europeans tend to be more independent. He suggests that this tendency might be genetically influenced by a serotonin transporter gene, found twice in its 'short' variant in two-thirds of East Asians, but in only one-fifth of Western Europeans. "These cultural Big Ideas appear to have migrated until they found the populations with the right neurochemistry to make them sticky," Lieberman says.

Psychologist Lera Boroditsky asks how language influences the way we think: "People who speak different languages do indeed think differently and … even flukes of grammar (such as the particular gender assigned to a noun in different languages) can profoundly affect how we see the world."

Modern science is highly specialized, but many scientists agree that interdisciplinary research is crucial to future work. Climatologist Gavin Schmidt asks why specialization has not led to the complete Balkanization of science, remarking that "It has been suggested that the physicist, physician, and Egyptologist Thomas Young (1773–1829) was the last person to 'know everything'." Schmidt seeks to distinguish between those fields that are separate out of necessity, and those that are mere artefacts of the way we organize our institutions. Fields become established because "the older, pioneering generation wants to employ and promote successors in its own mold; the younger generation wants to emulate its mentors", he says. But then "jargon excludes most of those outside the subfield". Schmidt warns us of the barriers that will have to be broken down, but he does not provide a good solution for doing so.

What Next? covers other ground, including global climate change, virus metagenomics, dark matter, entropy, the perception of time and the question of what makes us human. But there is surprisingly little coverage of computing or computational biology, engineering, space travel and colonization, personalized medicine, biotechnology, stem-cell biology, human cloning or artificial intelligence. The authorship is biased toward US male psychologists, and other voices might have broadened the perspective. Ideally, the book should have been twice the size, laden with references and with less repetition, and really pushing the edge. Scientists must not leave the future entirely in the hands of science-fiction writers. ■

Michael A. Goldman is professor and chair of the Department of Biology at San Francisco State University, California 94132-1722, USA.
e-mail: goldman@sfsu.edu

Galileo and the Pope

Two Men of Florence
by Richard N. Goodwin
Huntington Theatre, Boston, Massachusetts
6 March to 5 April 2009

Richard Goodwin's play explores the personal interactions between Galileo and Pope Urban VIII.

T. C. ERICKSON

Richard Goodwin's engrossing, witty and moving — if somewhat ahistorical — play about Galileo Galilei and Pope Urban VIII recently completed a run in Boston, Massachusetts, with a fine production mounted by the Huntington Theatre Company. Significantly reworked since its original production in 2003, *Two Men of Florence* deserves to become a classic.

Goodwin retains a reference to the title of his earlier version, *The Hinge of the World*, which plays on the alternative meaning of 'hinge' as 'cardinal'. Having completed the masterpiece that will destroy his standing with the Catholic Church, Galileo says to his daughter, "You and I, Maria, the saint and the philosopher, together we lean on the hinge of the world." Goodwin's depiction of their relationship, a role enhanced since the earlier production, deviates from the historical one that Dava Sobel described so compellingly in her 1999 book *Galileo's Daughter*.

Far from the play's portrayal of Maria as an essential collaborator in her father's work, the real Maria Celeste was placed in a convent at the age of 13, where she served as an apothecary. Although she read her father's books and letters, her assistance was limited to supplying him with medications and moral support. Galileo drew comfort from her letters as he endured his confrontation with the Inquisition and the Pope in 1633, following the publication of his controversial work '*Dialogo*'.

Goodwin's new title for the play focuses the audience's expectations more properly on the complicated relationship between Galileo the scientist and Urban the cleric. Both native Tuscans, they discover that they are intellectual compatriots. Galileo recognizes Urban as "a philosopher", an identification that Urban accepts with a slight modification: "Not a philosopher, Galileo, but at least a man of Florence, one who understands that the light of philosophy can enhance the radiance of faith." The Boston production was well served by the two actors in the starring roles: Edward Herrmann as Cardinal Maffeo Barberini, later Pope Urban VIII, and Jay O. Sanders as Galileo.

In Goodwin's rendering, the Pope encourages Galileo to write a "discussion between the followers of Copernicus and the supporters of Aristotle, a dialogue which contains arguments of all schools, in language which every educated man will comprehend". When Galileo tells Urban that his working title is *On the Flux and Reflux of the Tides,* the Pope advises him otherwise: "Since it is a dialogue, why not call it such — *A Dialogue on the Two Great World Systems*." Although renaming the book was a requirement of the Inquisition, it was dramatic licence to make it the suggestion of the Pope himself.

Both are men of towering ambition, and each ultimately feels betrayed by the other. After reading Galileo's pro-Copernican magnum opus, the Pope complains. He had, he thought, received a promise that the book would give equal treatment to each side, Copernican and Ptolemaic: "He deceived me. He betrayed his word. He makes a joke of Christianity." When Galileo learns that as a result of what he has written he will be under house arrest and his works banned, the scientist says, "This was never discussed. Never agreed."

In 1992, Pope John Paul II said, although falling short of an apology, "A tragic mutual

incomprehension has been interpreted as the reflection of a fundamental opposition between science and faith." Goodwin's Galileo refers to this mutual incomprehension when he tells Urban, "I thought you understood, but you do not understand. My philosophy rests on faith. I have been gifted by God to understand some of the language of His creation."

Goodwin beautifully dramatizes the two men's inability to reconcile by using on stage a piece of scientific equipment. Galileo demonstrates his *occhiale* — a compound microscope, a modification of his telescope — to the delighted Pope. Several scenes and many years later, the Pope picks up the same *occhiale* to let the scientist examine a communion wafer under its lens. Although Galileo admits he sees only an uneven surface, resembling that of the Moon, he accepts on faith the wafer as the body of Christ. The Pope accuses Galileo of being a heretic, "one who misreads the faith", and "the creator of a whole new faith whose trinity is the eyes … the mouth … and the brain".

Goodwin's earlier career as a speech writer for US presidents John F. Kennedy and Lyndon Johnson has clearly familiarized him with the loneliness of political leaders. At the end of the play, it is the victorious Pope who is more bereft than Galileo. Compelled by the responsibility of his office, Urban has jettisoned not only Galileo but also his own oldest friend, Monsignor Giovanni Ciampoli, who, as secretary of the briefs, gave Galileo "the licence to unloose his pestilential work on enfeebled Europe". Urban banishes Ciampoli to a remote village.

Galileo, deprived of the companionship of his daughter who has predeceased him, is joined in the final scene by Father Benedetto Castelli, a former student of Galileo and now a professor. Castelli's return restores hope to his former teacher, whose forced recantation has left him wondering: "How shall I continue — my proclamations of new wonders returned only by mocking echoes from the indifferent hills of Tuscany." As the two men set off for Arcetri, near Florence, where Galileo will be confined for the rest of his life, the older man tells his disciple, "Let us go home. There is much for us to do."

Throughout the play, the language and stage business were enhanced through the motifs of burning, revolution and rotation. The powerful opening scene depicts philosopher Giordano Bruno being burned at the stake in Rome's Campo dei Fiori in 1600 for heresy. Later, Maria burns Galileo's papers, as per her father's request. Goodwin used large, horizontally rotating devices on stage to demonstrate some of Galileo's scientific insights to great effect.

Following its brilliant run in Boston, *Two Men of Florence* deserves to find its place as a standard in the growing repertoire of plays dealing with science and scientists, including Bertolt Brecht's *Life of Galileo* and Michael Frayn's *Copenhagen*. ■

Jay M. Pasachoff is director of the Hopkins Observatory, Williams College; **Naomi Pasachoff** is a research associate at Williams College, Williamstown, Massachusetts 01267, USA.
e-mail: jay.m.pasachoff@williams.edu; naomi.pasachoff@williams.edu

Smells like green spirit

Green Aria: A ScentOpera
The Peter B. Lewis Theater, Guggenheim Museum, New York City
31 May and 1 June 2009

Audience members attending *Green Aria*, which has its world premiere at the Guggenheim Museum in New York this week, can leave their opera glasses at home. The brainchild of director and librettist Stewart Matthew, this aria will be performed in the dark. But while the eyes get a rest, the nose will be hard at work. *Green Aria* is a scent opera, a performance involving synchronized sounds and smells.

Each of the roughly two dozen characters in this musical tale about humanity's attempt to tame nature will have a distinct odour, transmitted for a few seconds at a time through a 'scent speaker' attached to each auditorium chair. The opera will tap into two perceptual pathways in the brain simultaneously.

"There have been other attempts to introduce smell technology in conjunction with various forms of art and entertainment," explains Matthew. But it is little used today. In 1960, a system called Smell-O-Vision released timed scents in a cinema to match the scenes unfolding in a film. And a 1981 movie by John Waters, *Polyester*, prompted audiences to use scratch-and-sniff Odorama cards at particular points in the plot.

Perfume perfectionist Christophe Laudamiel.

In the Guggenheim Museum's theatre, special tubing will connect the scent speakers on the chairs to a customized 'odour organ' jointly developed by Fläkt Woods, a company specializing in air-circulation systems, and Aeosphere, a fragrance media company. Aeosphere is headed by Matthew and fragrance designer Christophe Laudamiel, who created the opera's scents.

Laudamiel holds a master's degree in chemistry and has taught at both Harvard University and the Massachusetts Institute of Technology in Cambridge. He took pains to ensure that the voices in the opera have appropriately matched scents. For example, the fragrance of one character, Crunchy Green, is made up of ingredients that include a leafy-smelling chemical and a rooty scent known as gentian absolute, derived from gentian plants that grow on dormant volcanoes in France.

The unconventional music of the opera — written by composers Nico Muhly and Valgeir Sigurðsson — does not contain words but consists instead of sung sounds, notes played by orchestra instruments and electronic elements. When the music calls for a chord of voices, the scent speakers, too, release a 'chord' of corresponding odours.

Laudamiel notes that the team had to consider human biology when constructing the piece. Specifically, the succession of sounds and scents needed careful arrangement — a person can quickly become accustomed to one scent, such that it loses its impact if there is no variation. Moreover, "the perception of scent number two is going to depend on scent number one", he adds.

Stewart doesn't bar the possibility of someday wearing one of the scents developed for Green Aria in public, but explains that they would need fine-tuning. "The intent is to compose the scents for the opera. If there is a potential to apply the scents to the skin, that's a secondary consideration." ■

Roxanne Khamsi is news editor for *Nature Medicine* in New York.

NEWS & VIEWS

Marmosets — model prospects.

L. SEREBRENNIKOV/ALAMY

DEVELOPMENTAL BIOLOGY

Transgenic primate offspring

Gerald Schatten and Shoukhrat Mitalipov

Genetically engineered monkeys carrying a foreign gene that is passed on to their offspring provide a potentially valuable bridge between mouse models of disease and treatment for human disorders.

The development of transgenic mice — in which foreign DNA is inserted into the mouse genome — meant that the functions of human genes could be studied rigorously in living animals rather than in cell culture. Developments in transgene technology, complemented by advances in reproductive cloning, have followed in other mammals, including rats, rabbits, pigs, cows, and even cats and dogs. Sasaki *et al.*[1] (page 523 of this issue) now report a breakthrough in primate research — the generation of transgenic monkeys that pass the foreign gene on to their offspring and so could be used to establish specialized primate colonies for the study of human disease.

Mouse models of disease have been used in research into disorders ranging from anaemia and asthma to autism and schizophrenia. But not every human disease can be modelled faithfully in rodents. Mice engineered to express the cystic fibrosis gene, for example, do not develop the lung problems that typify this disorder (a pig model of cystic fibrosis[2] proved more useful). Disorders of higher brain function, such as Alzheimer's disease, are especially challenging to reproduce in rodents, and here, as with many other diseases, it is our closest animal relatives — the non-human primates — that offer potentially invaluable biological models.

In the United States, research with non-human primates can be conducted with stringent local and federal oversight. Consequently, this decade has seen the generation of a transgenic rhesus monkey[3] and the first primate model of a human disorder, Huntington's disease, also in rhesus monkeys[4]. In another study, the transfer of a transgenic rhesus embryo generated a transgene-expressing placenta[5]. But no study has shown transmission of foreign DNA to gametes — the sperm and egg — which is essential for the generation of transgenic offspring. These offspring could then be bred to create transgenic-primate strains.

Sasaki *et al.*[1] build on this work, and introduce several innovations. Instead of rhesus monkeys, the authors studied the common marmoset (*Callithrix jacchus*) — small creatures that reach sexual maturity in just over a year, and that often bear twins after a relatively short gestation period. They found that naturally produced embryos, flushed from the reproductive tracts of mated females, were better transgene carriers than embryos generated by *in vitro* fertilization (IVF). After injection of the transgene, which encoded a green fluorescent protein (GFP) as a reporter of gene expression, nearly 100% of the natural embryos expressed the gene compared with about 70% of the IVF embryos — four of the five transgenic marmosets developed from these natural embryos.

To improve the efficiency of transgene delivery, the authors shrank the egg within its outer coating by placing it in a sugar solution, freeing up space for the injection of more transgene-containing particles. After transferring 80 embryos to 50 surrogate females, seven pregnancies were established, resulting in five offspring. The GFP transgene was incorporated into several sites in the offspring's genome and was expressed in various tissues, as confirmed by the green glow of the fluorescent protein. Furthermore, Sasaki *et al.* followed these animals until sexual maturity and found that the transgene was in their gametes, affording the tantalizing prospect of obtaining transgenic offspring through germline transmission. This hope came to fruition when the first infant conceived by the GFP-transgenic founder member also expressed GFP in its skin.

The birth of this transgenic marmoset baby is undoubtedly a milestone. The cumbersome and often frustrating process of making a transgenic animal from scratch need now only occur with founder animals. Subsequent generations can be produced by natural propagation, with the eventual establishment of transgene-specific monkey colonies — a potentially invaluable resource for studying incurable human disorders, and one that may also contribute to preserving endangered primate species. The study of transgenic primates may also help to answer fundamental questions about stem-cell biology. Primate stem cells

Box 1 | Considerations before establishing colonies of primate disease models

- Optimize the initial protocol for disease modelling.
- Direct research primarily at incurable diseases for which there are potential treatments in the pipeline for preclinical testing.
- Ensure the disease under study cannot be modelled in transgenic mice or other non-primates.
- Attempt to develop transgenic animals with features that allow rapid and informative research[13,14], for example the use of:
  - Transgenes with inducible promoters, meaning that genes can be switched on or off.
  - Reporter transgenes that are sensitive to particular metabolic states.
  - Gene-trap sites in the target genome, similar to the *Rosa26* locus in mice, that allow efficient integration and strong expression of inserted sequences.
  - Cre–lox technology, which can be used to excise the transgene from the target genome.
  - Gene targeting by homologous recombination for the creation of animals in which specific genes are knocked out.
  - Transgene reporters suitable for non-invasive imaging by magnetic resonance imaging, positron emission tomography, luminescence and other whole-body approaches.
- Isolate primate colonies to prevent contamination with other research colonies.
- Clarify CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) and other regulatory practices to enable sharing of molecular and cellular research resources while still protecting endangered species.
- Foster public debate about the strengths and limits of these technologies[11,12].

have recently been generated from adult cells by nuclear cloning[6], and a comparison of these cells with patient-specific induced pluripotent stem cells — also derived from adult cells — will be enlightening.

Transgenic marmosets are potentially useful models for research into infectious diseases, immunology and neurological disorders, for example. Marmosets engineered to express single-gene defects, such as the mutated gene that causes muscular dystrophy, might accelerate the translation of discoveries from mouse research[7] to patients who have few treatment options. However, marmosets do have limitations as research models. They are New World primates, and are less closely related to humans than are Old World primates such as rhesus macaques and baboons. Because of biological differences, diseases such as HIV/AIDS, macular degeneration and tuberculosis can be studied only in Old World primates.

Also, despite the commendable success rates achieved by Sasaki *et al.*[1], their results pale in comparison with those achieved with mice. As in other primate studies, the authors use a virus vector to carry the transgene into the genome of the embryo after injection. Consequently, the transgene inserts into random sites in the target DNA. This is much less satisfactory than in transgenic mice, which are now routinely generated using embryonic stem cells. Here, the transgene is directly targeted to integrate into, or mutate, a specific site in the embryonic stem-cell genome by exploiting a natural genome repair process known as homologous recombination[8]. Random transgene integration probably resulted in some of the marmoset miscarriages; and as it could theoretically activate silent cancer-causing genes, or endogenous viral sequences that are part of the host genome, monitoring transgene inheritance in subsequent generations will be necessary.

As with all animal experimentation, genetic modification in primates raises concerns about animal welfare. We suggest that various considerations should be taken into account before colonies of primate disease models are established (Box 1). There are also bioethical concerns, which raise themselves anew with this work. Foremost among them is the prospect of unwarranted and unwise application of transgenic technologies to human gametes and embryos for reproductive purposes. Transgenic technologies are still primitive and inefficient, with unknown risks for animals, let alone people. Hence the very real need for the existing guidelines framed by professional societies and regulatory authorities (for example, those issued by the UK Human Fertilisation and Embryology Authority[9]), which prevent germline genetic modifications in humans. Perhaps even the use of techniques for generating embryonic stem cells from human embryos[10] that have been genetically altered to prevent implantation, and that are therefore devoid of reproductive potential, needs to be weighed carefully so as to avoid going down any kind of slippery slope involving human transgenesis.

With recent breakthroughs in stem-cell research and these latest advances in primate developmental biology, increased attention will naturally be focused on the practices of human assisted reproductive technologies — hence the need to consider calls[11,12] to establish realistic policies for governing work with human embryos. Although the future for using transgenic primates for medical and translational research looks bright, scientists need to engage with the public in informed bioethical debate about genetic modification and innovation in reproductive biology. ■

Gerald Schatten is in the Departments of Obstetrics, Gynecology and Reproductive Sciences, and of Cell Biology, Physiology and Bioengineering, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. Shoukhrat Mitalipov is in the Division of Reproductive Sciences, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, Oregon 97006, USA.
e-mails: gschatten@pdc.magee.edu; mitalipo@ohsu.edu

1. Sasaki, E. *et al. Nature* **459,** 523–527 (2009).
2. Rogers, C. S. *et al. Science* **321,** 1837–1844 (2008).
3. Chan, A. W. S., Chong, K. Y., Martinovich, C., Simerly, C. & Schatten, G. *Science* **291,** 309–312 (2001).
4. Yang, S.-H. *et al. Nature* **453,** 921–924 (2008).
5. Wolfgang, M. J. *et al. Proc. Natl Acad. Sci. USA* **98,** 10728–10732 (2001).
6. Byrne, J. A. *et al. Nature* **450,** 497–502 (2007).
7. Lim, L. E. & Rando, T. A. *Nature Clin. Pract. Neurol.* **3,** 149–158 (2008).
8. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo: A Laboratory Manual* 3rd edn (Cold Spring Harbor Lab. Press, 2003).
9. www.hfea.gov.uk/docs/SCAAC_Genetic_ModificationJan09.pdf.pdf
10. Hurlbut, W. B. *Stem Cell Rev.* **1,** 293–300 (2005).
11. Schatten, G. *Nature Cell Biol.* **4,** s19–s22 (2002).
12. Berg, P. *Nature* **455,** 290–291 (2008).
13. Raymond, C. S. & Soriano, P. *Dev. Dyn.* **235,** 2424–2436 (2006).
14. Rochefort, N. L., Jia, H. & Konnerth, A. *Trends Mol. Med.* **14,** 389–399 (2008).

See Editorial, page 483.

QUANTUM PHYSICS

Tailor-made quantum states

Yasunobu Nakamura

The ability to produce arbitrarily superposed quantum states is a prerequisite for creating a workable quantum computer. Such highly complex states can now be generated on demand in superconducting electronic circuitry.

As an innocent reductionist in elementary school, I dreamed of creating everything in the world by assembling atoms one by one, just like building with Lego blocks. Decades later, the dream has, to some extent, come true with the 'bottom-up' approach of nanotechnology in which, for example, single atoms can be manipulated and assembled using the tip of a scanning probe microscope. But physicists are now playing with even fancier — and often more fragile — 'quantum Lego blocks'. Using a bottom-up approach, Hofheinz *et al.*[1] (page 546 of this issue) report on-demand synthesis of arbitrary quantum states in a superconducting resonator circuit. Starting from a vacuum (zero-photon) state, the authors pile up

photons one by one in the resonator and create complex quantum states in an entirely deterministic way.

Quantum mechanics was founded and, to a great extent, developed during the last century. Despite its weird and counterintuitive predictions, such as the uncertainty principle and the superposition and entanglement of states, it has stood up to a number of tests, and has proved itself to be a rigorous foundation across a broad spectrum of physics fields, from particle physics to solid-state physics. But only relatively recently have people recognized that the paradoxical nature of quantum mechanics is in itself useful in many applications, such as quantum cryptography and quantum computation. This recognition has boosted research on technologies of quantum-state engineering in various types of physical setting, and the twenty-first century will hopefully be memorable for the implementation of such technologies.

Among physical systems currently being investigated, superconducting (zero-resistance) macroscopic circuits stand in a unique position. Although the naive expectation is that quantum mechanics is normally associated with single microscopic systems such as atoms, nuclei and electrons, it has been shown that quantum-mechanical behaviour can be observed and controlled in human-designed, superconducting circuits that are micrometres or even millimetres in size[2].

The simplest example of a superconducting quantum circuit is a linear resonator consisting of an inductor and a capacitor. If proper parameters are chosen, such a circuit can store a number of energy quanta (photons) at a microwave frequency. Another example is a quantum bit (or qubit), which is an effective two-state system. It can be implemented using a Josephson junction — a tunnel junction between two superconductors — as a nonlinear inductor; the two states are the ground and the first excited state of the nonlinear circuit. Coherent control of quantum states in such circuits and their combinations[3] is the basis of superconducting quantum-state engineering.

To synthesize quantum states in a resonator, Hofheinz *et al.*[1] use a circuit (see Fig. 1a on page 546) in which a resonator is coupled to a qubit. Because it is not possible to create arbitrary quantum states in resonator circuits using classical control signals alone, the qubit is used as a 'forklift' to load photons one by one into the resonator. Each cycle consists of two sequential steps. First, the qubit, initially detuned off-resonant with the resonator, is excited by a microwave pulse. Then, the qubit energy level is tuned into resonance with the resonator, enabling coherent transfer of energy quanta. A similar technique was proposed for an optical cavity with an atom inside[4], and was demonstrated for the motional states of an ion in a trap[5]. Hofheinz and colleagues have also previously reported[6] generation of states with a certain number of photons (*N*-photon or Fock states) — with up to 15 photons[7] — based on the same scheme. In their new study[1], they perfect the scheme to precisely control not only the amplitude but also the phase of each quantum loading. This allows them to synthesize, in a completely deterministic manner, quantum states that are the largest-ever arbitrary superpositions of multiple *N*-photon states.

Another distinctive aspect of Hofheinz and colleagues' experiment is the quantitative characterization and visualization (Fig. 1) of the generated quantum states, which they attained using Wigner tomography. This method fully characterizes, by means of the Wigner function, the resonator's quantum state: just as tomography is used in medical diagnoses such as magnetic-resonance and X-ray imaging, Wigner tomography allows the quantum state to be completely reconstructed from a large number of measurements. In this case, such measurements were taken by using the same qubit, now as a diagnostic probe, to unload energy quanta from the resonator. In the past year, an analogous technique was used to characterize the quantum state of a microwave field in a three-dimensional cavity, using atoms passing through the cavity as probes of the radiation field[8].

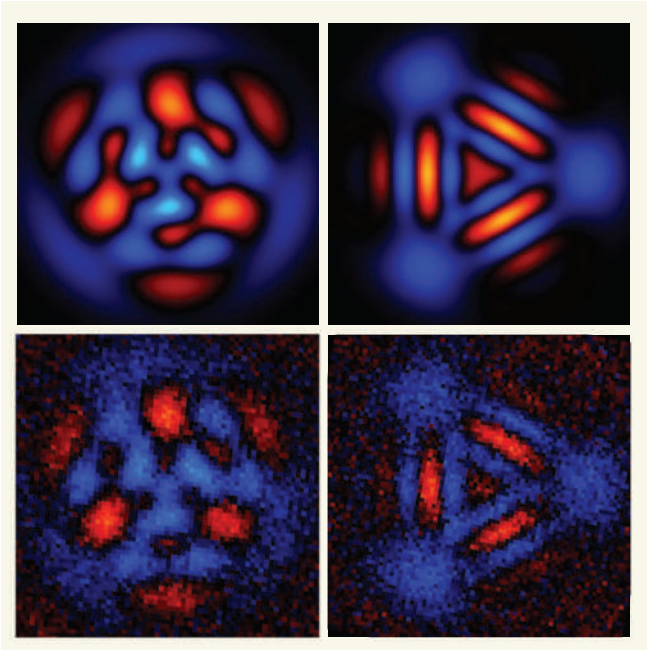


Figure 1 | Synthesizing quantum states. Hofheinz and colleagues[1] demonstrate on-demand generation of arbitrary quantum states in a superconducting resonator. In a bottom-up approach, which starts with a vacuum state, the target quantum state is constructed in a completely controlled manner by pumping photons one by one into the resonator using a superconducting quantum bit. The generated, superposed quantum states are fully characterized by the Wigner function, a representation of the resonator state, and are visualized by Wigner tomograms: top, simulated tomograms of two example states; bottom, experimentally measured tomograms. The right panel displays the 'voodoo cat' state (see Supplementary Information in ref. 1), which can be described by the superposition of three states: 'alive', 'dead' and 'zombie'. Positive values of the Wigner function are shown in blue; negative values, which are a distinctive signature of the quantum nature of the state, are shown in red.

Comparison[1] of the observed and simulated Wigner functions (Fig. 1) clearly indicates that the target quantum states were synthesized with high fidelity. The ability to accurately create and control superposed quantum states is the first requisite for quantum computing. Moreover, coupling between qubits and resonators, such as that achieved in this study, has already shown its value in the implementation of quantum gates — the analogues of logic gates in conventional computers — between remote qubits[9,10].

That said, the complexity and accuracy of the quantum states achieved by the authors is limited by decoherence — that is, the vulnerability of the quantum superposition. In superconducting circuits, quantum coherence tends to be lost more quickly than in atoms. This is not surprising if one considers the macroscopic nature of the circuits, which makes them interact more strongly with their surroundings.

Efforts to achieve longer coherence times are ongoing, and include improving circuit design and reducing the number of defects in the materials from which circuit components are made. Studying the decay of coherence in a variety of quantum states will be a valuable approach to understanding what mechanism triggers decoherence itself and the crossover from quantum to classical behaviour[7,8]. For now, Hofheinz and colleagues' experiment has set the stage for further developments in quantum-state engineering in superconducting electronic circuitry, and has brought physicists a step closer to realizing a workable quantum computer. ■

Yasunobu Nakamura is at the NEC Nano Electronics Research Laboratories, Tsukuba, Ibaraki 305-8501, and the Institute of Physical and Chemical Research (RIKEN), Wako, Saitama 351-0198, Japan.
e-mail: yasunobu@ce.jp.nec.com

1. Hofheinz, M. *et al. Nature* **459,** 546–549 (2009).
2. Clarke, J. & Wilhelm, F. K. *Nature* **453,** 1031–1042 (2008).
3. Wallraff, A. *et al. Nature* **431,** 162–167 (2004).
4. Law, C. K. & Eberly, J. H. *Phys. Rev. Lett.* **76,** 1055–1058 (1996).
5. Ben-Kish, A. *et al. Phys. Rev. Lett.* **90,** 037902 (2003).
6. Hofheinz, M. *et al. Nature* **454,** 310–314 (2008).
7. Wang, H. *et al. Phys. Rev. Lett.* **101,** 240401 (2008).
8. Deléglise, S. *et al. Nature* **455,** 510–514 (2008).
9. Sillanpää, M. A., Park, J. I. & Simmonds, R. W. *Nature* **449,** 438–442 (2007).
10. Majer, J. *et al. Nature* **449,** 443–447 (2007).

DIABETES

A virus–gene collaboration

Matthias von Herrath

Viral infection has long been implicated in the development of type 1 diabetes. Fresh evidence shows how some rare genetic variations might mediate a viral contribution to this autoimmune disorder.

Type 1 diabetes is an autoimmune disease in which the immune system attacks insulin-producing pancreatic β-cells. As is seen for many other disorders, genome-wide association studies have indicated that variations in certain genomic regions confer susceptibility to this disease, but the precise genes and genetic variants involved remain unknown. The genetic background to type 1 diabetes is complex: some genes protect against the disorder, whereas others promote its development; the two sets act in concert to confer an individual's degree of susceptibility. In addition, environmental factors such as viral infection are thought to contribute to the disease process[1]. Writing in *Science*, Nejentsev *et al.*[2] provide genetic evidence further implicating viral infection in the development of type 1 diabetes.

Of the 15 loci in the human genome that have previously been associated with type 1 diabetes, 10 candidate genes were subjected to a detailed analysis by the authors[2]. Specifically, they used the 454 sequencing technology for high-throughput, detailed sequencing of the genes' exons (coding sequences) and splice sites in 480 patients with type 1 diabetes and 480 healthy controls. They then tested whether the variants identified were associated with the disease in 30,000 patients and controls, including families with one or more children with the disorder.

This extensive analysis led Nejentsev *et al.* to identify four rare variations, each of which independently lowered the risk of developing type 1 diabetes by about 50%. Intriguingly, all four variants occur in the same gene, *IFIH1*, affecting the expression and structure of its protein product IFIH1. A helicase enzyme, IFIH1 (also known as MDA5) triggers the secretion of immune mediators known as interferons in response to viral infection. As disabling *IFIH1* expression lowers the risk of type 1 diabetes[2], this warrants a more detailed look at the role of viral infection as a potential culprit in at least some cases of the disorder.

Viruses, notably human enteroviruses, infect various organs, causing a range of disorders including damage to β-cells. It is unlikely, however, that they commonly cause type 1 diabetes in those with low genetic risk[3] and might, instead, contribute to this disorder by infecting β-cells and predisposing them to autoimmune attack. Animal studies have shown that human enteroviruses that replicate fast *in vivo* cause type 1 diabetes[4], whereas those that grow more slowly actually protect against it by activating various regulatory immune mechanisms[5].

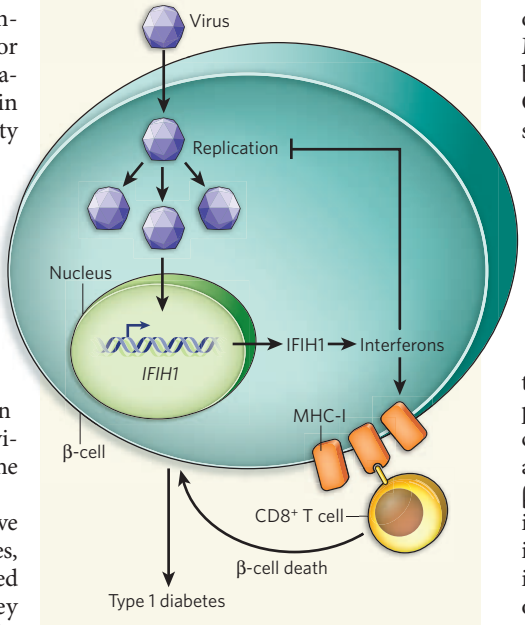


Figure 1 | Viral infection, the interferon response and diabetes. On infection with viruses such as human enteroviruses, interferon-response genes, including *IFIH1*, are activated in insulin-producing pancreatic β-cells, leading to increased levels of interferon. These immune mediators not only inhibit viral replication, but also enhance the expression of surface MHC-I molecules. Cytotoxic $CD8^+$ T cells recognize infected β-cells through the MHC-I molecules on their surface, damaging and eventually killing them. Thus, viral infection can contribute to the development of type 1 diabetes. In support of this proposal, Nejentsev *et al.*[2] show that rare variations in *IFIH1* protect against type 1 diabetes.

In the light of Nejentsev and colleagues' results, these observations are relevant to the hitherto rather tenuous link between infections with human enteroviruses and type 1 diabetes. Most, if not all, viral infections trigger the production of interferon by the host immune system (Fig. 1). Being an interferon-response gene, *IFIH1* allows the infected cell to sense the RNA genome of enteroviruses and increase interferon production. Interferons then primarily limit viral replication to prevent damage to the infected cell. They also, however, increase the visibility of the infected cell to the immune system by enhancing the expression of surface recognition molecules called major histocompatibility complex (MHC) molecules.

What do interferons do to the β-cells? Nothing good, based on animal studies. Together with other inflammatory factors, such as IL-1β and TNF-α, interferons directly contribute to the demise of β-cells[6]. Moreover, they induce the expression of MHC class I (MHC-I) molecules by these cells. Both effects promote the development of type 1 diabetes in mice[7], with enhanced expression of MHC-I proteins perhaps being particularly relevant. Normally, β-cells express low levels of these molecules. Following viral infections, and thus increased interferon production, β-cells express several 100-fold more MHC-I molecules, becoming highly susceptible to recognition and destruction by cytotoxic $CD8^+$ T cells, which recognize both viral and self antigens through MHC-I molecules[7]. At least in animal models, therefore, unmasking of β-cells to the immune system — associated with viral infection and mediated by interferons — could be an essential step in the development of type 1 diabetes[8].

As far as humans are concerned, accumulating evidence also supports a role for viral infection in the development of type 1 diabetes. Indeed, all the main ingredients seem to be present. First, $CD8^+$ T cells are the most common cell type present in human pancreatic islets[9], and clones of these cells that can kill human β-cells have been isolated[10]. Second, and more importantly, several laboratories have reported increased expression of MHC-I molecules (and in some cases, interferons) in pancreatic islets of patients with type 1 diabetes, but not those of controls[11]. Because many of the islets in these patients had not yet been targeted by the immune system, an underlying viral infection is a plausible explanation for these observations. Indeed, some studies have documented the presence of human enterovirus proteins in islets of patients with type 1 diabetes but not in those of controls[12], providing further backing for a role of viruses in the disorder (Fig. 1).

Together with Nejentsev and colleagues' observations, and another study[13] showing that *IFIH1* expression in peripheral blood cells is associated with type 1 diabetes, these earlier studies point us in a new direction for mechanistic studies of factors contributing to this disorder: virus-mediated interferon production. From an immunogenetic viewpoint, investigations similar to that of Nejentsev *et al.* should be carried out to define disease-associated variations in genes located within the MHC region, which shows the strongest association with human type 1 diabetes. Detecting structural variations in genes in different individuals will also allow us to devise tailored forms of immunotherapy for specific patient groups with the disorder.

From a virological and immunological standpoint, use should be made of well-preserved, freshly frozen human organ samples, such as those provided by nPOD (the Network for Pancreatic Organ Donors with Diabetes)[14], to directly isolate viral sequences, to clone viruses and to rigorously examine islets for

the presence of viral footprints such as interferons and increased MHC-I expression. If viral infection proves to be a contributing factor, we could aim to lower the risk of type 1 diabetes by designing suitable vaccines against human enteroviruses. ■

Matthias von Herrath is at the La Jolla Institute for Allergy and Immunology, San Diego, California 92121, USA.
e-mail: matthias@liai.org

1. Redondo, M. J., Jeffrey, J., Fain, P. R., Eisenbarth, G. S. & Orban, T. *N. Engl. J. Med.* **359,** 2849–2850 (2008).
2. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. *Science* **324,** 387–389 (2009).
3. Green, J., Casabonne, D. & Newton, R. *Diabet. Med.* **21,** 507–514 (2004).
4. Kanno, T. *et al. J. Virol.* **80,** 5637–5643 (2006); Erratum *J. Virol.* **80,** 8843 (2006).
5. Filippi, C. M., Estes, E. A., Oldham, J. E. & von Herrath, M. G. *J. Clin. Invest.* (in the press).
6. Eizirik, D. L., Colli, M. L. & Ortis, F. *Nature Rev. Endocrinol.* **5,** 219–226 (2009).
7. Seewaldt, S. *et al. Diabetes* **49,** 1801–1809 (2000).
8. von Herrath, M. G., Fujinami, R. S. & Whitton, J. L. *Nature Rev. Microbiol.* **1,** 151–157 (2003).
9. Pipeleers, D. *et al. Novartis Found. Symp.* **292,** 19–24; discussion 24–31, 122–129, 202–203 (2008).
10. Skowera, A. *et al. J. Clin. Invest.* **118,** 3390–3402 (2008).
11. Dotta, F. *et al. Proc. Natl Acad. Sci. USA* **104,** 5115–5120 (2007).
12. Richardson, S. J., Willcox, A., Bone, A. J., Foulis, A. K. & Morgan, N. G. *Diabetologia* **52,** 1143–1151 (2009).
13. Liu, S. *et al. Hum. Mol. Genet.* **18,** 358–365 (2009).
14. www.jdrfnPOD.org

ANIMAL BEHAVIOUR

Birdsong normalized by culture

W. Tecumseh Fitch

Both birdsong and human language are learned, requiring complex social input. New findings show, however, that bird populations 'seeded' with aberrant song input transform it to normal song in a few generations.

Around 650 BC, the Egyptian king Psammeticus reputedly had two children raised by a mute goatherd, isolated from all linguistic input, to determine what language they would speak spontaneously[1]. As legend has it, the children's first word was *bekos*, so the king judged the 'original language' to be Phrygian, in which *bekos* means bread (that *bekos* represented the bleat of a goat was apparently not considered). Today, such an experiment seems both unethical and naive, for we know from numerous examples that children deprived of linguistic input will not spontaneously speak any language. To acquire language, humans need rich linguistic input early in life[2,3]. In a study published on page 564 of this issue, Fehér and colleagues[4] greatly improve on Psammeticus's experiment, using laboratory-raised songbirds.

When isolated from normal input, humans develop only basic communication systems. Deaf children raised in hearing homes spontaneously create simple communication systems termed home-sign[5]. Lacking semantic and grammatical complexity, such systems cater only to simple, specific needs. Similarly, adults thrown together with no common language develop pidgins — communication systems with small vocabularies and simple grammar. But intriguingly, in both cases, later generations transform such simple systems into true languages, with a full range of grammatical expressiveness. Pidgins can become complex, stable languages known as creoles in just a few generations[6,7]: in Nicaragua, deaf children transformed a set of simple home-sign systems into the full Nicaraguan sign language (NSL) within a few decades, under the watchful eyes and video cameras of linguists[8].

But such fascinating examples raise more questions than they answer. Do the systems developed by isolated children reflect an original, biologically given form of language, as Psammeticus believed? Or is cultural transmission itself a necessary part of human language? Do children require rich lexical and grammatical input to fully activate the language-acquisition system? How do the neural networks underlying isolated languages such as home-sign differ from those that mediate communication with full languages? For ethical reasons, linguists cannot address these questions with controlled experiments, and interpretations of 'natural experiments' such as NSL remain controversial. A more biologically tractable model of cultural transmission is therefore needed.

One possibility involves investigating birdsong. Biologists have known since Darwin's time that, to sing properly, young songbirds require song input, typically during an early sensitive period preceding adulthood[9]. Deprived of such input, they sing only an abnormal song, with fewer syllable types and less structure than normal.

Darwin further recognized that learned birdsong nullifies traditional nature/nurture dichotomies. Birdsong is neither an instinct present at birth, nor an arbitrary cultural construction: it rests on biological foundations, but also requires specific inputs to develop properly. Songs are transmitted and transformed over several generations of learners, forming birdsong 'dialects' that vary geographically within a species[10]. Thus, both birdsong and language are generated by a species-typical 'instinct to learn' that constrains, but does not fully determine, the final outcome[11]. Birdsong acquisition provides, in this sense, an animal equivalent of human cultural transmission.

In their study, Fehér *et al.*[4] first raise young male zebra finches in isolation. At adulthood, these birds sing only a raspy, arrhythmic isolate (ISO) song. The authors then pair these adult male 'tutors' with young males. These first-generation young learners successfully imitate ISO song, but already begin to transform it in the direction of normal — wild-type — song. Iterating this process, using first-generation song as input to a second generation of birds and so on, the authors follow the birds for up to five generations and find that each generation's song moves steadily closer to normal zebra finch song.

Fehér and colleagues performed these initial experiments with otherwise-isolated males, tutored one-on-one. To explore the consequences of a more typical social environment, the authors[4] established a colony of females, socially housed and seeded with a single ISO male. Again, the songs of the descendant males transformed steadily in the direction of normal song. Intriguingly, however, these learner males' songs also incorporated their male siblings' song errors and female whistles. This 'cultural' experimental condition is more reminiscent of human sign languages such as NSL, in which a critical mass of young signers is needed to spark the development of a rich lexicon and grammatical complexity typical of true human language[12].

Of course, zebra-finch song is not language: birdsong is not used to express complex meanings; has only a simple syntax; and is typically produced only by males. But the techniques and model Fehér and colleagues introduce exploit a mechanism — cultural transmission of vocally learned signals — shared by both systems[13,14]. This new approach will allow detailed mechanistic explorations into how iterating the process of song acquisition can lead to something more diverse and structured than any individual bird creates on its own. What are the differences in brain circuitry between ISO birds and their fourth-generation descendants? In later generations, will we see better-developed song circuitry, or different activation patterns of the genes involved in vocal learning? What about females: do they prefer normal song, regardless of their upbringing, or does song resembling their father's ISO song excite them more?

Ultimately, addressing these issues will allow us to uncover the detailed biological bases of songbirds' instinct to learn song. If discoveries concerning genetic similarities in birdsong and human speech[15] are any guide, such work may also provide intriguing hints about the biological mechanisms underlying the language-acquisition system in humans. Thus, Fehér and colleagues' birdsong studies provide a pioneering approach for exploring the biology of culturally transmitted systems in the laboratory. ■

W. Tecumseh Fitch is in the Department of

Neurobiology and Cognition, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.
e-mail: tecfitch@mac.com

1. Herodotus *The Histories* (Penguin, 1964).
2. Lenneberg, E. H. *Biological Foundations of Language* (Wiley, 1967).
3. Kuhl, P. K. *Nature Rev. Neurosci.* **5,** 831–843 (2004).
4. Fehér, O., Wang, H., Saar, S., Mitra, P. P. & Tchernichovski, O. *Nature* **459,** 564–568 (2009).
5. Goldin-Meadow, S. & Mylander, C. *Nature* **391,** 279–281 (1998).
6. Mühlhäusler, P. *Pidgin and Creole Linguistics* revised edn (Univ. Westminster Press, 1997).
7. Sandler, W., Meir, I., Padden, C. & Aronoff, M. *Proc. Natl Acad. Sci. USA* **102,** 2661–2665 (2005).
8. Senghas, A., Kita, S. & Özyürek, A. *Science* **305,** 1779–1782 (2004).
9. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* 1st edn (John Murray, 1871).
10. Marler, P. & Slabbekoorn, H. *Nature's Music: The Science of Birdsong* (Academic, 2004).
11. Marler, P. in *The Epigenesis of Mind: Essays on Biology and Cognition* (eds Carey, S. & Gelman, R.) 37–66 (Lawrence Erlbaum Associates, 1991).
12. Senghas, A. & Coppola, M. *Psychol. Sci.* **12,** 323–328 (2002).
13. Jarvis, E. D. *Ann. NY Acad. Sci.* **1016,** 749–777 (2004).
14. Nottebohm, F. *Ann. NY Acad. Sci.* **280,** 643–649 (1976).
15. Haesler, S. *et al. PLoS Biol.* **5,** e321 (2007).

EARTH SCIENCE

Restoration of the noble gases

Tim Elliott

The noble gases emitted from deep inside the Earth have been sending mixed messages to those intent on deciphering them. A model that promises to help clear up the confusion is now on offer.

Geochemists have long taken a close interest in emissions from the bowels of the Earth. In particular, analyses of noble gases emanating from the mantle have been prominent in shaping models of Earth's structure. Lately, these data have seemed to be more paradoxical than illuminating, but Gonnermann and Mukhopadhyay (page 560 of this issue)[1] have revisited this puzzle and, by adding a new twist to an old concept, help us come to terms again with our planet's gassy innards.

Gas is continuously lost from Earth's interior, being carried to the surface in magmas produced by melting of the shallow mantle. The abundances and isotopic compositions of the trace amounts of noble gases thus erupted provide clues to Earth's evolution. In this field of study, the natural isotopic variability of helium has had an especially influential role. Helium has two stable isotopes, ^{3}He and ^{4}He. Earth's complement of ^{3}He was acquired during the planet's formation, whereas ^{4}He has been produced throughout Earth's history by the decay of the naturally occurring radionuclides of uranium and thorium. Thus, the ^{4}He/^{3}He ratio increases with time, with a magnitude dependent on the (U+Th)/He ratio.

It is well established that the ^{4}He/^{3}He ratios of magma from many ocean islands (such as Hawaii) are significantly lower than those of magma erupted at submarine mid-oceanic ridges. This observation has been thought to reflect 'degassing' of the upper mantle as a result of melting and crust formation at mid-ocean ridges, driven by the spreading of tectonic plates (Fig. 1a). Magma transports helium, uranium and thorium from the mantle to the surface. Although helium is ultimately lost to the atmosphere, non-volatile uranium and thorium remain in the crust to be subsequently returned to the mantle by plate subduction. Hence, the mantle involved in the plate-tectonic cycle attains a higher (U+Th)/He ratio and so evolves to higher ^{4}He/^{3}He ratios than any unmolested, 'primitive' mantle. The low ^{4}He/^{3}He ratio evident in some ocean islands was thus taken to reflect their derivation from such a primitive source that had been convectively isolated from the rest of the mantle. This meshed with the idea that ocean islands are the surface manifestation of mantle plumes that, like the wax in lava lamps, rise by thermal buoyancy from a deep, hot boundary layer (Fig. 1b).

Imagination can be allowed to run further riot within this conceptually appealing model. Early estimates of the composition of the continental crust suggested that it might have been derived from melting of the upper third of the mantle. This fraction spookily corresponds to the proportion of the mantle above a notable seismic feature at a depth of 660 kilometres. So the notion of a layered mantle was born, with the 660-km discontinuity dividing 'depleted' upper mantle from 'primitive' lower mantle (Fig. 2a). The calling card of the lower mantle was its low ^{4}He/^{3}He ratio, and this ratio acquired a mystical significance, such that it is even traditionally expressed in the opposite way to all other radiogenic isotope ratios (that is, as ^{3}He/^{4}He, with the isotope produced by radioactive decay in the denominator).

Although the layered-mantle model also helped to account for a series of other notable observations of noble gases (refs 2, 3, for example), it has proved inconsistent with geochemical constraints provided by much of the rest of the periodic table. There is little evidence for any extant primitive mantle as assessed from a wide range of non-volatile element abundance and isotope ratios (ref. 4, for example). Moreover, seismological studies convincingly demonstrated[5] that the 660-km discontinuity is not a boundary to mantle flow, and so the notion that primitive mantle can be preserved in the lower mantle became untenable (Fig. 2b). Thus, some researchers[6,7] have argued for an entirely different mechanism for creating low ^{4}He/^{3}He ratios in the mantle. Yet noble-gas mythology is deep-rooted, and others have rebranded the layered mantle with a deeper boundary[8], or suggested that we live in a unique time in which long-term layering has only just been breached[9]. So it has increasingly seemed that noble-gas

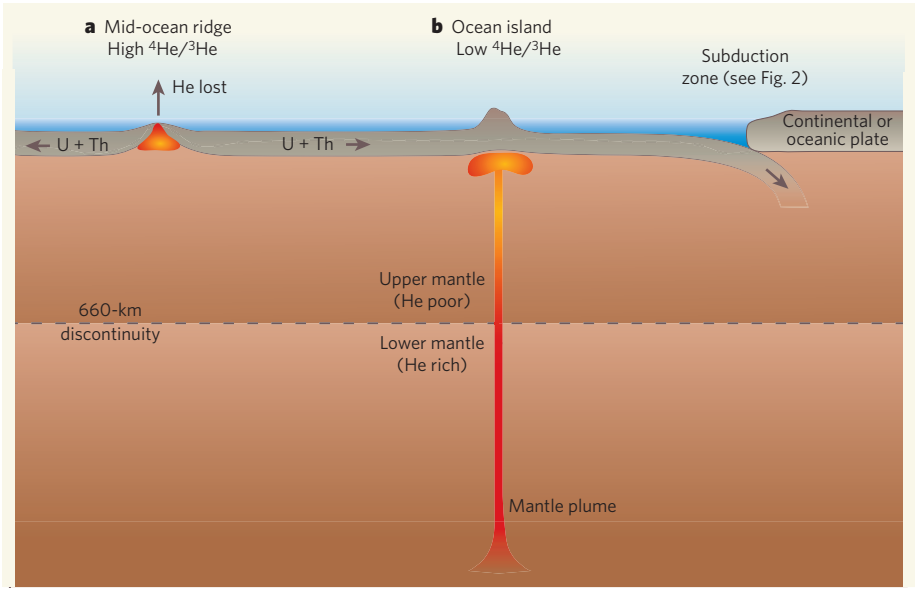


Figure 1 | Isotope ratios and Earth's mantle. a, As oceanic plates are pulled apart at mid-ocean ridges, the upper mantle rises in their place and (partially) melts. Uranium, thorium and helium in this portion of mantle are transferred to the magma, which migrates to the surface to form crust. Helium is lost during crystallization of the melts, but uranium and thorium are retained in the crust and are ultimately returned to the mantle by plate subduction (Fig. 2). Thus, the upper mantle becomes 'degassed' and the (U+Th)/He ratio increases, which with time translates into higher ^{4}He/^{3}H ratios. **b,** By contrast, ocean islands show a low ^{4}He/^{3}He ratio, thought to reflect a deep-mantle source of underlying mantle plumes. Graphic not to scale.

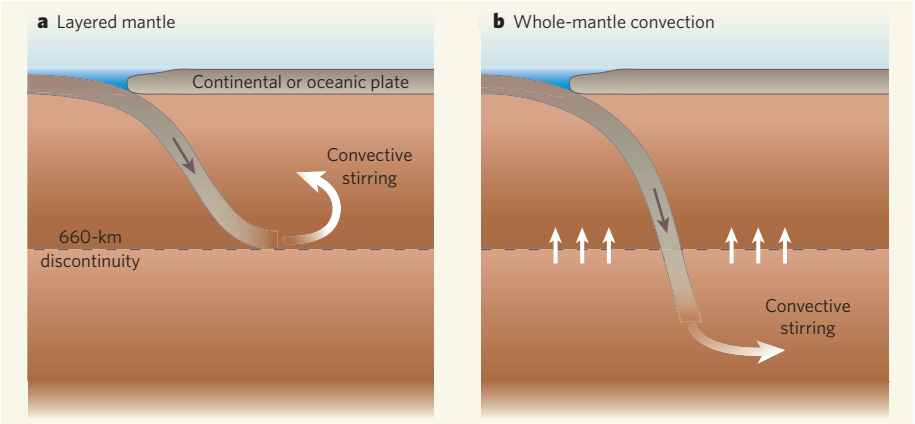


Figure 2 | The subduction connection. a, Models of noble-gas evolution have classically implied that only the upper mantle is effectively degassed[2,3], with the 660-kilometre seismic discontinuity producing a layered mantle by representing a boundary to the subduction of plates and the flow of material into the deep mantle. **b**, Seismological images[5], however, now provide strong evidence that plates can penetrate into the lower mantle, with the associated counterflow producing whole-mantle convection. The model of Gonnermann and Mukhopadhyay[1] shows that, contrary to many expectations, this mode of whole-mantle convection is quite compatible with observations of helium isotopes and other noble gases. Graphic not to scale.

measurements pose more questions than they provide answers for in our understanding of Earth's interior[10].

Gonnermann and Mukhopadhyay[1] tackle these issues with alarmingly simple finesse. They propose a solution in which the lower mantle is not isolated, but just sluggish. This is not unreasonable because mantle viscosity is believed to increase significantly with depth. The authors construct a model in which exchange between the shallow and the deep mantle is allowed, with plate subduction occurring through the 660-km discontinuity, and the associated counterflow (Fig. 2b) producing a more homogeneous, less-layered mantle. But this is not an all-or-nothing process, with the deep subduction being limited to a modest flux consistent with the seismological observations. The authors explore the parameter space in which their model can reproduce the helium isotope ratios of mid-ocean-ridge basalts and the ocean islands. The inputs to successful solutions are quite compatible with independent constraints on terrestrial gas budgets and the isotope ratios of non-volatile elements such as neodymium ($^{143}Nd/^{144}Nd$, for example).

The authors' model[1] consists of an elegant mass balance, similar to one that has been used to reproduce the natural range of isotopic compositions in other systems[11]. It is not a dynamic model, but the requirements of the successful model solutions do not seem onerous, even if they clearly need to be explored using numerical simulations of convection. The essential feature, that the mantle is not layered but not homogeneous, also seems eminently reasonable, and it is perhaps surprising that such an idea has previously not been sufficiently explored. Yet many ideas seem obvious only after they have been shown to be effective.

Before feeling too comfortable, however, it must be remembered that there are the mysteries of xenon and its nine isotopes to get to grips with[12]. And perhaps ironically, the work of Gonnermann and Mukhopadhyay[1] arrives at a time when the Earth science community is grappling with striking evidence[13] for a 'hidden reservoir' that formed at the base of the mantle within the first 30 million years of the birth of the Earth, some 4.5 billion years ago. So it seems that we now have an embarrassment of hide-outs for noble gases (see also ref. 14), and several possible solutions to our volatile dilemmas. ■

Tim Elliott is in the Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK.
e-mail: tim.elliott@bristol.ac.uk

1. Gonnermann, H. M. & Mukhopadhyay, S. *Nature* **459,** 560–563 (2009).
2. Porcelli, D. & Wasserburg, G. J. *Geochim. Cosmochim. Acta* **59,** 4921–4937 (1995).
3. Allègre, C. J., Hofmann, A. & O'Nions, K. *Geophys. Res. Lett.* **23,** 3555–3557 (1996).
4. Hofmann, A. W. *et al. Earth Planet. Sci. Lett.* **79,** 33–45 (1986).
5. van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. *Nature* **386,** 578–584 (1997).
6. Class, C. & Goldstein, S. L. *Nature* **436,** 1107–1112 (2005).
7. Parman, S. W. *et al. Nature* **437,** 1140–1143 (2005).
8. Kellogg, L. H., Hager, B. H. & van der Hilst, R. D. *Science* **283,** 1881–1884 (1999).
9. Allègre, C. J. *Phil. Trans. R. Soc. Lond. A* **360,** 2411–2431 (2002).
10. Ballentine, C. J. *et al. Phil. Trans. R. Soc. Lond. A* **360,** 2611–2631 (2002).
11. Kellogg, J. B., Jacobsen, S. B. & O'Connell, R. J. *Earth Planet. Sci. Lett.* **204,** 182–202 (2002).
12. Pepin, R. O. & Porcelli, D. *Earth Planet. Sci. Lett.* **250,** 470–485 (2006).
13. Boyet, M. & Carlson, R. W. *Science* **309,** 576–581 (2005).
14. Jephcoat, A. P., Bouhifd, M. A. & Porcelli, D. *Phil. Trans. R. Soc. Lond. A* **366,** 4295–4314 (2008).

OLFACTION

Noses within noses

Steven D. Munger

The mammalian olfactory system does more than just detect food odours and pheromones. The discovery of a novel class of olfactory receptor provides evidence that mammals can also sniff out cell damage and disease.

The mammalian olfactory system recognizes diverse chemical stimuli conveying information about such things as food quality, the genetic identity or sexual status of potential mates, and even stress[1,2]. An exciting paper by Rivière *et al.*[3] (page 574 of this issue) describes the identification of a previously unrecognized type of chemosensory neuron in the rodent nose that responds to stimuli associated with cell damage, disease and inflammation. These results should help us to understand how animals identify pathogens or assess the health status of potential partners.

Not so long ago, it was believed that the olfactory system of most mammals had only two divisions: a main olfactory system that detects environmental odours, for instance those emitted by food or predators, and an accessory (vomeronasal) olfactory system that detects pheromones — intraspecies chemical signals that elicit a stereotyped behavioural or hormonal change. It is now clear that the sense of smell is much more complex. Indeed, the main and accessory olfactory systems each respond to both general odours and pheromones[4–6]. Furthermore, each olfactory division contains several types of sensory cell identified by the receptors and other proteins they express, the connections they make in the olfactory part of the brain, and the chemical stimuli to which they respond[2]. This diversity of sensory cells in the nose has given rise to the concept of olfactory subsystems, each dedicated to a particular chemosensory role[2].

There is a growing literature indicating that animals use olfaction to assess whether other organisms may be dangerous, or even to judge the health status of potential partners. For example, mice use olfactory cues to avoid potential mates that are infected with parasites[7], whereas nematode worms develop aversions to odours given off by harmful bacteria, thereby avoiding toxic food[8]. However, although such olfactory-based aversion behaviours have been documented, no olfactory subsystem that is dedicated to the assessment of health status or disease has been identified in mammals. The findings of Rivière and colleagues[3] may provide this missing link.

In mammals, most olfactory receptors are

members of the G-protein-coupled receptor (GPCR) superfamily, which signal the presence of stimuli by initiating a cascade of biochemical changes within chemosensory cells[2,9]. Reasoning that novel vomeronasal receptors would similarly belong to this group of proteins, Rivière and colleagues screened mouse vomeronasal tissue for the expression of about 100 known and candidate GPCRs. They found that rodent vomeronasal sensory neurons (VSNs) express five formyl peptide receptor (FPR)-like genes (*Fpr-rs1*, *Fpr-rs3*, *Fpr-rs4*, *Fpr-rs6* and *Fpr-rs7*)[10]. These genes are expressed in subsets of VSNs that do not express other known vomeronasal receptors, indicating that *Fpr-rs*-expressing VSNs are a novel group of chemosensory neurons.

FPR and FPR-like receptors are also found in immune cells, where they mediate cellular responses to cell damage, disease and inflammation[10]. Rivière and colleagues[3] engineered cells to express the VSN-specific FPR-rs proteins on their surface and found that they responded to several of the same activators (agonists) that stimulate the immune-cell FPR and FPR-like proteins. These agonists included peptides and lipids derived from microorganisms, or involved in the inflammatory response. Each type of FPR-rs responded to a distinct, but overlapping, panel of agonists (Fig. 1a).

But do these diverse FPR-rs agonists actually activate VSNs? Rivière and colleagues addressed this question by exposing isolated VSNs and VSNs in intact vomeronasal epithelium to several of the FPR-rs agonists. They measured cellular activation by imaging changes in calcium levels within the VSNs. Responses to FPR-rs agonists were concentration dependent and were observed at nanomolar concentrations, reminiscent of the exquisite sensitivity of other VSN subpopulations to volatile and peptide stimuli[5]. Agonist selectivity also varied among individual VSNs, a finding consistent with the pattern of FPR-rs expression across the vomeronasal epithelium. This arrangement would ensure that VSNs can discriminate between different FPR-rs agonists, while at the same time allowing for the detection of diverse chemical stimuli. A similar strategy is seen in the retina, which uses several related GPCRs called opsins to detect light. The different opsins have overlapping spectral tuning curves that permit the detection of light across the visual spectrum. Furthermore, the non-overlapping expression of these different opsins in subsets of retinal photoreceptors, each with distinct connections to the brain, contributes to our ability to discriminate colour.

Rivière *et al.*[3] provide compelling evidence for a class of olfactory receptor that is expressed in a unique subpopulation of VSNs and that is responsive to novel chemosensory stimuli. However, it will be essential to show that behavioural or physiological responses to the presence of pathogens, spoiled food or diseased animals depend on activation of this group of VSNs before we can firmly conclude that they are part of an olfactory subsystem dedicated to discerning pathogenicity or health status.

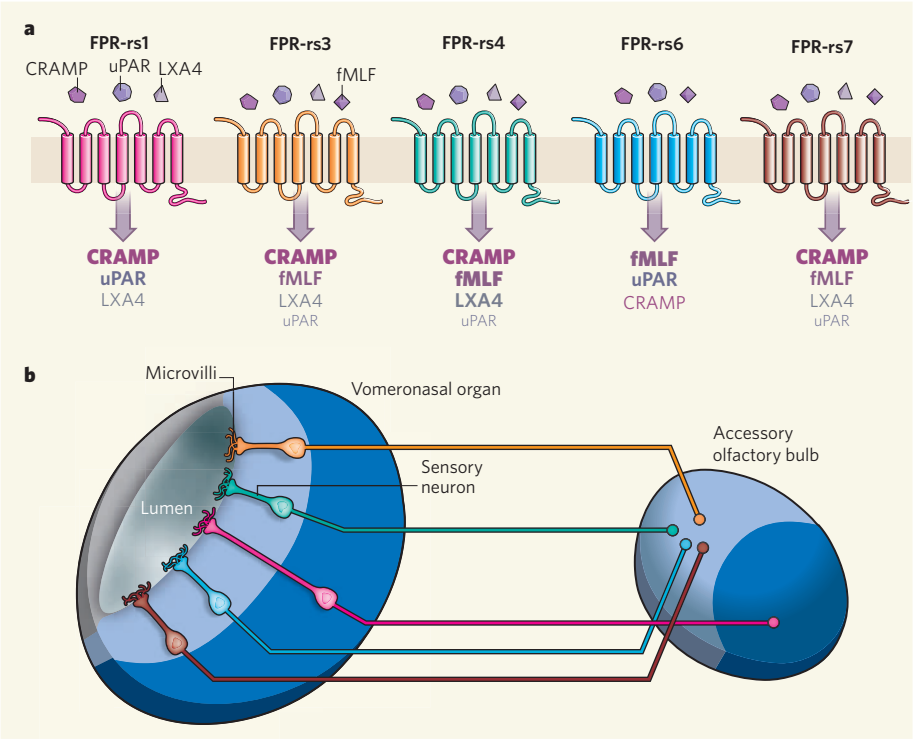


Figure 1 | Vomeronasal receptors discovered by Rivière and colleagues[3]. **a**, The formyl peptide receptor (FPR)-rs proteins are G-protein-coupled receptors; they have seven membrane-spanning domains and interact with G proteins (not shown). Rivière *et al.* found that five FPR-rs proteins that are present on sensory neurons in the vomeronasal epithelium are differentially responsive to several compounds involved in cell damage and inflammation: the peptides CRAMP, fMLF and uPAR, and the lipid lipoxin A4 (LXA4). The relative efficacy of each agonist for a given receptor is represented by the size and thickness of the text. **b**, Chemosensory stimuli entering the lumen of the vomeronasal organ interact with FPR-rs receptors on sensory neuron microvilli. Receptor activation initiates neuronal signals that travel to the accessory olfactory bulb. The sensory epithelium of the vomeronasal organ is segregated into two zones (light blue and dark blue) that are differentiated by physical position and molecular markers. This segregation is maintained in the accessory olfactory bulb of the brain. Rivière *et al.*[3] report that, unlike members of known vomeronasal-receptor families that are expressed in sensory neurons of either the apical or basal zone, FPR-rs receptors are found in sensory neurons in both zones (each colour represents the FPR-rs receptor expressed by that VSN as in **a**), suggesting that sensory signals detected by related receptors may be processed by different parts of the accessory olfactory system.

Intriguingly, there are already clues that some FPR-rs-expressing VSNs may have disparate chemosensory roles. Four FPR-rs proteins are restricted to VSNs found in the apical zone of the vomeronasal epithelium, whereas one (FPR-rs1) is instead found only in VSNs of the basal zone (Fig. 1b). The segregation of these two VSN populations, which can be identified by the G-protein subunits they express, is maintained in the olfactory part of the brain (Fig. 1b), suggesting that the apical and basal VSNs encode different types of chemosensory information[5,6]. Thus, stimuli that can activate both FPR-rs1 and at least one other FPR-rs protein could evoke two distinct perceptions, behavioural responses or physiological changes. There is evidence in the main olfactory system that the sensory information extracted from the same odour compound differs depending on which cells detect it[11]. Similar strategies may be at work in the accessory olfactory system.

The nose is a busy place for researchers right now, with the discovery of several families of olfactory receptors and a multitude of distinct sensory cell types. As the biological roles of individual olfactory subsystems are elucidated, we can begin to truly understand how animals detect and dissect their complex chemosensory worlds. ■

Steven D. Munger is in the Department of Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.
e-mail: smung001@umaryland.edu

1. Bargmann, C. I. *Nature* **444,** 295–301 (2006).
2. Munger, S. D., Leinders-Zufall, T. & Zufall, F. *Annu. Rev. Physiol.* **71,** 115–140 (2009).
3. Rivière, S., Challet, L., Fluegge, D., Spehr, M. & Rodriguez, I. *Nature* **459,** 574–577 (2009).
4. Baxi, K. N., Dorries, K. M. & Eisthen, H. L. *Trends Neurosci.* **29,** 1–7 (2006).
5. Brennan, P. A. & Zufall, F. *Nature* **444,** 308–315 (2006).
6. Halpern, M. & Martinez-Marcos, A. *Prog. Neurobiol.* **70,** 245–318 (2003).
7. Kavaliers, M., Choleris, E. & Pfaff, D. W. *Neurosci. Biobehav. Rev.* **29,** 1347–1359 (2005).
8. Zhang, Y., Lu, H. & Bargmann, C. I. *Nature* **438,** 179–184 (2005).
9. Migeotte, I., Communi, D. & Parmentier, M. *Cytokine Growth Factor Rev.* **17,** 501–519 (2006).
10. Mombaerts, P. *Nature Rev. Neurosci.* **5,** 263–278 (2004).
11. Kobayakawa, K. *et al. Nature* **450,** 503–508 (2007).

Asteroids and andesites

Arising from: J. M. D. Day *et al. Nature* **457,** 179–182 (2009)

The production of terrestrial andesites in subduction zones is well established. Day *et al.*[1] describe two examples of meteorites (GRA 06128 and GRA 06129) that they claim to represent "an entirely new mode of generation of andesite crust compositions" on asteroids; this suggestion has wide implications for the generation of andesitic planetary crusts in general. However, here we show that compositional data, particularly for the rare-earth elements (REEs) and other lithophile elements, presented in their paper[1] do not substantiate this claim. We conclude that existing mechanisms for andesite generation do not need revision.

Day *et al.*[1] present two relevant figures. The first (Fig. 1a in ref. 1) is a plot of Na_2O plus K_2O versus SiO_2 that falls marginally within the andesite field. We consider that this figure could be misleading, as it is dominated by an exceptionally high Na_2O content (6%) and should not have been used to infer that their samples have andesitic affinities.

The second figure (Fig. 1b in ref. 1) is a chondrite-normalized REE plot; we note that Day *et al.* also present data for the age, platinum group elements, Re–Os systematics and oxygen isotopes of the samples that do not directly address their thesis. REE values are very low (La varies between 0.95 and 1.48 p.p.m. or 2.6 to 4 times chondritic using volatile-free CI values for normalization). The chondrite-normalized REE patterns in these meteorites are flat but with a significant enrichment in Eu. These patterns are comparable, if a little lower than those of 'Eu-rich', albeit otherwise 'normal' eucrites[2] (basaltic meteorites derived from the asteroid 4 Vesta) and do not at all resemble that of the 'average terrestrial andesite' REE pattern (or the majority of terrestrial arc andesites) that is also shown in the same figure. This pattern has a strong enrichment in the light REEs (for example, La = 16 p.p.m. (ref. 3) compared to 0.95–1.48 p.p.m. La for the meteorites). We believe that differences by factors of more than 10 need to be considered.

From the data given in the Supplementary Information of ref. 1, the sodium content (around 6%) is a factor of two higher, and the potassium content (about 0.2%) a factor of six lower, than averages for andesites or for the continental crust of the Earth. Indeed, the elevated silica in these rocks results from the high albite composition of the plagioclase, which dominates the mineralogy. In addition to the REEs, other significant lithophile elements such as Li, Rb, Sr, Cs, Zr, Hf, Nb, Ba, Pb, U and Th also show major discrepancies compared to their values in the continental crust, with the most incompatible of these (Th, Cs) being depleted by factors of 40 or more.

So although these meteorites are very interesting examples with high silica values (around 55%) that arose early (4.52 Gyr) on an asteroid, their compositions are fundamentally different from terrestrial andesites, and this study does not demonstrate that andesitic crusts resembling anything like the continental crust of the Earth might have evolved on such asteroids. Labelling these samples as 'andesite' carries the implication that the origin of these rocks has lessons for the formation of the continental crust of the Earth or elsewhere, and could be misleading.

Richard Arculus[1], Ian H. Campbell[1], Scott M. McLennan[2] & Stuart Ross Taylor[1]

[1]Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory 0200, Australia.
e-mail: Ross.Taylor@anu.edu.au
[2]Department of Geosciences, State University of New York at Stony Brook, Stony Brook, New York 11794-2100, USA.

Received 2 February; accepted 24 April 2009.

1. Day, J. M. D. *et al.* Early formation of evolved asteroidal crust. *Nature* **457,** 179–182 (2009).
2. Mittlefehldt, D. W. & Lindstrom, M. M. Geochemistry of diverse lithologies in Antarctic eucrites. *Lunar Planet. Sci.* **XIX,** abstr. 790-1 (1988).
3. Taylor, S. R. & McLennan, S. M. *Planetary Crusts: Their Composition, Origin and Evolution* Table 11.4 (Cambridge Univ. Press, 2009).

doi:10.1038/nature08077

Day et al. *reply*

Replying to: R. Arculus, I. H. Campbell, S. M. McLennan & S. R. Taylor *Nature* **459,** doi:10.1038/nature08077 (2009)

Arculus *et al.*[1] raise an important question regarding the use of terrestrial rock nomenclature to characterize extraterrestrial materials. Here the issue relates to the definition of 'andesite', and whether the felsic achondrite meteorites GRA 06128 and GRA 06129 (GRA 06128/9) can and should be classified using this term[2]. More broadly, the question is whether extraterrestrial rocks should be classified using standard petrologic and geochemical criteria, such as mineralogy and major-element bulk composition, developed for the description of terrestrial rocks[3,4]. The approach of Arculus *et al.*[1] is that petrogenetic process is of equal or greater importance when classifying a rock. This question must ultimately be decided by the scientific community.

Examples of the difficulties associated with extraterrestrial rock nomenclature are numerous in planetary petrology. Materials derived from the lunar maria are termed 'basalt' despite enormous diversity in their trace-element compositions, and have a different mode-of-origin as compared to terrestrial basalts[5]. Furthermore, meteoritic 'basalts' from Mars and from the asteroid belt (basalt is a term used by Arculus *et al.*[1] to describe the eucrites) carry with them no inference that they formed under the same petrogenetic conditions as terrestrial basalts. Similarly, 3.9 to 4.4-Gyr-old lunar felsites displaying dry mineral assemblages have been commonly termed 'granites'[6], despite some opinion that granite can only form in water-rich environments[7]. The Moon is also populated by norite, gabbro, anorthosite, troctolite, granulite and so forth[5], all of which are nomenclature terms used to describe terrestrial rocks and whose lunar genesis differs from their terrestrial counterparts.

Terrestrial andesite is commonly defined using two classification schemes from the International Union of Geological Sciences[3,4]. First, with a petrological classification, as a rock chiefly composed of plagioclase feldspar with subordinate accessory minerals, such as pyroxene, olivine or hornblende. Second, using the chemical classification as shown in Fig. 1 of our paper[2]. GRA 06128/9 qualify as andesite using either of these classifications. The rare-earth elements (REEs) and other lithophile trace elements have not commonly been used to define terrestrial or extra-terrestrial rocks, as suggested by Arculus *et al.*[1]; however, we acknowledge that these elements are important petrogenetic indicators.

We suggest that some asteroids may have crustal material that shares compositional and mineralogical similarities with the Earth's average continental crust, but we do not suggest a similar mechanism for its generation[2]. The generation of GRA 06128/9 is probably not plate tectonics, but through low degrees of partial melting of a relatively undifferentiated asteroidal body, as we demonstrated, and for which the Re–Os and oxygen isotope, age and platinum-group element data provide important evidence. This process may have been an important mechanism for evolved crust formation in the early Solar System[2]. Indeed, if rocks with andesitic compositions are discovered on planetary bodies other than Earth (for example, Mars or asteroids), they might have origins that differ from terrestrial andesites, which are commonly products of water-charged subduction processes.

Given the prior applications of terrestrial rock nomenclature to extraterrestrial rocks, as well as the common petrologic and geochemical definitions of the term andesite, we believe it is an appropriate term to use to describe the achondrites GRA 06128/9.

J. M. D. Day[1], R. D. Ash[1], Y. Liu[2], J. J. Bellucci[1], D. Rumble III[3], W. F. McDonough[1], R. J. Walker[1] & L. A. Taylor[2]

[1]Department of Geology, University of Maryland, College Park, Maryland 20742, USA.
e-mail: jamesday@geol.umd.edu
[2]Department of Earth and Planetary Sciences, Planetary Geosciences Institute, University of Tennessee, Knoxville, Tennessee 37996, USA.
[3]Geophysical Laboratory, Carnegie Institution of Washington, Washington DC 20015, USA.

1. Arculus, R., Campbell, I. H., McLennan, S. M. & Taylor, S. R. Asteroids and andesites. *Nature* **459**, doi:10.1038/nature08077 (2009).
2. Day, J. M. D. *et al.* Early formation of evolved asteroidal crust. *Nature* **457**, 179–182 (2009).
3. Le Maitre, R. W. (ed.) *A Classification of Igneous Rocks and Glossary of Terms: Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks* (Blackwell Scientific, 1989).
4. Le Bas, M. J. & Streckeisen, A. L. The IUGS systematics of igneous rocks. *J. Geol. Soc. Lond.* **148**, 825–833 (1991).
5. Jolliff, B. L., Wiezorek, M. A., Shearer, C. K. & Neal, C. R. (eds) *New Views of the Moon* (Reviews in Mineralogy and Geochemistry, Vol. 60, Mineralogical Society of America, 2006).
6. Bonin, B., Bébien, J. & Masson, P. Granite: a planetary point of view. *Gondwana Res.* **5**, 261–273 (2002).
7. Campbell, I. H. & Taylor, S. R. No water, no granites — no oceans, no continents. *Geophys. Res. Lett.* **10**, 1061–1064 (1983).

doi:10.1038/nature08078

ARTICLES

Generation of transgenic non-human primates with germline transmission

Erika Sasaki[1], Hiroshi Suemizu[1], Akiko Shimada[1], Kisaburo Hanazawa[2], Ryo Oiwa[1], Michiko Kamioka[1], Ikuo Tomioka[1,3], Yusuke Sotomaru[5], Reiko Hirakawa[1,3], Tomoo Eto[1], Seiji Shiozawa[1,4], Takuji Maeda[1,4], Mamoru Ito[1], Ryoji Ito[1], Chika Kito[1], Chie Yagihashi[1], Kenji Kawai[1], Hiroyuki Miyoshi[6], Yoshikuni Tanioka[1], Norikazu Tamaoki[1], Sonoko Habu[7], Hideyuki Okano[4] & Tatsuji Nomura[1]

The common marmoset (*Callithrix jacchus*) is increasingly attractive for use as a non-human primate animal model in biomedical research. It has a relatively high reproduction rate for a primate, making it potentially suitable for transgenic modification. Although several attempts have been made to produce non-human transgenic primates, transgene expression in the somatic tissues of live infants has not been demonstrated by objective analyses such as polymerase chain reaction with reverse transcription or western blots. Here we show that the injection of a self-inactivating lentiviral vector in sucrose solution into marmoset embryos results in transgenic common marmosets that expressed the transgene in several organs. Notably, we achieved germline transmission of the transgene, and the transgenic offspring developed normally. The successful creation of transgenic marmosets provides a new animal model for human disease that has the great advantage of a close genetic relationship with humans. This model will be valuable to many fields of biomedical research.

The use of transgenic mice has contributed immensely to biomedical science. However, the genetic and physiological differences between primates and mice—including their neurophysiological functions, metabolic pathways, and drug sensitivities—hamper the extrapolation of results from mouse disease models to direct clinical applications in humans. Thus, the development of non-human primate models that mimic various human systems would accelerate the advance of biomedical research. In particular, genetically modified primates would be a powerful human disease model for preclinical assessment of the safety and efficacy of stem-cell or gene therapy.

The common marmoset (*Callithrix jacchus*) is a small New World primate that, because of its size, availability, and unique biological characteristics[1], has attracted considerable attention as a potentially useful biomedical research animal in fields such as neuroscience, stem cell research, drug toxicology, immunity and autoimmune diseases, and reproductive biology. Marmosets have a relatively short gestation period (about 144 days), reach sexual maturity at 12–18 months, and females have 40–80 offspring during their life. Therefore, the application of transgenic techniques to marmosets may be feasible, and would greatly facilitate the study of human disease. In contrast, the more commonly used Old World primates, such as the rhesus monkey (*Macaca mulatta*) and cynomolgus monkey (*Macaca fascicularis*), show slow sexual maturation (about 3 years) and have fewer offspring (around 10) over the female lifespan. Thus, even though marmosets are less closely related to humans than either apes or Old World primates, their potential as transgenic primate models of human disease means they may be uniquely valuable.

Obtaining large numbers of oocytes from primates for transgenic experiments is limited by ethical and economic constraints. However, because retroviral vectors allow the efficient integration of a provirus into the host genome[2–4], their use requires fewer oocytes than some other techniques. Furthermore, the injection of a lentiviral vector into the perivitelline space of a pre-implantation embryo, which is less invasive than injection into the pronucleus, is an advantageous method for generating transgenic animals. In fact, transgenic modification of rhesus monkeys using retroviral vectors and a lentiviral vector[5–7] has been attempted. In these studies, genomic integration and expression of the transgene was observed in the placenta, but not in the infants' somatic tissues, by objective analyses such as PCR with reverse transcription (RT–PCR) or western blotting.

The recombinant adeno-associated virus has been used for the targeted knockout of the cystic fibrosis transmembrane conductance receptor gene in swine fetal fibroblasts, and targeted gene knockout pigs have been generated by somatic cell nuclear transfer (SCNT) of the fibroblast nuclei into oocytes[8,9]. Although conceptually this method could be used to make targeted gene-knockout primates, marmoset SCNT techniques are not available at present.

Here we successfully produced transgenic marmosets, by injecting a lentiviral vector containing an enhanced green fluorescent protein (EGFP) transgene[10] into marmoset embryos. Four out of five transgenic marmosets expressed the EGFP transgene in neonatal tissues; the fifth expressed it in the placenta. Two showed transgene expression in the germ cells, and one fathered a healthy transgenic neonate. Our method for producing transgenic primates promises to be a powerful tool for studying the mechanisms of human diseases and developing new therapies.

Production of transgenic marmosets

In a pilot study, we showed that pre-implantation marmoset embryos obtained through natural intercourse had much better developmental potential than embryos obtained by *in vitro* fertilization (IVF). Therefore, both natural and IVF embryos were used in this study.

[1]Central Institute for Experimental Animals, 1430 Nogawa, Miyamae-ku, Kawasaki, Kanagawa 216-0001, Japan. [2]Department of Urology, Juntendo University Nerima Hospital 3-1-10 Takanodai, Nerima-ku, Tokyo 177-8521, Japan. [3]Center for Integrated Medical Research, [4]Department of Physiology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. [5]Natural Science Centre for Basic Research and Development, Hiroshima University 1-2-3, Kasumi, Minami-ku, Hiroshima 734-8551, Japan. [6]Subteam for Manipulation of Cell Fate, RIKEN BioResource Centre, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan. [7]Department of Immunology, Tokai University School of Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan.

To introduce the EGFP gene into the marmoset embryo, three kinds of self-inactivating lentiviral vectors were constructed on the basis of human immunodeficiency virus type 1 (HIV-1), and each carried a different promoter, CAG, CMV or EF1-α. The self-inactivating lentiviral vectors were named CAG-EGFP, CMV-EGFP and EF1-α-EGFP, respectively.

All lentiviral vector injections were performed at the earliest embryonic stage possible using an Eppendorf FemtoJet express and a Narishige micromanipulator. Twenty-seven IVF embryos and 64 natural embryos were injected with a high titre of the lentiviral vector, from 5.6×10^9 to 5.6×10^{11} transducing units per ml (Table 1). Because the perivitelline space of the marmoset early embryo is rather small, 16 of the 27 IVF embryos, and 49 of the 64 natural embryos, at the pronuclear-to-morula stage, were first placed in 0.25 M sucrose in PB1 medium (0.25 M sucrose medium), which made the perivitelline space expand 1.2–7.5-fold (data not shown). The lentiviral vector was then injected into the perivitelline space (Supplementary Data 1). Virus was injected into the blastocoel of the remaining 11 IVF and 15 natural embryos at the blastocyst stage, without the 0.25 M sucrose treatment (Supplementary Data 1).

Immediately after injection, 4 of the IVF and 12 of the natural embryos were transferred to recipient females. The rest were examined for the expression of EGFP, starting 48 h after injection. Among the sucrose-treated IVF and natural early embryos at 48 h after injection, 68.8% and 97.7% expressed EGFP, respectively; of the non-sucrose-treated IVF and natural embryos injected with lentivirus as blastocysts, 85.7% and 87.5% expressed EGFP, respectively (Supplementary Data 1). Therefore, 61 of the natural embryos and 19 of the IVF embryos were transferred to surrogate mothers (Table 1). For the transfers, the recipients were synchronized with the donor oocyte cycle; each recipient received 1–3 embryos per cycle, and 50 surrogate mother animals were used.

Of the surrogate mothers, seven that received natural or IVF embryos became pregnant. Three recipients miscarried on days 43, 62 and 82, and the other four delivered five healthy offspring (three singletons, one pair of twins), one male (number 666) and four females, on days 144–147 after ovulation (Fig. 1). For the infants, the lentiviral vector injection had been performed at the four-cell stage (584), the pronuclear stage (587), and the morula stage (588, 594 and 666). The EGFP transgene was driven by the CAG promoter in three newborns (584, 587 and 588) and by the CMV promoter in the other two (594 and 666; Supplementary Data 1).

Table 1 | Production rates of transgenic marmosets

	Artificial reproductive technique	Natural
Number of GV oocytes	460	No data
Number of matured oocytes (only MII)	201	No data
Number of IVFs performed (including MI)	272	No data
Number of fertilized oocytes	121	No data
Fertilization rate (fertilization per GV)	26.3%	No data
Fertilization rate (fertilization per IVF)	44.5%	No data
Lentiviral injections	27	64
EGFP expression confirmed after 48 h or later	23	52
EGFP expression	17	50
EGFP expression rate	73.9%	96.2%*
ETs	19	61
Number of surrogates	13	37
Number of pregnancies	1	6
Number of deliveries	1	3
Births	1	4
Birth rate (birth per ET)	5.20%	6.55%
Number of Tgs	1	4
Production rate (Tg per injection)	3.70%	6.25%
Production rate (Tg per ET)	5.26%	6.25%
Production rate (Tg per birth)	100	100

ET, embryo transfer; GV, germinal vesicle; MI, metaphase I; MII, metaphase II; Tg, transgenes.
*$P < 0.01$, chi-squared analysis.

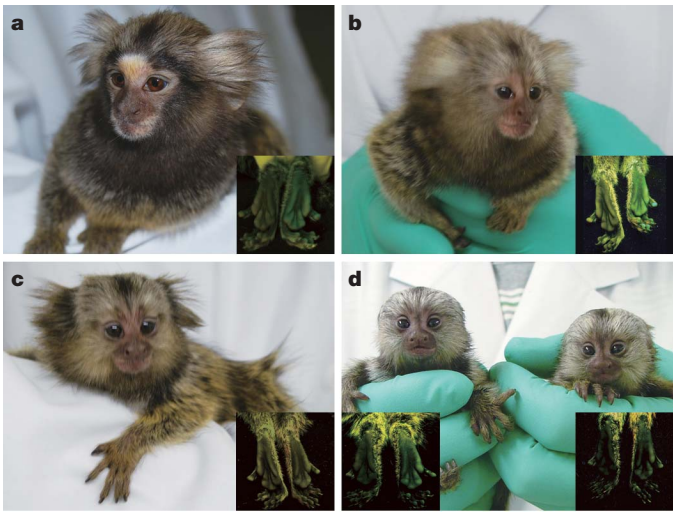


Figure 1 | Self-inactivating lentiviral vector-derived EGFP transgenic marmosets. a–d, The transgenic marmoset infants are shown. Shown are 584 (Hisui) (**a**), 587 (Wakaba) (**b**), 588 (Banko) (**c**), and twin infants 594 (Kei)/666 (Kou) (**d**). 584, 587 and 588 contained CAG-EGFP and 594/666 carried CMV-EGFP. Inset boxes in each panel show epifluorescent images of the paw of a transgenic animal (right), compared to a wild-type animal's foot pad (left). All animals except 588 expressed EGFP in their paw. 666 expressed EGFP at a slightly lower level.

EGFP transgene integration in the genome

The integration, transcription and expression of the transgene in the infant marmosets were examined using tissues that could be acquired noninvasively (placenta, hair roots, skin and peripheral blood cells). Because marmosets usually eat the placenta after delivery, only three placentae (584, 588 and that shared by twins 594/666) were collected and available for analysis[11].

The placental DNA from infants 584 and 588 showed high levels of the transgene content by real-time PCR, whereas that from 594/666 showed a relatively low level (Supplementary Data 2). The transgene was detected in the hair roots, skin and peripheral blood from infants 584, 587, 594 and 666.

Copy numbers of the integrated transgene were determined by Southern blotting analysis. At least four copies of the transgene were integrated into the genome of animal 584, and two copies were present in the genome of animal 587 (Fig. 2). Several integration sites in the genomic DNA of skin fibroblast cells, peripheral blood, the placenta of 594 and 666, and the placenta of 588 were found. Infant 588 showed transgene integration only in the placenta (Fig. 2).

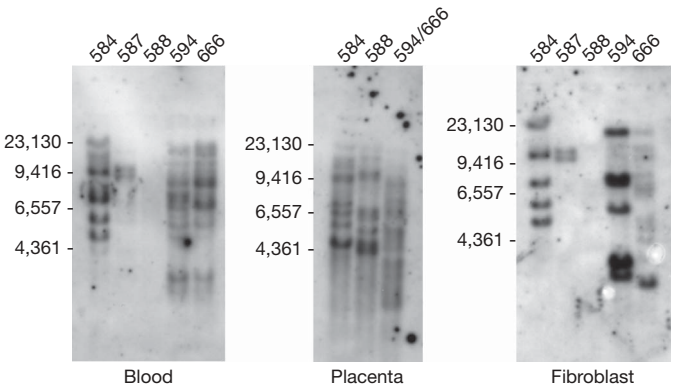


Figure 2 | Transgene insertions in several infant tissues. Southern blot analysis. All infants except 588 showed transgene integration in the skin fibroblast cells and blood, whereas 588 showed transgene integration in the placenta. The lane markers on the left of each gel represent base pairs.

To identify the chromosomal transgene integration sites, fluorescence *in situ* hybridization (FISH) was performed. Consistent with the Southern blotting analysis, the FISH results showed several integration sites in the chromosomes of peripheral blood mononuclear cells (MNCs), and further showed that each infant had different transgene integration patterns with patterns the sometimes varied among different MNCs (Supplementary Fig. 1 and Supplementary Data 3). In 584, four transgene integration sites were seen, on chromosomes 2, 7 and 13; in 587, two distinct signals were recognized in the peripheral blood lymphocyte DNA, on chromosomes 3 and 12. No signal was detected in the peripheral blood lymphocyte samples from 588, and several transgene integration patterns were seen in 594 and 666. Infant 594 had at least three different transgene integration patterns, and more than six patterns may have occurred. Infant 666 showed the largest number of integration patterns, up to 13. Moreover, although this animal was male, of the 13 investigated karyograms, eight samples were of the female karyotype, owing to haematopoietic chimaerism caused by blood exchange with his twin, 594.

Expression of the EGFP transgene

EGFP messenger RNA was detected in the hair roots of all the infants except 588 and in the peripheral blood cells of 584 and 587, by RT–PCR. Transcription of the EGFP gene was indicated in all of the placental samples, 584, 588 and 594/666 (Fig. 3a–c).

To assess EGFP expression in tissues, EGFP fluorescence was examined directly by fluorescence microscopy, and immunohistochemical analysis of the hair roots, frozen sections of a small piece of ear tissue, and placenta samples was performed (Fig. 3d–g). EGFP was strongly expressed in the epidermal cells of the ear tissue and stromal cells of the placenta. In all of the animals except 588, EGFP expression was observed in the hair roots and skin. Placental samples from 584 and 588 also showed high levels of EGFP, but it was undetectable in 594/666 (Supplementary Figs 2–4).

Peripheral blood samples were subjected to flow cytometric analysis using a FACScant. FACS analysis showed EGFP-positive peripheral blood MNCs in 584 and 587. The proportion of EGFP-positive cells was 15.7 and 19.1%, respectively (Fig. 3h). The flow cytometry results corresponded well with those from RT–PCR. Among the peripheral blood cells, the EGFP-positive percentage of granulocytes, lymphocytes and monocytes was 34.5, 3.3 and 18.0% in 584, and 47.7, 4.6 and 20.0% in 587, respectively (Supplementary Fig. 5).

Germline transmission of the transgene

At the moment when two of the animals (666 and 584) became sexually mature, the transgene expression in their gametes was analysed. Semen samples were collected from 666, and live spermatozoa were obtained by the swim-up method in TYH medium. RT–PCR analysis demonstrated the presence and expression of the transgene in the germ cells of 666 (Fig. 4a). IVFs were then performed using semen collected from 666 and wild-type oocytes to analyse the fertility of the germ cells carrying the transgene. Fluorescence microscopy showed that 20–25% of the IVF embryos strongly expressed EGFP, as shown in Fig. 4b. Furthermore, three pre-implantation live natural embryos were collected from female animal 584, and one of these embryos strongly expressed EGFP. The IVF embryos from 666 and two of the natural blastocyst embryos from 584 were shown to express the EGFP transgene by RT–PCR (Fig. 4a). Three EGFP-positive IVF embryos from the male animal (666) were then transferred into a surrogate mother. One neonate (687) was delivered at full term by caesarean section, and this neonate carried the EGFP gene and expressed the transgene in skin (Fig. 4c–e), but not in the placenta and hair.

Discussion

To our knowledge, this is the first report of transgenic non-human primates showing not only the transgene expression in somatic tissues, but also germline transmission of the transgene with the full, normal

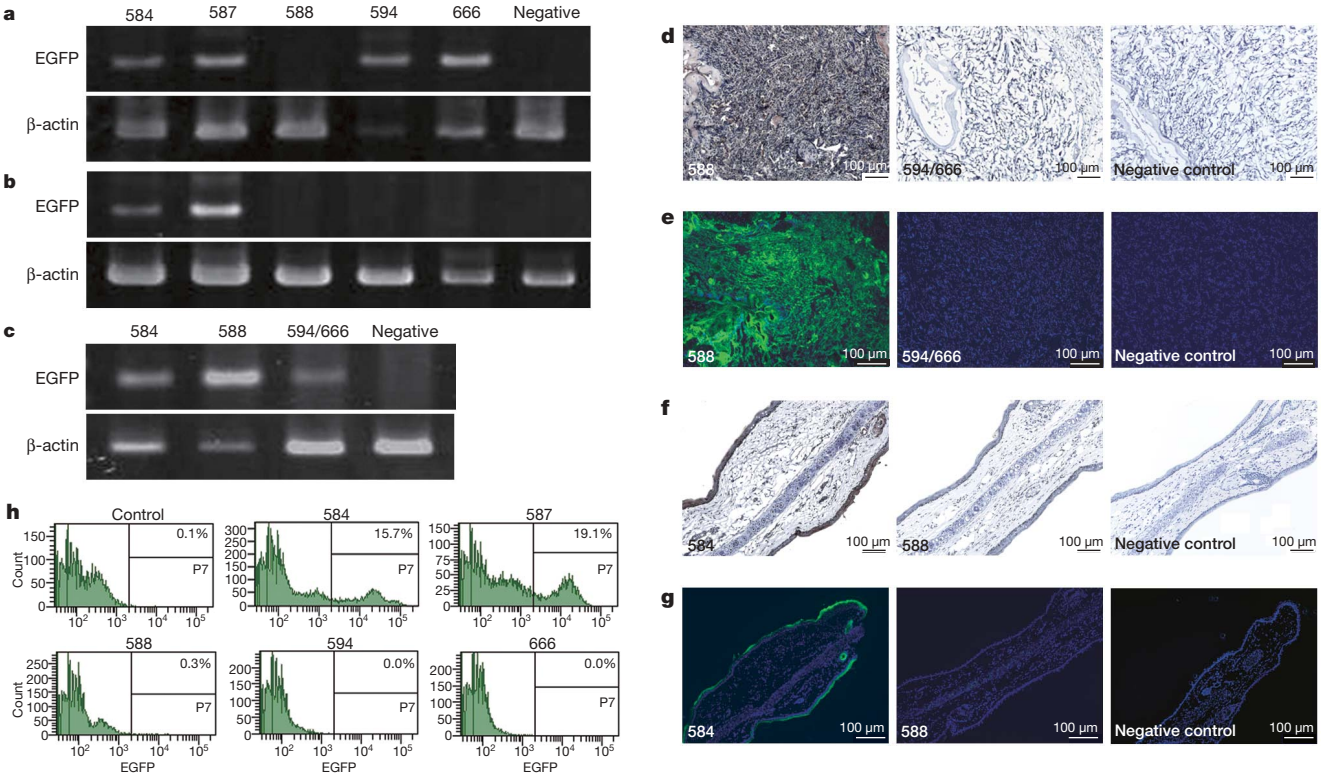


Figure 3 | Transgene transcription and expression in several infant tissues. **a**–**c**, RT–PCR results from hair roots (**a**), peripheral blood (**b**) and placenta (**c**). Each lane indicates the animal number. **d**–**g**, Immunohistochemical (**d**, **f**) and epifluorescence (**e**, **g**) analyses using an anti-EGFP antibody, of frozen ear tissues (**f**, **g**) and placentae (**d**, **e**). Scale bars, 100 μm. **h**, Results of FACS analysis using whole peripheral blood cells. The percentage of EGFP-positive cells is shown in the top right of each panel.

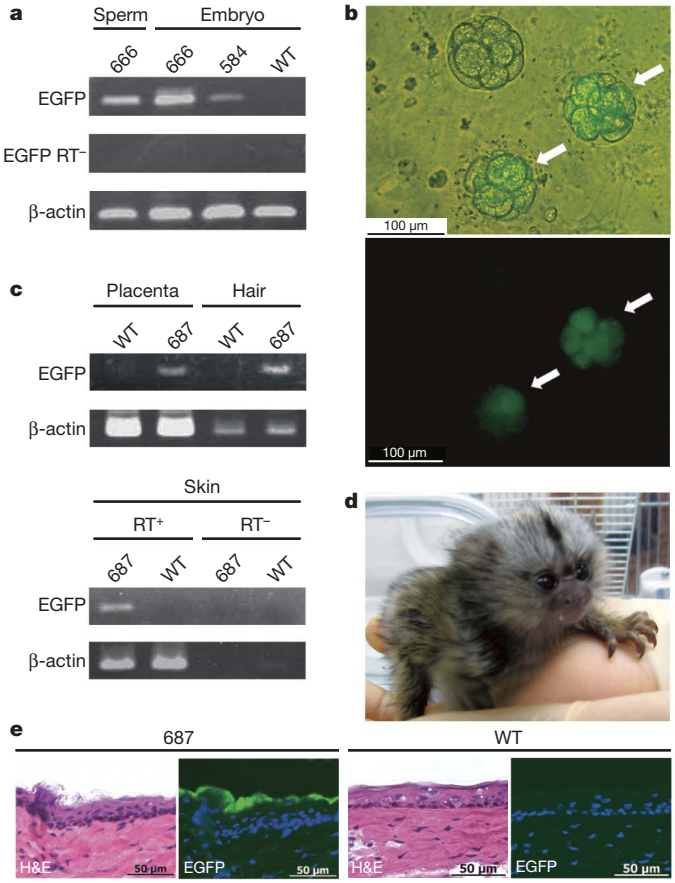


Figure 4 | Germline transmission of the transgene. **a**, RT–PCR analysis of spermatozoa and IVF embryos from 666, and natural embryos from 584. RT− denotes the absence of reverse transcriptase as a control. **b**, Bright-field and dark-field of epifluorescence images of IVF embryos. EGFP-positive IVF embryos produced with 666 spermatozoa are indicated by white arrows. **c**, PCR (top panel) and RT–PCR (bottom panel) analysis of the tissues from the F_1 neonate. **d**, Photograph of the F_1 offspring (687) from 666. **e**, Haematoxylin and eosin (H&E) staining and epifluorescence imaging of frozen skin tissue from the neonate. WT, wild-type control.

development of the embryo. We obtained five transgenic marmosets, four of which expressed the transgene in several somatic cell lineages, such as hair root, skin fibroblast and peripheral blood cells. The remaining animal expressed the transgene only in the placenta. Two of these animals reached sexual maturity and showed the transgene insertion and expression in germ cells. Epifluorescence microscopic observation and RT–PCR analysis of embryos generated by transgene-bearing gametes strongly indicated that the transgenic germ cells from animals 666 and 584 were fertile, and this was proved for the male (666) who fathered one healthy, transgenic infant (687) with the transgene expression in the somatic cells. These findings suggest that it should be possible to establish transgenic non-human primate colonies, opening the door to their use in biomedical research.

Because the manipulation of embryos for viral injection and their subsequent culture may affect embryonic development, the birth rate after embryo transfer (6.25%) was lower than that for normal embryos (30.7%, data not shown). The miscarriage rates were not significantly different between embryonic transfers performed using normal embryos (28.6%) and transgenic embryos (42.6%). Despite considerable effort, transgenic marmosets have not been produced by DNA pronuclear microinjection. The production rate that we obtained using lentivirus (5.26–6.25%) suggests that our technique is sufficiently effective for the production and use of genetically modified marmosets as human disease models.

The 100% birth rate of transgenic marmosets achieved in the present study could be due to several technical advantages. First, we used EGFP as the transgene, enabling us to monitor the presence and expression of the transgene at each experimental step in live embryos and the transgenic animals. Accordingly, we were able to select unambiguously EGFP-expressing embryos for transfer into surrogate mothers. This selection was effective, not only for increasing the birth rate of transgenic animals, but also for reducing the number of surrogate mother animals needed.

Second, we used pre-implantation embryos obtained by natural intercourse, high-titre lentiviral vectors, and 0.25 M sucrose solution as a medium for injection. Even though the birth rates (birth per embryonic transfer) were no different between IVF and natural embryos, the fertilization rate of the germinal vesicle-stage oocytes was quite low. Because it is difficult to collect large quantities of oocytes, it was advantageous to use marmoset natural embryos. To inject as much lentiviral vector as possible into the perivitelline space, the embryos were placed in 0.25 M sucrose medium at the time of lentiviral vector injection, which expanded the volume of the perivitelline space 1.2–7.5-fold. For example, the estimated volume of the perivitelline space of one marmoset pronuclear stage embryo was approximately 31.5 pl, but when placed in 0.25 M sucrose medium, it expanded to about 231 pl. A high titre of the lentiviral vector solution was used so that many lentiviral vector particles were injected into the expanded perivitelline space; approximately 1.3×10^3–1.3×10^5 transducing units of lentiviral vector were injected in this study. Each of these steps probably contributed to the successful production of transgenic marmosets.

The high number of injected lentiviral vector particles resulted in several transgene integrations, as observed by Southern blot analysis and FISH. The embryos injected with the transgene before the four-cell stage (584 and 587) showed fewer than four copies of the transgene per genome by Southern blotting and FISH. The three other embryos (588, 594 and 666), which received the injection at the morula stage, exhibited several integrations of the transgene by Southern blotting and FISH. As the FISH analysis was performed using only peripheral blood MNCs, other patterns of transgene integration cells may have existed in other tissues. The FISH results for 666 were consistent with this hypothesis, as the integration sites in the chimaeric blood MNCs from his twin, 594, were different from those in the blood MNCs of 594.

The lentiviral vector used in the present study can be used to transmit only relatively small transgenes, 8.5 kilobases of DNA or less. Therefore, further study will be necessary to enable the introduction of larger transgenes into marmoset embryos. Furthermore, to study human diseases involving the malfunctions of specific genes, targeted gene-knockdown marmosets could be developed using RNA interference (RNAi) lentiviruses.

The results of the present study indicate that transgenic marmosets may be used as experimental animals for biomedical research. Recently, somatic cell nuclear-transferred embryonic stem cells from the rhesus macaque and induced pluripotent stem cells from adult human fibroblasts were reportedly established[12–15]. Those studies indicated that the obstacle caused by immunogenetic incompatibility has at least theoretically been resolved, and that a new era of regenerative medicine using somatic cell nuclear-transferred embryonic stem cells in primates[14] or human induced pluripotent stem cells[12,13,15] has become possible. However, before such stem cells can be used in clinical applications, preclinical assessments of their safety and efficacy are essential. We previously reported that marmosets with injured spinal cords can recover motor function after the transplantation of human neural stem/progenitor cells[16], highlighting the usefulness of the marmoset for assessing the safety and efficacy of, not only these cells, but also of other stem cells, such as human embryonic stem cells[17] or induced pluripotent stem cells. Human disease models in non-human primates have so far been limited to mechanical injury models (for example, spinal cord injury[18]) and drug administration models (for example, MPTP-induced Parkinson's[19,20]). The only transgene-induced primate disease model is of Huntington's disease[7], in rhesus

monkeys expressing a mutant human huntingtin gene. In that report, although the transgene was inserted into the genome of founder infants and its expression was detected in post-mortem animals, the germline transmission of the transgene has not yet been confirmed[7]. Thus, at this point, it is not certain how reproducible the effects of various therapeutic interventions would be using a large number of animals.

The technique by which we achieved transgene expression in several tissues, along with germline transmission, may provide the means to obtain genetically modified non-human primate models for translational research, investigations of regenerative medicine and gene therapy, and clarification of the scientific gaps among transgenic mice, human disease models, and real human diseases.

METHODS SUMMARY

All animal experiments were approved by the institutional animal care and use committee, and were performed in accordance with Central Institution for Experimental Animal (CIEA) guidelines.

To obtain oocytes, recombinant human follicle stimulating hormone (r-hFSH; 50 international units (IU); Fertinome, Serono) was administered daily by intramuscular injection for 11 days. Human chorionic gonadotropin (hCG; 75 IU; Gonatropin, Teikoku-zouki) was administered by intramuscular injection at 17:30 on day 12. On day 13, the animals were anaesthetized and follicular aspiration was performed surgically. Oocytes were incubated for 24 h at 38 °C, 5% CO_2 in air, for *in vitro* maturation. After incubation, only matured oocytes (metaphase II) were collected and used for IVF.

Ejaculated semen was collected in TYH medium (Mitsubishi Kagaku Iatron), using a Ferti Care personal vibrator. Hyaluronidase-treated oocytes were placed in 70-μl drops of TYH, and an aliquot of sperm (4×10^5) was added to each oocyte incubation drop. After 26–30 h of insemination, the fertilized oocytes were placed into ISM1 (Medicult) medium, and lentiviral vector injection was performed in 0.25 M sucrose.

Natural embryo collection was performed as previously described[21]. Embryos at the pronuclear-to-morula stage were placed in 0.25 M sucrose supplemented PB1 medium (Mitsubishi Chemical Medience Corporation) and injected with lentiviral vector. Blastocysts were not treated with sucrose. Lentiviral vector injection was performed using an Eppendorf FemtoJet express and a Narishige micromanipulator. The embryos were cultured until GFP expression was confirmed.

The ovulation cycles of donor and recipient animals were synchronized, and EGFP-expressing embryos were transferred as previously described[22,23]. After embryo transfer, the recipients were tested for pregnancy by plasma progesterone once a week. The resulting infants were analysed for transgene integration, transcription and expression, by real-time PCR, Southern blot analysis, RT–PCR, immunohistochemical analysis, FACS and FISH.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 September 2008; accepted 30 April 2009.

1. Mansfield, K. Marmoset models commonly used in biomedical research. *Comp. Med.* **53,** 383–392 (2003).
2. Chan, A. W., Homan, E. J., Ballou, L. U., Burns, J. C. & Bremel, R. D. Transgenic cattle produced by reverse-transcribed gene transfer in oocytes. *Proc. Natl Acad. Sci. USA* **95,** 14028–14033 (1998).
3. Hofmann, A. *et al.* Efficient transgenesis in farm animals by lentiviral vectors. *EMBO Rep.* **4,** 1054–1060 (2003).
4. Hofmann, A. *et al.* Generation of transgenic cattle by lentiviral gene transfer into oocytes. *Biol. Reprod.* **71,** 405–409 (2004).
5. Chan, A. W., Chong, K. Y., Martinovich, C., Simerly, C. & Schatten, G. Transgenic monkeys produced by retroviral gene transfer into mature oocytes. *Science* **291,** 309–312 (2001).
6. Wolfgang, M. J. *et al.* Rhesus monkey placental transgene expression after lentiviral gene transfer into preimplantation embryos. *Proc. Natl Acad. Sci. USA* **98,** 10728–10732 (2001).
7. Yang, S. H. *et al.* Towards a transgenic model of Huntington's disease in a non-human primate. *Nature* **453,** 921–924 (2008).
8. Rogers, C. S. *et al.* Disruption of the *CFTR* gene produces a model of cystic fibrosis in newborn pigs. *Science* **321,** 1837–1841 (2008).
9. Rogers, C. S. *et al.* Production of *CFTR*-null and *CFTR-ΔF508* heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. *J. Clin. Invest.* **118,** 1571–1577 (2008).
10. Miyoshi, H., Blomer, U., Takahashi, M., Gage, F. H. & Verma, I. M. Development of a self-inactivating lentivirus vector. *J. Virol.* **72,** 8150–8157 (1998).
11. Ross, C. N., French, J. A. & Orti, G. Germ-line chimerism and paternal care in marmosets (*Callithrix kuhlii*). *Proc. Natl Acad. Sci. USA* **104,** 6278–6282 (2007).
12. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
13. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
14. Byrne, J. A. *et al.* Producing primate embryonic stem cells by somatic cell nuclear transfer. *Nature* **450,** 497–502 (2007).
15. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnol.* **26,** 101–106 (2008).
16. Iwanami, A. *et al.* Transplantation of human neural stem cells for spinal cord injury in primates. *J. Neurosci. Res.* **80,** 182–190 (2005).
17. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282,** 1145–1147 (1998).
18. Iwanami, A. *et al.* Establishment of graded spinal cord injury model in a nonhuman primate: the common marmoset. *J. Neurosci. Res.* **80,** 172–181 (2005).
19. Eslamboli, A. Marmoset monkey models of Parkinson's disease: which model, when and why? *Brain Res. Bull.* **68,** 140–149 (2005).
20. Ando, K. *et al.* Neurobehavioral protection by single dose l-deprenyl against MPTP-induced parkinsonism in common marmosets. *Psychopharmacology (Berl.)* **195,** 509–516 (2008).
21. Sasaki, E. *et al.* Establishment of novel embryonic stem cell lines derived from the common marmoset (*Callithrix jacchus*). *Stem Cells* **23,** 1304–1313 (2005).
22. Lopata, A., Summers, P. M. & Hearn, J. P. Births following the transfer of cultured embryos obtained by *in vitro* and *in vivo* fertilization in the marmoset monkey (*Callithrix jacchus*). *Fertil. Steril.* **50,** 503–509 (1988).
23. Summers, P. M., Shephard, A. M., Taylor, C. T. & Hearn, J. P. The effects of cryopreservation and transfer on embryonic development in the common marmoset monkey, *Callithrix jacchus*. *J. Reprod. Fertil.* **79,** 241–250 (1987).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Toyoda, S. Ohba, T. Inoue, Y. Sawada and M. Yokoyama for technical assistance with the animal experiments and care. E.S. is an associate professor of the Global COE program for human metabolomic systems biology assigned to Keio University. This study was also supported by the Global COE program for Education and Research Centre for Stem Cell Medicine from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japanese Government to Keio University. This study was also supported by funds from Solution-Oriented Research for Science and Technology (SORST) of the Japan Science and Technology Agency and grants from MEXT to H.O. and from Special Coordination Funds for Promoting Science and Technology of MEXT to S.H.

Author Contributions E.S. designed the experiments, conducted the project, and wrote the paper. A.S., Y.S., T.E., I.T. and R.H. assisted in embryological technique development. K.H., R.O. and M.K. developed surgical techniques for embryo collection and transfer. H.S., C.K. and C.Y. performed or assisted with the real-time PCR and parentage evaluation test. S.S. and T.M. assisted with the Southern blot analysis and tissue collection. M.I. raised the anti-marmoset CD45 antibody. R.I. performed the FACS analysis, and K.K. performed the immunohistochemical analysis. H.M. provided the lentiviral vectors. Y.T., H.O., S.H., N.T. and T.N. designed the project, and H.O., S.H. and N.T. also participated in writing the paper. The whole project was supervised by E.S. and H.O.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.S. (esasaki@ciea.or.jp) or H.O. (hidokano@sc.itc.keio.ac.jp).

METHODS

Animals. Adult common marmosets more than 2 years old were obtained from a marmoset breeding colony for experimental animals. Female marmosets with normal ovarian cycles were paired with intact males for natural embryo collection. Recipient females were paired with vasectomised males or intact females. This study was approved by the Institutional Animal Care and Use Committee of CIEA, and was performed in accordance with CIEA guidelines.

***In vitro* fertilization.** Semen was collected as previously described for common marmosets[24]. Ejaculated semen was collected in TYH medium (Mitsubishi Kagaku Iatron) and washed twice with TYH. The semen was placed in a CO_2 incubator for 10 min in a test tube inclined at a 30° angle to allow the sperm to swim up. Hyaluronidase-treated metaphase-II-arrested oocytes were inseminated with a final concentration of 5×10^6 sperm ml^{-1} for 26–30 h. Fertilized embryos were cultured in ISM medium (Medicult, Nosan Corp.).

Embryo collection and transfer. Embryo collection and transfer were performed as previously described[25]. After embryo transfer, the recipients were monitored for pregnancy by measuring their plasma progesterone once a week until the pregnancies could be monitored by transabdominal palpation of the uterus.

Lentiviral vector preparation and transduction. The lentiviral vectors were produced as previously described[26]. The medium containing viral particles was spun at 4 °C, 50,000*g* for 4 h, and the viral pellet was then resuspended in ISM2 medium, in 1/1,000 of the volume of the original lentiviral vector supernatant. To measure the lentivirus titre, serially diluted (10^{-2} to 10^{-8}) lentiviral vector was used to infect 10^5 293T cells. The number of EGFP-positive cells was counted by FACS to quantify the titre.

Pronuclear-to-morula stage embryos were placed in 0.25 M sucrose supplemented PB1 medium (Mitsubishi Chemical Medience Corporation), and the virus was injected into the perivitelline space. For blastocyst embryos, the viral vector was injected into the blastocoel. All viral injections were performed using an Eppendorf FemtoJet Express and a Narishige micromanipulator.

Southern blot analysis. Five micrograms of genomic DNA was digested with BamHI for animals that had been injected with CAG-EGFP, and with EcoRI for those that had been injected with CMV-EGFP. The digested genomic DNA was separated on a 0.8% agarose gel and transferred to a Hybond-N+ nylon membrane (GE Healthcare Biosciences). Southern blot analysis was performed using the DIG system (Roche Diagnostics K.K.), according to the manufacturer's protocol. CMV-EGFP was digested with EcoRI and then labelled with DIG using the PCR DIG probe synthesis kit, according to the manufacturer's instructions (Roche Diagnostics K.K.).

RT–PCR. To detect EGFP gene expression, EGFP5-5 (5′-GCACAAGCTGGAGTACAACTACAACAGC-3′) and EGFP3-1 (5′-TCACGAACTCCAGCAGGACCAT-3′) primers were used. To detect β-actin expression, β-actin 001 (5′-TCCTGACCCTGAAGTACCCC-3′) and β-actin 002 (5′-GTGGTGGTGAAGCTGTAGCC-3′) primers were used. PCR was performed for 35 cycles of denaturation at 94 °C for 30 s, annealing for 30 s at 58 °C for EGFP primers or 62 °C for β-actin primers, and elongation at 72 °C for 30 s.

To detect EGFP gene expression in germ cells and neonatal tissues, PCR was performed using the EGFP5-4 (5′-CAAGGACGACGGCAACTACAAGACC-3′) and EGFP3-3es (5′- GCTCGTCCATGCCGAGAGTGA-3′) primers. Then, 1 μl of the PCR products was re-amplified with the EGFP5-6 (5′-TCGAGCTGAAGGGCATCGAC-3′) and EGFP3-1 (5′-TCACGAACTCCAGCAGGACCAT-3′) primers. To detect β-actin expression, the PCR primers β-actin 003 (5′-TGGACTTCGAGCAGGAGAT-3′) and β-actin 006R (5′-CCTGCTTGCTGATCCACATG-3′) were used. Then, 0.5 μl of the PCR products was re-amplified with the 004 (5′-TCCCTGGAGAAGAGCTATG-3′) and 005R (5′-GAGCCACCAATCCACACTGA-3′) primers. PCR was performed for 30 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 10 s, and elongation at 72 °C for 30 s.

Immunohistochemical analysis. Tissues were embedded in OCT compound, frozen in liquid nitrogen, and sliced into 5-μm sections, which were fixed in 4% paraformaldehyde for 30 min at 4 °C. Endogenous peroxidase activity was quenched using 0.03% hydrogen peroxidase for 30 min at room temperature. The slides were blocked with 10% goat serum (Nichirei) for 10 min at room temperature and then reacted with the rabbit anti-GFP polyclonal antibody (A.v. peptide antibody, BD Bioscience) overnight at 4 °C. The slides were incubated with the biotinylated secondary antibody Simple Stain Mouse MAX PO (Nichirei) for 30 min at room temperature. The bound antibodies were detected with DAB (3,3-diaminobenzidine tetrahydrochloride) horseradish peroxidase complex. The samples were then stained with H&E and examined by microscopy.

FACS analysis. Whole blood cells were washed with PBS and suspended in 0.13 M NH_4Cl. The pellet was incubated with the mouse IgG1 anti-marmoset CD45, 6C9 antibody for 30 min on ice[27], then mixed with an allophycocyanin (APC)-labelled anti-mouse IgG antibody, and incubated for 30 min on ice. The sample was washed with PBS and resuspended in 200 μl of propidium iodide solution. FACScant analysis was then performed.

Fluorescent *in situ* hybridization. Peripheral blood samples were cultured in RPMI 1640 containing phytohaemagglutinin, concanavalin A, lipopolysaccharide, and 2-mercaptoethanol for 2–3 days. After 2–3 h of incubation with BrdU (final concentration 30 $\mu g\,ml^{-1}$), colcemid (final concentration 0.02 $\mu g\,ml^{-1}$) was added to the medium, and the samples were incubated for another 2 h. After lymphocyte fixation, the cells were spread on slides and air-dried overnight, then stained with Hoechst 33258 and treated with ultraviolet light. CAG-EGFP was labelled with digoxigenin-11-dUTP as a probe, and hybridized at 37 °C overnight. After stringent washes, the bound label was detected using anti-Dig-Cy3. For karyotyping, Leica CW4000 FISH and Leica CW4000 Karyo were used.

24. Kuederling, I., Schneiders, A., Sonksen, J., Nayudu, P. L. & Hodges, J. K. Non-invasive collection of ejaculates from the common marmoset (*Callithrix jacchus*) using penile vibrostimulation. *Am. J. Primatol.* **52,** 149–154 (2000).
25. Summers, P. M., Shephard, A. M., Taylor, C. T. & Hearn, J. P. The effects of cryopreservation and transfer on embryonic development in the common marmoset monkey, *Callithrix jacchus*. *J. Reprod. Fertil.* **79,** 241–250 (1987).
26. Bai, Y. *et al.* Effective transduction and stable transgene expression in human blood cells by a third-generation lentiviral vector. *Gene Ther.* **10,** 1446–1457 (2003).
27. Ito, R. *et al.* Novel monoclonal antibodies recognizing different subsets of lymphocytes from the common marmoset (*Callithrix jacchus*). *Immunol. Lett.* **121,** 116–122 (2008).

Vol 459|28 May 2009|doi:10.1038/nature07999

ARTICLES

Common genetic variants on 5p14.1 associate with autism spectrum disorders

Kai Wang[1]*, Haitao Zhang[1]*, Deqiong Ma[2]*, Maja Bucan[3], Joseph T. Glessner[1], Brett S. Abrahams[4], Daria Salyakina[2], Marcin Imielinski[1], Jonathan P. Bradfield[1], Patrick M. A. Sleiman[1], Cecilia E. Kim[1], Cuiping Hou[1], Edward Frackelton[1], Rosetta Chiavacci[1], Nagahide Takahashi[5], Takeshi Sakurai[5], Eric Rappaport[6], Clara M. Lajonchere[7], Jeffrey Munson[8], Annette Estes[8], Olena Korvatska[8], Joseph Piven[9], Lisa I. Sonnenblick[4], Ana I. Alvarez Retuerto[4], Edward I. Herman[4], Hongmei Dong[4], Ted Hutman[4], Marian Sigman[4], Sally Ozonoff[10], Ami Klin[11], Thomas Owley[12], John A. Sweeney[12], Camille W. Brune[12], Rita M. Cantor[13], Raphael Bernier[8], John R. Gilbert[2], Michael L. Cuccaro[2], William M. McMahon[14], Judith Miller[14], Matthew W. State[11], Thomas H. Wassink[15], Hilary Coon[14], Susan E. Levy[6], Robert T. Schultz[6], John I. Nurnberger Jr[16], Jonathan L. Haines[17], James S. Sutcliffe[18], Edwin H. Cook[12], Nancy J. Minshew[19], Joseph D. Buxbaum[5,20], Geraldine Dawson[8], Struan F. A. Grant[1,6], Daniel H. Geschwind[4], Margaret A. Pericak-Vance[2], Gerard D. Schellenberg[21] & Hakon Hakonarson[1,6]

Autism spectrum disorders (ASDs) represent a group of childhood neurodevelopmental and neuropsychiatric disorders characterized by deficits in verbal communication, impairment of social interaction, and restricted and repetitive patterns of interests and behaviour. To identify common genetic risk factors underlying ASDs, here we present the results of genome-wide association studies on a cohort of 780 families (3,101 subjects) with affected children, and a second cohort of 1,204 affected subjects and 6,491 control subjects, all of whom were of European ancestry. Six single nucleotide polymorphisms between cadherin 10 (*CDH10*) and cadherin 9 (*CDH9*)—two genes encoding neuronal cell-adhesion molecules—revealed strong association signals, with the most significant SNP being rs4307059 ($P = 3.4 \times 10^{-8}$, odds ratio = 1.19). These signals were replicated in two independent cohorts, with combined *P* values ranging from 7.4×10^{-8} to 2.1×10^{-10}. Our results implicate neuronal cell-adhesion molecules in the pathogenesis of ASDs, and represent, to our knowledge, the first demonstration of genome-wide significant association of common variants with susceptibility to ASDs.

ASDs encompass a range of clinically defined conditions, including autism and pervasive developmental disorder not otherwise specified, which are more common and severe, as well as Asperger's syndrome, which appears less frequently and is milder[1]. ASDs are about four times more common in boys than girls, and at present around 1 in 150 children in the United States have a diagnosis of an ASD[2]. Several sources of evidence suggest that strong genetic components are involved in susceptibility to ASDs: there are much higher concordance rates of ASDs in monozygotic twins (92%) than dizygotic twins (10%)[3], and recent estimate of the sibling recurrence risk ratio (λ_s) is 22 for autism[4]. Despite being highly heritable, ASDs show heterogeneous clinical symptoms and genetic architecture, which have hindered the identification of common genetic susceptibility factors[5]. Although previous linkage studies, candidate gene association studies and cytogenetic studies have implicated several chromosomal regions for the presence of autism susceptibility loci[6–9], they have failed to consistently identify and replicate common genetic variants that increase the risk of ASDs.

Besides well-known genetic conditions reported in ASDs, recent studies have identified a growing number of distinct and individually rare genetic causes, suggesting that the genetic architecture of ASDs may have a significant contribution from heterogeneous rare variants. For example, rare *de novo* copy number variants have been implicated in 7% of families with ASDs, but only in 1% of control families[10]. In addition, 16p11.2 microdeletions and microduplications have been found in approximately 1% of autism cases[11,12]. Several hundred rare structural variations have also been catalogued in families with ASDs[13]. Although these reported variants indicate a role for rare genomic

[1]Center for Applied Genomics, Children's Hospital of Philadelphia, Pennsylvania 19104, USA. [2]The Miami Institute for Human Genomics and Department of Human Genetics, University of Miami, Miami, Florida 33101, USA. [3]Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19014, USA. [4]Department of Neurology and Center for Autism Research and Treatment, University of California Los Angeles, Los Angeles, California 90095, USA. [5]Seaver Autism Center for Research and Treatment, Department of Psychiatry, Mount Sinai School of Medicine, New York, New York 10029, USA. [6]Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19014, USA. [7]Department of Biomedical Engineering, University of Southern California and Autism Speaks, Los Angeles, California 90089, USA. [8]Department of Psychiatry, University of Washington, Seattle, Washington 98105, USA. [9]Department of Psychiatry, University of North Carolina and Carolina Institute for Developmental Disabilities, Chapel Hill, North Carolina 27599, USA. [10]The M.I.N.D. Institute and Department of Psychiatry and Behavioral Sciences, University of California Davis, Sacramento, California 95817, USA. [11]Department of Genetics and Child Study Center, Yale University, New Haven, Connecticut 06520, USA. [12]Institute for Juvenile Research and Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60608, USA. [13]Department of Human Genetics, University of California Los Angeles, Los Angeles, California 90095, USA. [14]Department of Psychiatry, University of Utah, Salt Lake City, Utah 84132, USA. [15]Department of Psychiatry, University of Iowa, Iowa City, Iowa 84112, USA. [16]Institute of Psychiatric Research, Indiana University, Indianapolis, Indiana 46202, USA. [17]Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee 37232, USA. [18]Center for Molecular Neuroscience and Vanderbilt Kennedy Center, Vanderbilt University, Nashville, Tennessee 37232, USA. [19]Departments of Psychiatry and Neurology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [20]Departments of Psychiatry, Neuroscience, and Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA. [21]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19014, USA.
*These authors contributed equally to this work.

variation in a proportion of families, no common variants have been previously associated with ASDs in genome-wide studies. The latter is consistent with reports from previous genome-wide association studies of other neuropsychiatric disorders, including bipolar disorder[14,15], schizophrenia[16] and attention deficit/hyperactivity disorder[16], all of which have failed to identify common susceptibility loci with genome-wide significance, when individual data sets with small sample sizes were analysed. However, recent meta-analysis reported evidence for common variants in both schizophrenia[17] and bipolar disorder[18], suggesting that the search for common genetic variation that confer susceptibility to ASDs may benefit from the combined analysis of several studies.

Genome-wide association studies on ASDs

To identify common genetic risk factors for ASDs, we carried out a genome-wide association study on 943 ASDs families (4,444 subjects) from the Autism Genetic Resource Exchange (AGRE cohort, Table 1)[19]. The subjects with ASDs in the AGRE cohort were diagnosed using both the Autism Diagnostic Interview-Revised (ADI-R)[20] and Autism Diagnostic Observation Schedule (ADOS)[21] diagnostic tools, which are the gold standard diagnostic tools for individuals with ASDs. All subjects were genotyped using the Illumina HumanHap550 BeadChip with over 550,000 single nucleotide polymorphism (SNP) markers. We applied stringent quality control criteria (Supplementary Methods), including call rates, Mendelian inconsistencies and genetically inferred ancestry, to identify a set of 3,101 subjects of European ancestry in 780 AGRE families for association tests. We performed analysis with the Pedigree Disequilibrium Test (PDT)[22] for autosomes, and with X-APL[23] for the X chromosome, using genotypes from 486,864 markers. The complete sets of SNP genotype data and signal intensity data were released to the academic research community in April 2008 (http://www.agre.org).

We did not observe genome-wide significant association ($P < 5 \times 10^{-8}$) to ASDs in the AGRE cohort, but we proposed that meaningful associations were contained within the lowest *P* values. To boost power for identifying these associations, we examined an Autism Case-Control cohort (ACC cohort, Table 1), comprising 1,453 subjects with ASDs from several US sites, and 7,070 control subjects without ASDs from the Children's Hospital of Philadelphia, who were also genotyped on the same platform. The subjects with ASDs in this cohort were diagnosed using the ADI and ADOS tools. After conducting thorough quality control measures on the genotypes, association analyses were conducted on 1,204 subjects with ASDs and 6,491 control subjects of inferred European ancestry. We did not detect genome-wide significant association ($P < 5 \times 10^{-8}$) to ASDs in the ACC cohort either. Therefore, we subsequently performed a combined analysis of these two independent data sets using recommended meta-analysis approaches[24]. From examining autosomes and the X chromosome, one SNP located on 5p14.1 reached genome-wide significance (rs4307059, $P = 3.4 \times 10^{-8}$), and five further SNPs at the same locus had *P* values below 1×10^{-4} (Table 2 and Fig. 1a). Several other loci contain SNPs with suggestive association signals (Table 3), such as 13q33.3 (near *MYO16* (myosin XVI)), 14q21.1 (between *FBXO33* (F-box protein 33) and *LRFN5* (leucine rich repeat and fibronectin type III domain containing 5)) and Xp22.32 (between *PRKX* (protein kinase, X-linked) and *NLGN4X* (neuroligin 4, X-linked)). We also analysed ten markers on the Y chromosome in the ACC cohort, with the most significant SNP being rs2032597 ($P = 1.1 \times 10^{-4}$) located within *USP9Y* (ubiquitin specific protease 9, Y-linked) (Supplementary Table 1). Furthermore, we have analysed 15 markers in pseudoautosomal regions of sex chromosomes in the two discovery cohorts, but no markers showed evidence of association (Supplementary Table 2).

To identify other variants that associate with ASDs but were not captured by the SNP genotyping array, we analysed the discovery cohorts using whole-genome imputed genotypes on autosomes generated by the MACH software (Supplementary Methods). The most significant association signals were still those in the 5p14.1 region. However, several other genomic loci, such as 10q21.3 (within *CTNNA3* (catenin, alpha 3)) and 16p13.2 (between *A2BP1* (ataxin 2-binding protein 1) and *C16orf68* (chromosome 16 open reading frame 68)), contain imputed SNPs with suggestive association signals (Table 3). Follow-up studies with larger sample sizes are required to determine whether these represent genuine ASD susceptibility loci.

Replication of the association signals

To replicate our genome-wide significant association signals at the 5p14.1 locus, we examined the association statistics for these markers in a third independently generated and analysed cohort, including 1,390 subjects from 447 autism families genotyped with ~1 million markers on the Illumina HumanHap1M BeadChip (CAP cohort, Table 1). The association signals for all the aforementioned SNPs were replicated in this cohort with the same direction of association, with *P* values ranging from 0.01 to 2.8×10^{-5} (Table 2). To seek further evidence of replication, we examined association statistics from a fourth independent cohort of 108 ASD cases and 540 genetically matched control subjects, genotyped on the Illumina HumanCNV370 BeadChip, an array with over 300,000 SNP markers (CART cohort, Table 1). Because rs7704909 and rs10038113 were not present in this array platform, we analysed association on imputed genotypes. Most of the SNPs were replicated ($P < 0.05$) in the CART cohort with the same direction of association (Table 2). Combined analysis on all four data sets indicates that all six SNPs are associated with ASDs, with *P* values ranging from 7.4×10^{-8} to 2.1×10^{-10} (Table 2 and Supplementary Table 3). Taken together, several sources of converging evidence firmly established that common genetic variants on 5p14.1 confer susceptibility to ASDs.

Genomic features of the 5p14.1 region

Closer examination of the 5p14.1 region indicated that all genotyped and imputed SNPs with *P* values below 1×10^{-7} reside within the same ~100 kilobase (kb) linkage disequilibrium block, suggesting that these SNPs are tagging the same variants (Supplementary Figs 1 and 2). The linkage disequilibrium block is located within a 2.2-megabase (Mb) intergenic region between *CDH10* (cadherin 10) and *CDH9* (cadherin 9) (Fig. 1b, c). Both *CDH10* and *CDH9* encode type II classical cadherins from the cadherin superfamily, which represent transmembrane proteins that mediate calcium-dependent cell–cell adhesion. To search for other types of variants, including copy number variations (CNVs), in the intergenic region, we used the PennCNV software[25] on the signal intensity data and identified five CNV loci (Supplementary Fig. 3). All of these CNVs are present in control subjects in our study, and three of the five CNVs are also reported in the Database for Genomic Variants that annotates healthy

Table 1 | Description of the four data sets used in the study

Data set	Purpose	Study design	Measurements	Subjects				Analysis	
				Cases	Controls	Families	Total	Genotyped markers	Imputed markers
AGRE	Discovery	Family-based	550K Illumina	1,299		780	3,101	PDT, X-APL	
ACC	Discovery	Case-control	550K Illumina	1,204	6,491		7,695	PLINK	SNPTEST
CAP*	Replication	Family-based	1M Illumina	504		447	1,390	PDT	
CART	Replication	Case-control	300K Illumina	108	540		648	PLINK	SNPTEST

* The CAP data set has been fully described in ref. 51.

Table 2 | The most significantly associated SNPs ($P < 1 \times 10^{-4}$ in the discovery phase) between *CDH10* and *CDH9* on 5p14.1

SNP	Position*	Minor/major allele*	Discovery cohorts								Replication cohorts		*P* value (combined)
			MAF† (AGRE)	*P* value (AGRE)	*Z* score‡ (AGRE)	Case MAF (ACC)	Control MAF (ACC)	*P* value (ACC)	Odds ratio§ (ACC)	*P* value (discovery cohorts)	*P* value (CAP)	*P* value (CART)	
rs4307059	26003460	C/T	0.38	1.1×10^{-5}	4.40	0.35	0.39	2.2×10^{-4}	1.19	3.4×10^{-8}	1.2×10^{-2}	1.6×10^{-2}	2.1×10^{-10}
rs7704909	25934678	C/T	0.39	1.6×10^{-5}	4.31	0.36	0.40	6.2×10^{-4}	1.17	1.4×10^{-7}	9.1×10^{-3}	4.0×10^{-2}	9.9×10^{-10}
rs12518194	25987318	G/A	0.39	1.3×10^{-5}	4.36	0.36	0.39	1.0×10^{-3}	1.16	2.0×10^{-7}	9.3×10^{-3}	1.8×10^{-2}	1.1×10^{-9}
rs4327572	26008578	T/C	0.39	2.2×10^{-5}	4.24	0.36	0.39	2.0×10^{-3}	1.15	6.2×10^{-7}	7.3×10^{-3}	1.5×10^{-2}	2.7×10^{-9}
rs1896731	25934777	C/T	0.34	1.7×10^{-3}	−3.14	0.38	0.34	1.7×10^{-3}	0.87	1.7×10^{-5}	7.7×10^{-5}	9.9×10^{-1}	4.8×10^{-8}
rs10038113	25938099	C/T	0.40	1.4×10^{-3}	−3.19	0.43	0.39	2.4×10^{-3}	0.87	2.1×10^{-5}	2.8×10^{-5}	4.5×10^{-1}	7.4×10^{-8}

* The chromosome coordinates and allele designation are on the basis of the forward strand of the NCBI 36 genome assembly.
† The minor allele frequencies (MAF) are calculated on the basis of AGRE parents of European ancestry.
‡ Positive *Z* score indicates overtransmission of the major allele.
§ The odds ratio is calculated with respect to the major allele.

individuals (Supplementary Fig. 4), suggesting that CNVs in the region are unlikely to be causal variants for ASDs.

We next focused on the ~100 kb linkage disequilibrium block containing the most significant SNPs, and determined whether other transcripts or functional elements are located in the block. By examining the UCSC Genome Browser annotations[26], we did not identify predicted genes, predicted transcription start sites, spliced human expressed sequence tag (EST) sequences, known microRNA genes or predicted microRNA targets that overlap with the linkage disequilibrium block (Supplementary Fig. 5). However, we note that the linkage disequilibrium block contains several highly conserved genomic elements, including a 849-base pair (bp) element that ranks as the top 0.026% most-conserved elements in the entire human genome (log odds (LOD) score = 3,480 by PhastCons[27], Fig. 1b). Consistent with previous reports that large stable gene deserts typically contain regulatory elements for genes involved in development or transcription[28], we

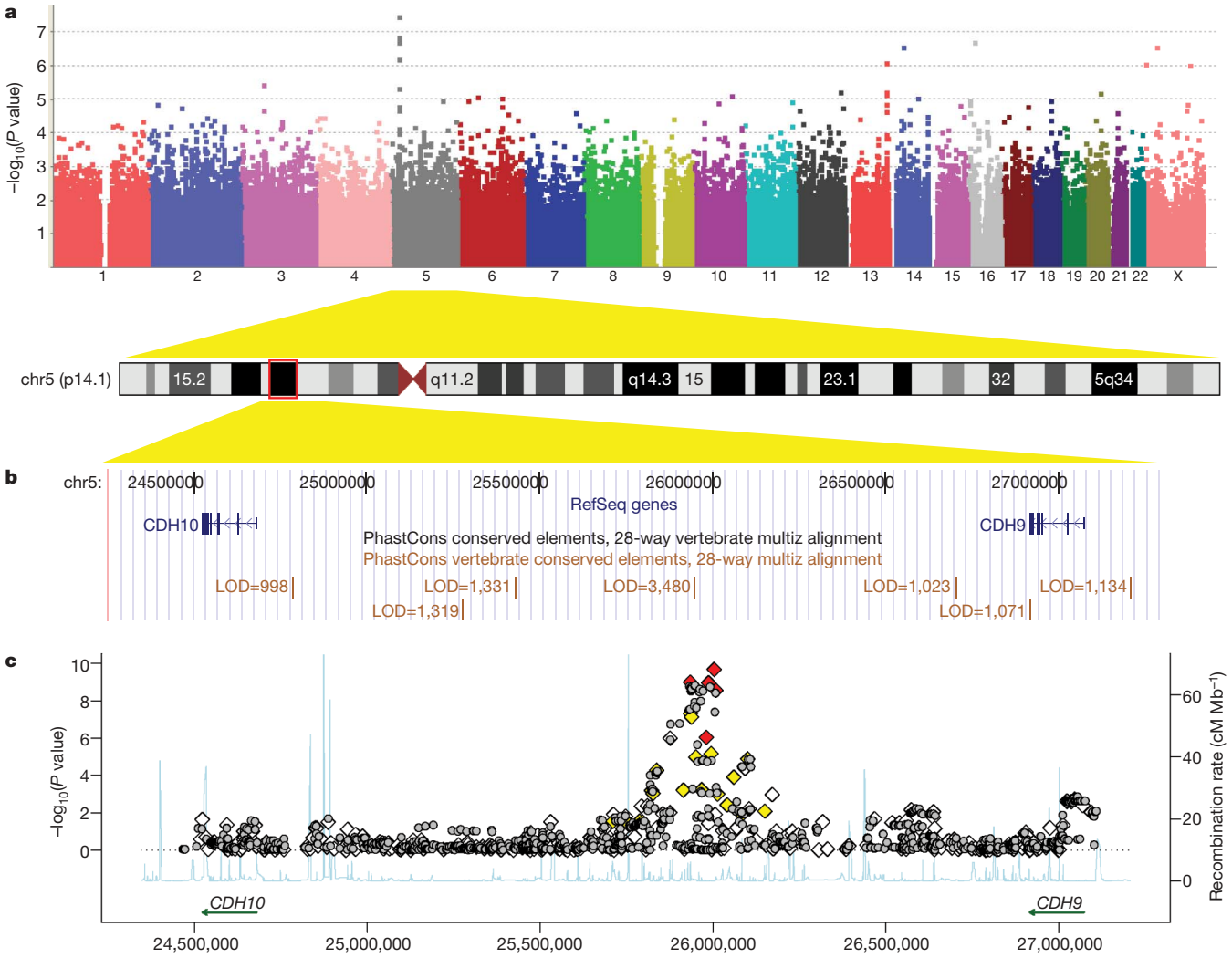


Figure 1 | Genome-wide association results for the 5p14.1 region. **a**, A Manhattan plot showing the $-\log_{10}(P$ values) of SNPs from the combined association analysis of the AGRE and ACC cohorts. **b**, The 5p14.1 genomic region is displayed in UCSC Genome Browser, with conserved genomic elements in the PhastCons track. **c**, Both genotyped (diamonds) and imputed (grey circles) SNPs are plotted with their combined *P* values in all four cohorts. Genotyped SNPs were coloured on the basis of their correlation with rs4307059 (red: $r^2 \geq 0.5$; yellow: $0.2 \leq r^2 < 0.5$; white: $r^2 < 0.2$). Estimated recombination rates from HapMap data are plotted to reflect the local linkage disequilibrium structure.

Table 3 | Genotyped and imputed markers (other than those on 5p14.1) with suggestive association signals ($P < 1 \times 10^{-5}$) in the two discovery cohorts

Chr	SNP	Position	Locus	Gene	Type	MACH Rsq*	Minor / major allele	MAF (AGRE)	*Z* score† (AGRE)	Case MAF (ACC)	Control MAF (ACC)	Odds ratio† (ACC)	*P* value (combined)
3	rs3755827	62335411	3p14.2	Near *FEZF2*	Genotyped		C/T	0.13	3.37	0.14	0.11	1.27	3.5×10^{-6}
6	rs9395885	53853436	6p12.1	*LRRC1*	Imputed	0.99	T/C	0.08	2.27	0.10	0.07	1.43	5.8×10^{-6}
6	rs9349688	53870051	6p12.1	*LRRC1*	Genotyped		G/A	0.09	2.10	0.10	0.07	1.43	8.1×10^{-6}
6	rs9384952	116066757	6q22.1	Between *HS3ST5* and *FRK*	Genotyped		C/T	0.41	2.33	0.42	0.37	1.20	9.4×10^{-6}
10	rs9651325	68649859	10q21.3	*CTNNA3*	Imputed	0.97	A/G	0.11	−3.24	0.09	0.12	0.80	9.2×10^{-6}
10	rs807023	102760072	10q24.31	Near *LZTS2*	Genotyped		A/C	0.14	1.12	0.15	0.10	1.48	7.8×10^{-6}
12	rs10774538	118888180	12q24.23	Near *CCDC64*	Genotyped		T/C	0.13	−2.51	0.12	0.15	0.77	6.2×10^{-6}
13	rs9521337	108823637	13q33.3	Near *MYO16*	Imputed	0.98	T/G	0.13	2.41	0.14	0.11	1.34	6.8×10^{-6}
13	rs4771632	108846743	13q33.3	Near *MYO16*	Imputed	0.99	A/G	0.13	2.32	0.14	0.11	1.32	9.5×10^{-6}
13	rs7996916	108855628	13q33.3	Near *MYO16*	Imputed	0.99	A/C	0.12	3.07	0.15	0.12	1.27	5.5×10^{-6}
13	rs1328250	108856632	13q33.3	Near *MYO16*	Imputed	1.00	T/C	0.12	3.07	0.14	0.12	1.27	5.9×10^{-6}
13	rs4771633	108858788	13q33.3	Near *MYO16*	Imputed	1.00	G/A	0.12	2.98	0.14	0.12	1.26	9.2×10^{-6}
13	rs9521354	108865125	13q33.3	Near *MYO16*	Genotyped		C/A	0.13	3.38	0.14	0.12	1.25	$5.8x10^{-6}$
13	rs9521355	108865183	13q33.3	Near *MYO16*	Genotyped		T/C	0.13	3.20	0.14	0.12	1.25	7.1×10^{-6}
13	rs1328244	108881899	13q33.3	Near *MYO16*	Genotyped		C/T	0.08	3.92	0.10	0.08	1.31	8.2×10^{-7}
14	rs12897470	39895590	14q21.1	Between *FBXO33* and *LRFN5*	Imputed	0.99	A/G	0.50	3.39	0.51	0.47	1.17	2.6×10^{-6}
14	rs7147817	39901754	14q21.1	Between *FBXO33* and *LRFN5*	Genotyped		G/A	0.54	3.38	0.54	0.50	1.20	2.7×10^{-7}
14	rs17783432	76141161	14q24.3	Between *ESRRB* and *VASH1*	Genotyped		T/G	0.15	1.97	0.15	0.11	1.36	9.4×10^{-6}
16	rs7206043	8412954	16p13.2	Between *A2BP1* and *C16orf68*	Imputed	0.98	G/A	0.41	−3.03	0.39	0.43	0.84	4.8×10^{-6}
16	rs9932538	19116070	16p12.3	*SYT17*	Genotyped		G/A	0.16	1.97	0.17	0.12	1.47	1.9×10^{-7}
20	rs6131030	44241393	20q13.12	*CDH22*	Genotyped		A/G	0.42	3.24	0.41	0.37	1.17	6.5×10^{-6}
X	rs11798405	4940801	Xp22.32	Between *PRKX* and *NLGN4X*	Genotyped		G/A	0.09	2.71	0.10	0.06	1.66	9.0×10^{-7}
X	rs5972577	32390211	Xp21.1	*DMD*	Genotyped		A/G	0.33	3.90	0.35	0.30	1.26	2.7×10^{-7}
X	rs6646569	119125802	Xq24	Near *RHOXF1*	Genotyped		T/C	0.13	2.49	0.14	0.09	1.56	9.7×10^{-7}

* MACH Rsq estimates the squared correlation between imputed and true genotypes, and a value less than 0.3 flags poorly imputed SNPs.
† The *Z* score and odds ratio are calculated with respect to the major allele.

hypothesized that these tagging SNPs were capturing the association of functional variants that regulate the expression and action of either *CDH10* or *CDH9*.

Expression of *CDH10* and *CDH9* in brain

Because *CDH10* and *CDH9* are expressed at low levels in non-neural tissues (Supplementary Figs 6 and 7), we evaluated their messenger RNA distribution in human fetal brain by *in situ* hybridization. Multiple sagittally sectioned human fetal brains, each between 19 and 20 weeks gestation, were hybridized with riboprobes against *CDH10* or *CDH9*. Results for *CDH9*, showing uniformly low levels of expression at the time points evaluated, were largely uninformative. In contrast, a marked pattern of enrichment for *CDH10* was observed in the frontal cortex (Fig. 2a)—a region known to be important in ASDs. The expression pattern was similar to that for *CNTNAP2* (contactin-associated protein-like 2)[29], a molecule now well-established to be involved in the ASDs[1]. These results are consistent with previous work showing high levels of *CDH10* in the human fetal brain[30] and a prominent enrichment of *Cdh10* mRNA in the anterior cortical plate of the developing mouse brain[31].

To examine whether the SNP genotypes associate with gene expression for *CDH10* and *CDH9*, we next examined the SNPExpress database[32] that profiles gene expression in 93 human cortical brain tissues from genotyped subjects. However, none of the SNPs in Table 2 was associated with expression levels for either *CDH9* ($P = 0.92$ for rs4307059) or *CDH10* ($P = 0.86$ for rs4307059) (Fig. 2b). Although the small sample size may not have sufficient power to detect subtle effect sizes, it is also possible that the causal variants regulate gene expression only in the developing brain, or that the causal variants target an unidentified functional element, similar to the variants reported in the intergenic region on 8q24, which have been implicated in various cancers[33,34].

Pathway analysis of cell-adhesion genes

Recent genetic studies have identified several neuronal cell-adhesion genes, including *NRXN1* (neurexin 1)[35,36], *CNTNAP2* (refs 37–39) and *PCDH10* (protocadherin 10)[40], as potentially disrupted in rare ASD cases. Cadherins represent a large group of transmembrane proteins that are involved in cell adhesion and the generation of synaptic complexity in the developing brain[41]. In light of the information described earlier, we note that several other cadherin genes were also tagged by the top 1,000 most significant SNPs of the combined discovery cohorts (Supplementary Table 4). In addition, SNPs surrounding several prominent ASD candidate loci[1], including

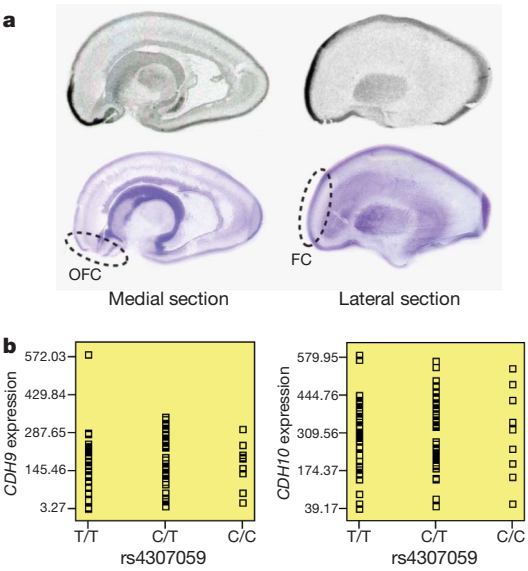


Figure 2 | Examination of brain expression for *CDH10* and *CDH9*. **a**, The *in situ* hybridization of *CDH10* in human fetal brain sectioned in the sagittal plane. Medial and lateral sections from a representative sample are shown above corresponding cresyl-violet-stained marker slides. Orbitofrontal cortex (OFC) and frontal cortex (FC) are highlighted, with marked expression enrichment. **b**, The SNP genotypes of rs4307059 are not associated with *CDH9* or *CDH10* transcript levels in 93 cortical brain tissues.

CACNA1C (L type voltage-gated calcium channel), *CNTNAP2*, *GRIK2* (glutamate receptor, ionotropic, kainate 2), *NRXN1* and *NLGN4X*, also show suggestive evidence of association (Supplementary Table 5). These sources of evidence indicate a potential role for cell-adhesion molecules in the pathogenesis of ASDs.

To examine if cell-adhesion molecules, as a gene family, associate with ASDs, we applied two pathway-based association approaches on the genotype data (Supplementary Methods). First, we assign each SNP to the overlapping or the closest gene, summarize the significance of each gene using the Simes-adjusted *P* value[42] from its SNPs, and then test whether the distribution of *P* values differ between a group of genes and all other genes using a nonparametric rank sum test. Using the combined *P* values from the two discovery cohorts, we found that a group of 25 related cadherin genes show more significant association with ASDs than all other genes ($P = 0.02$), whereas a stronger enrichment signal ($P = 0.004$) was obtained when the 25 cadherin genes were combined with eight neurexin family genes (*NRXN1* to *NRXN3*, *CNTNAP1* to *CNTNAP5*). Second, we analysed the ACC cohort using a formal pathway-association method for case-control data sets[43]. This method examines whether statistics for a group of genes have modest yet consistent deviation from what is expected by chance, through shuffling case/control labels many times, each time recalculating *P* values for all SNPs. We confirmed that the set of cadherin genes is associated with ASDs (permutation $P = 0.02$), whereas the combined cadherin/neurexin genes show more significant association (permutation $P = 0.002$). Therefore, our pathway analysis suggests that neuronal cell-adhesion molecules may be collectively associated with ASDs.

Discussion

Besides recent genetic findings supporting the role of neuronal cell-adhesion molecules in the pathogenesis of autism, an increasing number of functional neuroimaging studies have suggested the presence of cortical underconnectivity in subjects with ASDs[44,45]. Furthermore, neuroanatomy studies have implicated abnormal brain development of the frontal lobes in autism[46,47]. The genetic findings, when coupled with anatomical and functional imaging studies, convergently indicate that ASDs may result from structural and functional disconnection of brain regions that are involved in higher-order associations[48–50], suggesting that ASDs may represent a neuronal disconnection syndrome.

In the current study, we have completed a genetic analysis in a large number of ASD cases and families, with a combined sample set of more than 10,000 subjects of European ancestry. We have identified and replicated common genetic variants on 5p14.1 that are associated with susceptibility to ASDs. Besides the potential roles of the nearby *CDH10* and *CDH9* genes, pathway-based association analysis lend further support to neuronal cell-adhesion molecules in conferring susceptibility to ASDs, suggesting that specific genetic variants in this gene class may be involved in shaping the physical structure and functional connectivity of the brain, that leads to the clinical manifestations of ASDs. Apart from highlighting the genetic complexity of ASDs and the need for large sample sizes in unveiling their genetic causes, our study represents a successful application of the genome-wide association approach in identifying common susceptibility alleles, as part of a larger effort to interrogate the complex genetic architecture of ASDs. Because the genetic aetiologies of ASDs may be linked to the neurobiological components that build and modify connectivity of the brain, by comprehensively identifying the relevant genes, genomic variants and genetic pathways, more focused analysis on gene expression, as well as structural and functional imaging, can be performed on subjects carrying specific genetic defects. Together with studies addressing epigenetic modifications and comprehensive analysis of environmental risk factors, these pieces of information can be better integrated to improve our understanding of the molecular basis of ASDs, and foster the development of early preventive and corrective strategies.

METHODS SUMMARY

All genome-wide SNP genotyping for the discovery cohorts was performed using the Illumina HumanHap550 BeadChip at the Center for Applied Genomics at the Children's Hospital of Philadelphia. For family-based cohorts, the association tests for markers in autosomes and pseudoautosomal region of sex chromosomes were performed by PDT, whereas tests for markers in the X chromosome were performed by X-APL. For case-control cohorts, the association tests were performed by PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). Pathway association analysis was performed by GenGen (http://www.openbioinformatics.org/gengen/), using the genotype data. The whole-genome genotype imputation was performed by MACH (http://www.sph.umich.edu/csg/abecasis/MaCH/) on the autosomal markers, on the basis of phased haplotypes (release 22) for the HapMap CEU population (http://ftp.hapmap.org/phasing/2007-08_rel22/). We removed all markers with MACH Rsq measure of less than 0.3, and zeroed out imputed genotypes with a posterior probability of less than 0.9. The case-control association tests for imputed genotypes were performed by SNPTEST (http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html), which can handle genotype imputation uncertainty. CNV calls were generated by PennCNV (http://www.openbioinformatics.org/penncnv/) on genotyping signal intensity data. For CNV validation by multiplex ligation-dependent probe amplification (MLPA), we used the Universal Probe Library system from Roche, and all reactions were performed in triplicate with an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). For CNV validation by quantitative PCR (qPCR), TaqMan probes were custom-designed using Primer Express 3.0 (Applied Biosystems). For *in situ* hybridization, multiple sagittally sectioned human fetal brains were obtained from the Developmental Brain and Tissue Bank at the University of Maryland. Riboprobes against *CDH9* or *CDH10* were used for hybridization. The SNPExpress database and software (http://people.genome.duke.edu/~dg48/SNPExpress/) were used to examine the genotype-expression relationships.

Received 28 November 2008; accepted 18 March 2009.
Published online 28 April 2009.

1. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9,** 341–355 (2008).
2. Autism and Developmental Disabilities Monitoring Network. ⟨http://www.cdc.gov/mmwr/pdf/ss/ss5601.pdf⟩ (2007).
3. Bailey, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* **25,** 63–77 (1995).
4. Lauritsen, M. B., Pedersen, C. B. & Mortensen, P. B. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. *J. Child Psychol. Psychiatry* **46,** 963–971 (2005).
5. Sykes, N. H. & Lamb, J. A. Autism: the quest for the genes. *Expert Rev. Mol. Med.* **9,** 1–15 (2007).
6. Gupta, A. R. & State, M. W. Recent advances in the genetics of autism. *Biol. Psychiatry* **61,** 429–437 (2007).
7. Freitag, C. M. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry* **12,** 2–22 (2007).
8. Veenstra-VanderWeele, J. & Cook, E. H. Jr. Molecular genetics of autism spectrum disorder. *Mol. Psychiatry* **9,** 819–832 (2004).
9. Vorstman, J. A. S. *et al.* Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry* **11,** 18–28 (2006).
10. Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316,** 445–449 (2007).
11. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
12. Kumar, R. A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17,** 628–638 (2008).
13. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–488 (2008).
14. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–678 (2007).
15. Sklar, P. *et al.* Whole-genome association study of bipolar disorder. *Mol. Psychiatry* **13,** 558–569 (2008).
16. Alkelai, A. *et al.* The XVth world congress of psychiatric genetics, October 7–11, 2007: rapporteur summaries of oral presentations. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **174B,** 233–277 (2008).
17. O'Donovan, M. C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genet.* **40,** 1053–1055 (2008).
18. Ferreira, M. A. *et al.* Collaborative genome-wide association analysis supports a role for *ANK3* and *CACNA1C* in bipolar disorder. *Nature Genet.* **40,** 1056–1058 (2008).
19. Geschwind, D. H. *et al.* The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* **69,** 463–466 (2001).

20. Lord, C., Rutter, M. & Le Couteur, A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **24,** 659–685 (1994).
21. Lord, C. *et al.* The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30,** 205–223 (2000).
22. Martin, E. R., Monks, S. A., Warren, L. L. & Kaplan, N. L. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.* **67,** 146–154 (2000).
23. Chung, R. H., Morris, R. W., Zhang, L., Li, Y. J. & Martin, E. R. X-APL: an improved family-based test of association in the presence of linkage for the X chromosome. *Am. J. Hum. Genet.* **80,** 59–68 (2007).
24. de Bakker, P. I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17,** R122–R128 (2008).
25. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17,** 1665–1674 (2007).
26. Kuhn, R. M. *et al.* The UCSC genome browser database: update 2009. *Nucleic Acids Res.* **37,** D755–D761 (2009).
27. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).
28. Ovcharenko, I. *et al.* Evolution and functional classification of vertebrate gene deserts. *Genome Res.* **15,** 137–145 (2005).
29. Abrahams, B. S. *et al.* Genome-wide analyses of human perisylvian cerebral cortical patterning. *Proc. Natl Acad. Sci. USA* **104,** 17849–17854 (2007).
30. Kools, P., Vanhalst, K., Van den Eynde, E. & van Roy, F. The human cadherin-10 gene: complete coding sequence, predominant expression in the brain, and mapping on chromosome 5p13–14. *FEBS Lett.* **452,** 328–334 (1999).
31. Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* **32,** D552–D556 (2004).
32. Heinzen, E. L. *et al.* Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.* **6,** e1000001 (2008).
33. Witte, J. S. Multiple prostate cancer risk variants on 8q24. *Nature Genet.* **39,** 579–580 (2007).
34. Ghoussaini, M. *et al.* Multiple loci with different cancer specificities within the 8q24 gene desert. *J. Natl. Cancer Inst.* **100,** 962–966 (2008).
35. Kim, H. G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 199–207 (2008).
36. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39,** 319–328 (2007).
37. Arking, D. E. *et al.* A common genetic variant in the neurexin superfamily member *CNTNAP2* increases familial risk of autism. *Am. J. Hum. Genet.* **82,** 160–164 (2008).
38. Alarcon, M. *et al.* Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82,** 150–159 (2008).
39. Bakkaloglu, B. *et al.* Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am. J. Hum. Genet.* **82,** 165–173 (2008).
40. Morrow, E. M. *et al.* Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321,** 218–223 (2008).
41. Redies, C. Cadherins in the central nervous system. *Prog. Neurobiol.* **61,** 611–648 (2000).
42. Simes, R. J. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73,** 751–754 (1986).
43. Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* **81,** 1278–1283 (2007).
44. Just, M. A., Cherkassky, V. L., Keller, T. A. & Minshew, N. J. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain* **127,** 1811–1821 (2004).
45. Koshino, H. *et al.* fMRI investigation of working memory for faces in autism: visual coding and underconnectivity with frontal areas. *Cereb. Cortex* **18,** 289–300 (2008).
46. Amaral, D. G., Schumann, C. M. & Nordahl, C. W. Neuroanatomy of autism. *Trends Neurosci.* **31,** 137–145 (2008).
47. Courchesne, E. & Pierce, K. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Curr. Opin. Neurobiol.* **15,** 225–230 (2005).
48. Hughes, J. R. Autism: the first firm finding = underconnectivity? *Epilepsy Behav.* **11,** 20–24 (2007).
49. Frith, C. Is autism a disconnection disorder? *Lancet Neurol.* **3,** 577 (2004).
50. Geschwind, D. H. & Levitt, P. Autism spectrum disorders: developmental disconnection syndromes. *Curr. Opin. Neurobiol.* **17,** 103–111 (2007).
51. Ma, D. Q. *et al.* A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann. Hum. Genet.* doi:10.1111/j.1469-1809.2008.00523.x (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully thank all the children with ASDs and their families at the participating study sites who were enrolled in this study and all the control subjects who donated blood samples to Children's Hospital of Philadelphia (CHOP) for genetic research purposes. We also acknowledge the resources provided by the AGRE Consortium (D. H. Geschwind, M. Bucan, W. T. Brown, J. D. Buxbaum, R. M. Cantor, J. N. Constantino, T. C. Gilliam, C. M. Lajonchere, D. H. Ledbetter, C. Lese-Martin, J. Miller, S. F. Nelson, G. D. Schellenberg, C. A. Samango-Sprouse, S. Spence, M. State, R. E. Tanzi) and the participating families. AGRE is a program of Autism Speaks and is at present supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to C. M. Lajonchere (PI), and formerly by grant MH64547 to D. H. Geschwind (PI). We thank the technical staff at the Center for Applied Genomics at CHOP for producing the genotypes used for analyses, and the nursing, medical assistant and medical staff for their help with recruitment of patient and control subjects for the study. We thank R. Liu and I. Lindquist for helping with CNV validation. We thank D. J. Hedges, H. N. Cukier, J. L. McCauley, G. W. Beecham, H. H. Wright, R. K. Abramson, E. R. Martin and J. P. Hussman for their comments, advice and statistical support, and the laboratory core and the autism clinical personnel at the Miami Institute for Human Genomics and the autism clinical staff at the Vanderbilt Center for Human Genetics Research. A subset of the CAP participants was ascertained while M.A.P.-V. was a faculty member at Duke University. We thank the National Institutes of Health (NIH)-funded Developmental Brain and Tissue Bank at University of Maryland for access to the fetal brain tissues used in these studies (National Institute of Child Health and Human Development Contract no. NO1-HD-4-3368 and NO1-HD-4-3383). All genotyping of the AGRE and ACC cohort was supported by an Institutional Development Award to the Center for Applied Genomics (H.H.) at the Children's Hospital of Philadelphia. The study was supported in part by a Research Award from the Margaret Q. Landenberger Foundation (H.H.), a Research Development Award from the Cotswold Foundation (H.H. and S.F.A.G), UL1-RR024134-03 (H.H.), an Alavi-Dabiri fellowship from Mental Retardation and Developmental Disability Research Center at CHOP (K.W.), the Beatrice and Stanley A. Seaver Foundation (J.D.B.), the Department of Veterans Affairs (G.D.S.), NIH grants HD055782-01 (J.Munson, A.E., O.K., G.D. and G.D.S.), MH0666730 (J.D.B.), MH061009 and NS049261 (J.S.S.), HD055751 (E.H.C.), MH69359, M01-RR00064 and the Utah Autism Foundation (H.C., J.Miller and W.M.M.), MH64547, MH081754 (D.H.G.), HD055784 (D.H.G. and M.S.), NS26630, NS36768, MH080647 and a gift from the Hussman Foundation (M.A.P.-V.), the Autism Genome Project Consortium (B.S.A., J.P., C.W.B., D.H.G., T.H.W., W.M.M., H.C., J.I.N., J.S.S., E.H.C., J.Munson, A.E., O.K., J.D.B., B.D. and G.D.S.) funded by Autism Speaks, the Medical Research Council (UK) and the Health Research Board (Ireland). We also acknowledge the partial support to CAP cohort from the Autism Genome Project.

Author Contributions H.H. and G.D.S. designed the study and H.H. supervised the genotyping, data analysis and interpretation. K.W., H.Z. and D.M. analysed the AGRE, ACC/CART and CAP data sets, respectively. K.W. drafted the manuscript, and H.H., G.D.S. and other authors edited the manuscript. M.B., J.T.G, M.I., J.P.B., P.M.A.S., C.E.K., C.H., E.F., R.C., C.M.L., R.M.C. and S.F.A.G. helped generate data and assisted with data analysis of the AGRE cohort. D.M., D.S., J.R.G. and M.L.C. generated data for the CAP replication cohort. B.S.A., L.I.S., A.I.A.R., E.I.H., H.D., T.H., M.S., S.O. and A.K. performed *in situ* hybridization, and generated data for the CART replication cohort. K.W., C.E.K. and E.R. performed qPCR validation of CNVs. N.T., T.S. and J.D.B. preformed MLPA validation of CNVs. J.Munson, A.E., O.K., J.P., T.O., J.A.S. C.W.B., R.B. J.R.G., W.M.M., J.Miller, M.W.S., T.H.W., H.C., S.E.L., R.T.S., J.I.N., J.L.H., J.S.S., E.H.C., N.J.M., J.D.B., G.D., D.H.G., M.A.P.-V. and G.D.S. collected samples, contributed phenotype data for the study, helped with interpretation of data, and assisted with manuscript preparation. D.H.G. and M.A.P.-V. contributed equally to this work.

Author Information The AGRE data set was genotyped by the Center for Applied Genomics at CHOP, and the complete sets of genotype and signal intensity data have been released to the public domain. AGRE-approved academic researchers can acquire the data sets from AGRE (http://www.agre.org). In addition, the summary statistics for the full data set will be made available in the repository of the NIH Genotype and Phenotype database (dbGAP; http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.H. (hakonarson@chop.edu).

ARTICLES

Hippocampal theta oscillations are travelling waves

Evgueniy V. Lubenov[1] & Athanassios G. Siapas[1]

Theta oscillations clock hippocampal activity during awake behaviour and rapid eye movement (REM) sleep. These oscillations are prominent in the local field potential, and they also reflect the subthreshold membrane potential and strongly modulate the spiking of hippocampal neurons. The prevailing view is that theta oscillations are synchronized throughout the hippocampus, despite the lack of conclusive experimental evidence. In contrast, here we show that in freely behaving rats, theta oscillations in area CA1 are travelling waves that propagate roughly along the septotemporal axis of the hippocampus. Furthermore, we find that spiking in the CA1 pyramidal cell layer is modulated in a consistent travelling wave pattern. Our results demonstrate that theta oscillations pattern hippocampal activity not only in time, but also across anatomical space. The presence of travelling waves indicates that the instantaneous output of the hippocampus is topographically organized and represents a segment, rather than a point, of physical space.

Theta oscillations are a prominent 4–10-Hz rhythm in the hippocampal local field potential (LFP) of all mammals studied to date[1–3], including humans[4]. During wakefulness they are associated with different species-specific behaviours, and they are invariably present during REM sleep[2,3]. In the rat, theta oscillations always accompany voluntary movement and active exploration[2,5]. Theta oscillations are essential for the normal functioning of the hippocampus, because manipulations that disrupt them produce behavioural impairments that mimic hippocampal lesions[6,7]. The importance of theta oscillations is underscored by the fact that they reflect subthreshold membrane potentials[8–10] and strongly modulate the spiking[5,11,12,13] of hippocampal neurons. Furthermore, theta oscillations gate synaptic plasticity, because the timing of stimulation with respect to the phase of theta is important in determining the magnitude and direction of synaptic change[14,15]. Theta oscillations therefore offer macroscopic access to the internal clock of the hippocampal circuit, responsible for temporally patterning its operation. Such clocking is essential for the temporal coding of spatial information by place cells[5,16], as evidenced by theta phase precession[17–19]. In addition to coding position, theta phase precession ensures that the order of place-cell firing over behavioural timescales (seconds) is preserved and compressed within individual theta cycles and inside the window of plasticity[18]. In the presence of spike-timing-dependent plasticity[20], the resulting compression of temporal sequences offers a mechanism for the formation of hippocampal memory traces[21]. Furthermore, theta oscillations modulate activity not only in the hippocampus, but also in several subcortical, limbic and cortical structures[22–24].

If theta oscillations can be thought of as a clock, what time is it in different parts of the hippocampus? In other words, how does the phase of theta oscillations depend on the anatomical location of the recorded LFP? It is well known that the amplitude and phase of theta oscillations depend on the depth, or more precisely the lamina, of the observed signal[25–27], because several current dipoles located in different layers contribute to the macroscopic oscillation[12,28–32]. Conversely, it is believed that theta oscillations are robustly synchronized within each layer across the anatomical extent of the hippocampus[28,32–34], despite some early evidence to the contrary[26]. The dominant view is therefore that theta oscillations are a global clock, and the anatomical extent of the hippocampus is synchronized with zero delay or, figuratively speaking, contained in a single time zone[32]. In contrast to this view, we provide direct evidence that theta oscillations are not synchronized within stratum oriens of CA1, but are travelling waves that propagate predominantly along the septotemporal (long) hippocampal axis. Furthermore, we show that the firing of CA1 neurons is modulated in a wave pattern that is predicted by the LFP waves. Therefore, time in the hippocampus, as clocked by theta oscillations, is anatomically organized in a way similar to time on Earth—in a progression of local time zones.

Depth profile of theta oscillations

How can the depth dependence of theta phase be disentangled from systematic variations across the hippocampal surface? The solution requires recordings from a lamina in which the phase of theta oscillations shows little or no dependence on depth. We sought to determine whether such a lamina exists, and if so, whether it is sufficiently thick to allow reliable electrode positioning. To investigate this issue, we used multisite silicon probes with 50 μm vertical site spacing (Supplementary Information, section 1.1), and quantified with high resolution the depth profile of theta oscillations in behaving (Fig. 1) and urethane-anesthetized rats (Supplementary Fig. 2). In agreement with several previous studies[27,30,34], we found that theta oscillations start reversing their phase slightly below the pyramidal cell layer, as determined by the peak amplitude of sharp-wave-associated ripples and the reversal of sharp-wave polarity (Fig. 1e and Supplementary Fig. 4d). Moreover, there is a notable phase gradient throughout the dorsoventral (vertical) extent of stratum radiatum. In contrast, the phase of theta oscillations is practically constant throughout the vertical extent of a large slab of tissue (at least 400 μm thick), just dorsal to the CA1 pyramidal cell layer and including the entire stratum oriens (Fig. 1b). Therefore, we can reliably study the spatial organization of theta oscillations by comparing LFPs recorded at any depth within stratum oriens of CA1, but at systematically different septotemporal and transverse coordinates in the hippocampus.

Topography of theta phase offsets

To sample theta oscillations topographically across the CA1 circuitry, we used microdrive arrays targeting rectangular grids spanning

[1]Division of Biology, Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125, USA.

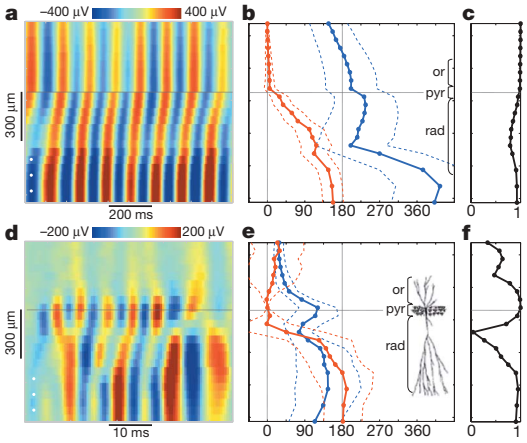


Figure 1 | Interlaminar depth profile of theta oscillations and ripples. **a**, Simultaneous CA1 recordings of LFPs filtered in the 4–10 Hz band during theta oscillations. In all panels data are ordered from most superficial (top) to deepest (bottom), with the pyramidal cell layer (pyr) marked by a horizontal line. White dots mark three inoperable sites where the image was interpolated. **b**, Theta amplitudes (blue, μV) and phase deviations (red, deg) with respect to the top signal. Solid and dashed lines are the means and s.d., respectively. or, oriens; rad, radiatum. **c**, Theta coherence with respect to the top signal. **d**–**f**, Same as **a**–**c** but for LFPs filtered in the 80–250 Hz band during a ripple. Phase deviations and coherence are computed with respect to the signal at the pyramidal cell layer.

approximately 5 mm^2 of the hippocampus (0.5-mm spacing) and allowing the independent depth adjustment of 24 tetrodes and four single reference electrodes (Supplementary Fig. 3). The arrays were chronically implanted and tetrodes were positioned in stratum oriens of CA1 on the basis of sharp-wave polarity and ripple power (Supplementary Information, section 1.2 and Supplementary Fig. 4). Figure 2 illustrates the anatomical position of one adjusted recording grid with respect to the hippocampus, and shows example recordings obtained while the rat ran on a linear track (Supplementary Figs 18–27).

Clear phase offsets were present in theta oscillations recorded in stratum oriens of CA1 from different locations on the grid (Fig. 2g), whereas their coherence across the grid was almost perfect (Supplementary Fig. 5). This is in contrast to the phase constancy of theta oscillations as a function of depth within stratum oriens discussed earlier (Fig. 1).

Next, we investigated whether the observed variability in the phase of theta oscillations is systematically organized across the hippocampus. To address this question (see Supplementary Information, section 1.3) we computed the instantaneous phase differences between theta oscillations recorded at each location on the grid and a reference tetrode in the centre of each grid (Fig. 3). These intralaminar phase differences exposed notable order and consistency across animals (Fig. 3). In particular, the relative phase of theta oscillations progressively advances from medial to lateral, whereas it is approximately constant along the rostrocaudal axis. Furthermore, the equiphase contours are roughly linear and are therefore well approximated by a planar (constant) phase gradient (analysis of variance (ANOVA) for multiple regression $P < 10^{-9}$, $R^2 > 0.88$, in all three animals). The magnitude of the phase gradient in the three examples from different rats (Fig. 3a–c) was 21.4, 21.7 and 20.3° mm^{-1}, and its anatomical orientation was 1°, 16° and 21°, respectively, with 0° corresponding to the mediolateral and −90° to the rostrocaudal direction. The parsimonious explanation for the observed intralaminar theta phase profiles is that theta oscillations are in fact travelling plane waves that propagate with consistent speeds and directions across animals.

Theta oscillations are travelling waves

We analysed the instantaneous organization of theta oscillations and observed clear travelling plane waves (Fig. 4a, Supplementary Information, section 2 and Supplementary Movies 1–3). In fact, whenever robust theta oscillations were present in the hippocampus, they appeared as travelling waves, similar to the example shown in Fig. 4. The spatial wavelength and the direction of wave propagation were estimated at every point in time by fitting a plane to the instantaneous theta phase deviations (Supplementary Information, section 1.3 and Fig. 4b). This showed that theta waves can propagate not only in their

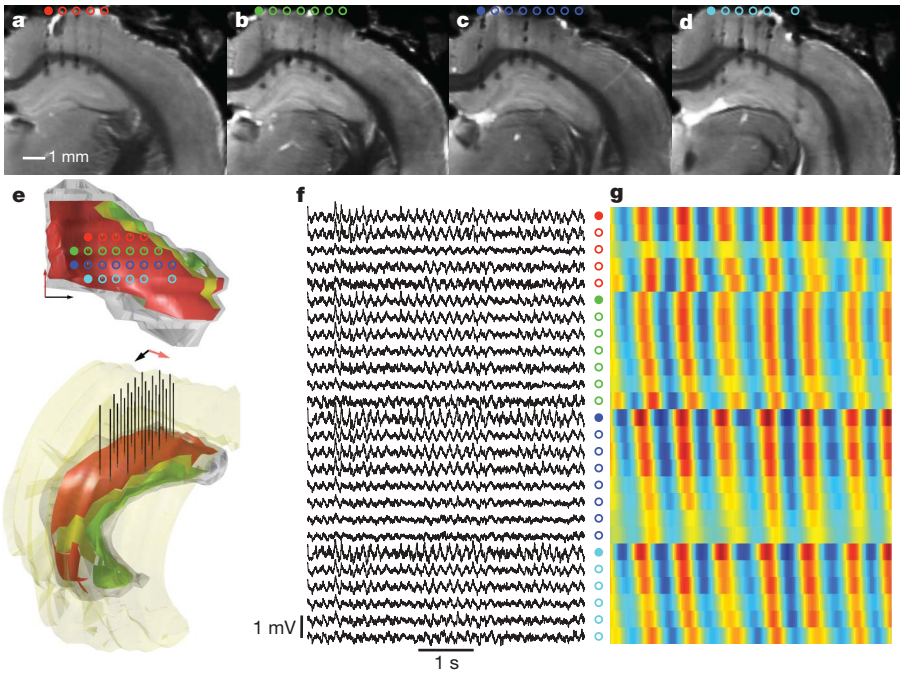


Figure 2 | Example of theta phase deviations. **a**–**d**, MRI coronal sections at bregma −3.8 (**a**), −4.3 (**b**), −4.8 (**c**) and −5.3 (**d**) (mm), showing final tetrode positions in the 4×8 grid. **e**, Bottom, three-dimensional model of the right hippocampus, with the CA1 pyramidal cell layer in red and tetrodes in black. The red arrow points rostrally and the black arrow points laterally. Top, top-down view with tetrodes marked by coloured circles. **f**, LFPs recorded from stratum oriens of CA1 using the grid in **e** during theta oscillations. Signals from each grid row are ordered from medial to lateral and stacked from rostral to caudal as indicated by the colour circles. **g**, LFPs filtered in the 4–10 Hz band from the interval marked by the horizontal bar in **f**.

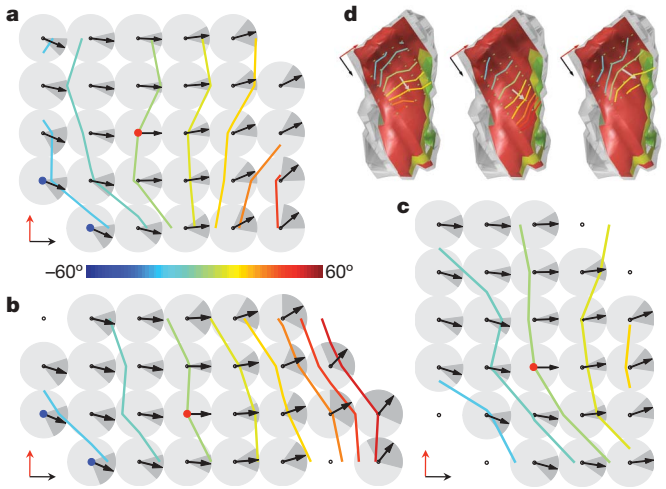


Figure 3 | Topographic intralaminar theta phase profiles. a–c, Grey circles mark grid points where LFPs were recorded over 45 min while three rats ran on a linear track. Black arrows show the means of the instantaneous theta phase deviations relative to a channel from the middle of each grid (red dots). Dark grey sectors indicate the corresponding s.d. The coloured lines are the contours of equal mean theta phase deviation. Grid points (blue dots) that may be in the subiculum have theta phase deviations consistent with CA1. **d**, The contours from **a–c** are replotted over a top-down view of the brain. Grey arrows show the mean direction and speed of wave propagation.

dominant direction but also within a range of ±45° around the dominant direction (Fig. 4c). The speed of theta wave propagation, v (Fig. 4d), and the spatial wavelength, λ (Fig. 4e), also showed a moderate level of variability (coefficient of variation ~0.5). The median values of the propagation direction (with respect to the mediolateral axis), speed, and spatial wavelength for each animal were as follows: $\theta = 22°$, 3° and −17°, $v = 87$, 80 and 107 mm s^{-1}, and $\lambda = 13.6$, 11.1 and 14.9 mm. The mean propagation direction with respect to the septotemporal axis was 32°, with its transverse component pointing from distal to proximal CA1 (towards CA3). Wave parameter distributions were very similar across animals, although the within animal similarities were even higher, indicating the presence of small systematic individual differences (Fig. 4c–e).

Travelling waves of spiking activity

Because the LFP recorded at each point on the grid reflects the collective behaviour of the local neuronal population, the fact that theta oscillations are travelling waves indicates that neurons in CA1 must spike in a wave pattern consistent with the LFP wave (Supplementary Information, section 1.4). We experimentally tested this prediction by positioning the tetrode grid within the CA1 pyramidal cell layer and recording the spiking activity of several neurons while rats ran on a linear track (Supplementary Figs 6, 8 and 9). To overcome the sparseness of firing on the timescale of a theta cycle (Supplementary Fig. 6b) we computed the mean phase of firing of spikes recorded at each point of the grid with respect to a single theta reference oscillation (Supplementary Figs 12 and 13). If spiking in CA1 is synchronized, the mean phase of firing at each point of the grid should be identical. In contrast, if spiking is organized as a travelling wave, the mean phase of firing should systematically advance across the grid in the direction of wave propagation. The latter is indeed the case when all recorded spikes are considered (Fig. 5, ANOVA for multiple regression, 5×6 grid: $P < 0.007$, $R^2 = 0.41$; 4×8 grid: $P < 0.0006$, $R^2 = 0.59$), as well as when only the spikes of active pyramidal cells (Supplementary Figs 7–12) are analysed (Supplementary Fig. 14, ANOVA for multiple regression $P < 0.03$, $R^2 = 0.51$, two-sample test of mean angles $P < 0.001$). Furthermore, the spiking wave parameters inferred from the topographic organization of mean phases match the characteristics

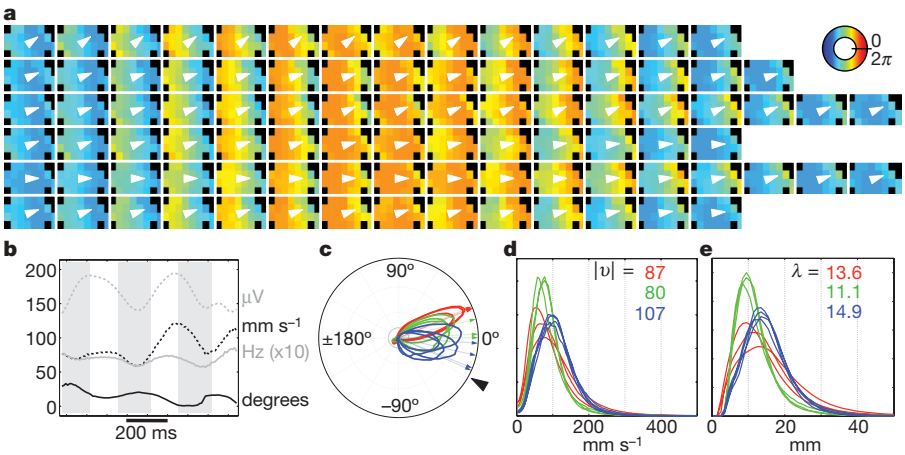


Figure 4 | Travelling wave example and wave parameter distributions. **a**, Frames show consecutive snapshots of theta oscillations across the recording grid, and white arrowheads mark the direction of wave propagation. Colours indicate the phase of theta at each grid location according to the circular colour map on the right. **b**, Instantaneous wave parameters for the data in **a**: amplitude (dashed grey line), theta frequency (solid grey line), speed (dashed black line) and direction (solid black line) of wave propagation. The alternating white and grey stripes denote the six theta cycles shown as the six rows of frames in **a**. **c–e**, Parameter distributions from ten separate recording sessions: wave propagation direction (0°, mediolateral; black triangle, septotemporal) (**c**), wave propagation speed (**d**), and spatial wavelength (**e**). Different colours correspond to different animals, and the numbers show the median parameter values for each animal.

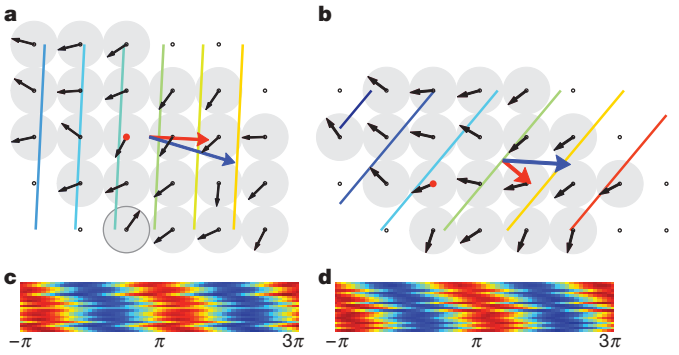


Figure 5 | Topographic profile of spiking theta phase preference. **a**, **b**, Multiunit mean phase of spiking at each grid location (black arrows) with respect to a single reference theta trace (red dot) in the 5×6 (**a**) and 4×8 (**b**) grids. The coloured lines are the level curves of the best linear fit to the unfolded mean phases. Large arrows mark the direction of wave propagation and spatial wavelength of spiking (red) and LFP (blue) waves. **c**, **d**, Each row shows the normalized multiunit theta phase distribution at a grid location and rows are ordered in the direction of spiking wave propagation. The slanted appearance demonstrates that the spiking probability is modulated in a wave pattern. Synchronous spiking would have resulted in vertical stripes.

of the theta wave seen in the LFPs (Fig. 5a, b). Parameter values were as follows (5×6 grid) LFP wave: $\theta = -17^\circ$, $\lambda = 23.2$ mm, ANOVA $P < 10^{-5}$, $R^2 = 0.62$; spiking wave: $\theta = -3^\circ$, $\lambda = 15.4$ mm, ANOVA after marked outlier exclusion $P < 0.007$, $R^2 = 0.41$; (4×8 grid) LFP wave: $\theta = -3^\circ$, $\lambda = 17.4$ mm, ANOVA $P < 0.0004$, $R^2 = 0.57$; spiking wave: $\theta = -40^\circ$, $\lambda = 9.3$ mm, ANOVA $P < 0.0006$, $R^2 = 0.59$. Notice that the LFP and spiking-wave parameters are extracted from non-overlapping bands of the recorded signals and therefore represent independent measurements.

Functional implications

We have demonstrated that within the septal third of the hippocampus theta oscillations and CA1 firing are organized as travelling waves, propagating in a direction with a dominant septotemporal component, and wavelength commensurate with the long hippocampal axis. We do not know yet if the waves travel throughout the entire septotemporal axis of the hippocampus. Assuming, for now, that they do (Supplementary Fig. 15), we examine their potential functional implications. The anatomical projections of CA1 are systematically organized along the septotemporal axis with pronounced differences between the two poles[35]. For example, the septal third of CA1 projects to retrosplenial and perirhinal cortex, whereas projections to the hypothalamus, lateral septum, amygdala and medial prefrontal cortex preferentially originate from the temporal third of CA1 (refs 35–37). Travelling waves ensure that these distinct hippocampal targets receive peak CA1 input in a particular order. This may be important for the integration of heterogeneous information in downstream networks. Furthermore, there is growing evidence that the size of hippocampal place fields systematically increases along the septotemporal axis[38–40]. In the presence of travelling waves, this indicates that CA1 output may cycle from finer to coarser spatial resolution within each theta period.

Although the peak of CA1 activity travels septotemporally, an individual hippocampal neuron can still fire throughout the theta cycle, albeit with modulated intensity. Therefore, at each instant there are neurons discharging at every septotemporal level. A question remains as to how travelling waves inform what is represented by these populations of co-active cells. In addition to place specificity of firing[16] (Supplementary Fig. 11), pyramidal neurons in the rat hippocampus show a strong correlation between place and phase of firing with respect to the local theta cycle, that is, theta phase precession[17,19] (Supplementary Fig. 10). The phase precession relation allows one to infer the location of the place-field centres of the neurons that fire in the course of a theta cycle (Fig. 6a). Consider, for example, a population of neurons with similar field sizes and four successive time windows in a theta cycle (Fig. 6a). Place specificity dictates that only the neurons with fields that overlap the rat's current location will fire (Fig. 6b), but phase precession further suggests that they will fire in sequence, rather than at once (Fig. 6c). Neurons with field centres behind the rat's location fire early in the theta cycle, whereas neurons with field centres ahead of the rat fire late (Fig. 6c). In this manner, a segment of space the length of a place field is traced in the course of a cycle. This effect, known as "compression of temporal sequences"[18], has been established experimentally[18,41] and presumably occurs at every septotemporal level. If theta oscillations are zero-lag synchronized, the compression of temporal sequences occurs in phase along the septotemporal axis and the instantaneous output of CA1 encodes a point of physical space (Fig. 6c). When the septotemporal expansion of place fields is considered[38–40] and the nonlinearity of phase precession is accounted for[17,18], the range of physical space represented by the hippocampus under synchronous oscillations depends on the phase of theta, shrinks to a point at least once, and is not consistently mapped on the septotemporal axis (Supplementary Figs 16 and 17a, c, e, bottom panel). Travelling theta waves introduce consistent theta phase offsets between different septotemporal levels, and so the compression of temporal sequences is systematically staggered along the septotemporal axis (Fig. 6d). In the presence of waves, at every instant a segment of physical space of characteristic length is topographically mapped along the septotemporal axis (Fig. 6d and Supplementary Figs 16, 17b, d, f, bottom panel). Time circularly shifts the encoded segment along the wave propagation direction, making it wrap around as if the septal and temporal poles were connected (Fig. 6d). So the fact that theta oscillations are travelling waves suggests that the instantaneous output of the hippocampus represents a segment, rather than a point, of physical space.

Possible mechanisms

Theoretical analysis of travelling waves in the cortex has suggested at least three distinct mechanisms[42], and experimental data indicate that all three may be involved in theta oscillations. The first and simplest mechanism is delayed excitation from a single oscillator. This mechanism may apply to the medial septum, which is considered to be the principal

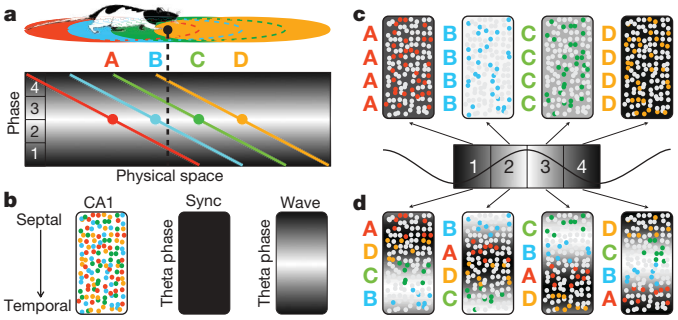


Figure 6 | Hippocampal representation of physical space. **a**, CA1 place fields with centres A–D overlapping the position of the rat (black line). The diagonal lines below illustrate the corresponding phase precession relations. The *x* axis is physical location, and the *y* axis is the local phase of theta oscillations (grey-scale values). The numbered squares mark four time windows within the theta cycle. **b**, Unfolded map of CA1 with dots showing all place cells with fields overlapping the rat's position (colours correspond to field centres). The grey-scale panels show snapshots of the instantaneous theta phase throughout CA1 in the case of synchronized (sync) oscillations and travelling waves. **c**, Patterns of CA1 activity if theta oscillations are synchronized. Each panel is as in **b**, but only the cells active within each time window are coloured. At each time a single phase of theta is mapped across CA1, and consequently cells representing a single point in physical space are co-activated. **d**, If theta oscillations are travelling waves, at each time point a range of local theta phases is mapped across the hippocampus and consequently cells representing the segment ABCD are co-activated.

pacemaker of theta oscillations and whose projections to the hippocampus are topographically organized along the septotemporal axis and travel through four separate routes with potentially different transmission delays[35]. The second mechanism is propagation of activity in an excitable network. This is relevant for the contribution of area CA3, because in disinhibited longitudinal slices pulses of activity propagate unattenuated throughout the length of CA3 (10 mm) with speeds of 0.1–0.15 m s^{-1} (refs 43, 44), matching the direction and speed of wave propagation we describe. The third mechanism requires a gradient of natural frequencies in a network of weakly coupled oscillators. Neurons in the entorhinal cortex have intrinsically oscillating membranes with natural frequencies in the theta range[45,46]. Furthermore, the natural frequencies are topographically organized and decrease in the dorsoventral direction[47], therefore providing the conditions for intrinsic travelling waves within the entorhinal cortex. Similarly, neurons within CA1 itself exhibit theta frequency resonance[48] and it has been argued that resonant frequencies decrease along the septotemporal axis[39], consistent with the observed wave propagation direction. Finally, it is important to emphasize that the brain areas exhibiting theta oscillations or containing theta phase-locked neurons are not connected in a simple feedforward network, but are part of a feedback loop with elaborate interconnection topographies. Therefore, theta resonance at the level of individual neurons[45–48], local interactions between neuronal populations[49], and global coupling between areas through connections with various transmission delays[50] all play a part in determining the spatiotemporal characteristics of theta oscillations throughout the loop. Given the above considerations, we predict that theta oscillations are travelling waves not only in CA1, but also in dentate gyrus, CA3, subiculum and the entorhinal cortex.

Travelling waves with consistent propagation parameters within a circuit represent topographic maps of temporal activation order. Therefore, in conjunction with other known topographies and local circuit properties, they can elucidate not only the global features of information representation but also the direction of information flow and the nature of information processing.

METHODS SUMMARY

Long–Evans male rats, 3–6 months old, were used in all experiments. Chronic recordings were obtained while rats ran on a 180-cm linear track for water reward. Acute recordings were obtained under urethane (1.3 g kg^{-1}) anaesthesia. Two skull screws above the contralateral cerebellum and cortex served as ground and reference in all recordings.

Multisite silicon probe experiments. Rats were implanted chronically ($n = 2$) and acutely ($n = 2$) with multisite silicon probes targeting bregma = −3.9, lateral = 2.7 (NeuroNexus Technologies, probe 2b: 54 sites in two staggered columns 43 μm apart, 50 μm vertical site spacing). During surgery, probes were stereotactically lowered and secured so that the middle of their recording vertical extent was approximately at the CA1 pyramidal cell layer.

Two-stage chronic microdrive array experiments. Rats ($n = 3$) were implanted with custom arrays targeting rectangular grids (6×5, 4×8 and 5×6), allowing the independent depth adjustment of 24 tetrodes and four reference electrodes. In all grids, probe spacing was 0.5 mm and the target of the caudomedial grid corner was bregma = −5.3, lateral = 1.0. In phase 1, tetrodes were positioned in stratum oriens of CA1 and the data used in the LFP theta wave analysis were obtained. In phase 2, tetrodes were lowered to the CA1 pyramidal cell layer and the data used in the spiking wave analysis were recorded.

Electrode position reconstruction. Sharp-wave polarity analysis was used during phase 1 to confirm the depth of all tetrodes within stratum oriens above CA1 (Supplementary Information, section 1.2.2 and Supplementary Fig. 1a). At the conclusion of phase 2, the final location of each tetrode tip was marked by an electrolytic lesion (10 μA anodal current for 15 s), brains were fixed, and a high-resolution structural MRI scan was performed to reconstruct electrode tracks in three-dimensions (Supplementary Fig. 1c).

Received 7 February; accepted 18 March 2009.
Published online 17 May 2009.

1. Green, J. D. & Arduini, A. A. Hippocampal electrical activity in arousal. *J. Neurophysiol.* **17,** 533–557 (1954).
2. Vanderwolf, C. H. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr. Clin. Neurophysiol.* **26,** 407–418 (1969).
3. Winson, J. Interspecies differences in the occurrence of theta. *Behav. Biol.* **7,** 479–487 (1972).
4. Arnolds, D. E., Lopes da Silva, F. H., Aitink, J. W., Kamp, A. & Boeijinga, P. The spectral properties of hippocampal EEG related to behaviour in man. *Electroencephalogr. Clin. Neurophysiol.* **50,** 324–328 (1980).
5. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
6. Winson, J. Loss of hippocampal theta rhythm results in spatial memory deficit in the rat. *Science* **201,** 160–163 (1978).
7. Mitchell, S. J., Rawlins, J. N., Steward, O. & Olton, D. S. Medial septal area lesions disrupt theta rhythm and cholinergic staining in medial entorhinal cortex and produce impaired radial arm maze behavior in rats. *J. Neurosci.* **2,** 292–302 (1982).
8. Fujita, Y. & Sato, T. Intracellular records from hippocampal pyramidal cells in rabbit during theta rhythm activity. *J. Neurophysiol.* **27,** 1012–1025 (1964).
9. Leung, L. S. & Yim, C. Y. Intracellular records of theta rhythm in hippocampal CA1 cells of the rat. *Brain Res.* **367,** 323–327 (1986).
10. Vertes, R. P. & Kocsis, B. Brainstem-diencephalo-septohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience* **81,** 893–926 (1997).
11. Ranck, J. B. Jr. Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. I. Behavioral correlates and firing repertoires. *Exp. Neurol.* **41,** 461–531 (1973).
12. Buzsaki, G., Leung, L. W. S. & Vanderwolf, C. H. Cellular bases of hippocampal EEG in the behaving rat. *Brain Res.* **287,** 139–171 (1983).
13. Fox, S. E., Wolfson, S. & Ranck, J. B. Jr. Hippocampal theta rhythm and the firing of neurons in walking and urethane anesthetized rats. *Exp. Brain Res.* **62,** 495–508 (1986).
14. Pavlides, C., Greenstein, Y. J., Grudman, M. & Winson, J. Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. *Brain Res.* **439,** 383–387 (1988).
15. Huerta, P. T. & Lisman, J. E. Heightened synaptic plasticity of hippocampal CA1 neurons during a cholinergically induced rhythmic state. *Nature* **364,** 723–725 (1993).
16. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34,** 171–175 (1971).
17. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3,** 317–330 (1993).
18. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6,** 149–172 (1996).
19. Huxter, J., Burgess, N. & O'Keefe, J. Independent rate and temporal coding in hippocampal pyramidal cells. *Nature* **425,** 828–832 (2003).
20. Debanne, D., Gahwiler, B. H. & Thompson, S. M. Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *J. Physiol. (Lond.)* **507,** 237–247 (1998).
21. Lisman, J. E. Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron* **22,** 233–242 (1999).
22. Pedemonte, M., Pena, J. L. & Velluti, R. A. Firing of inferior colliculus auditory neurons is phase-locked to the hippocampus theta rhythm during paradoxical sleep and waking. *Exp. Brain Res.* **112,** 41–46 (1996).
23. Alonso, A. & Garcia-Austt, E. Neuronal sources of theta rhythm in the entorhinal cortex of the rat. II. Phase relations between unit discharges and theta field potentials. *Exp. Brain Res.* **67,** 502–509 (1987).
24. Siapas, A. G., Lubenov, E. V. & Wilson, M. A. Prefrontal phase locking to hippocampal theta oscillations. *Neuron* **46,** 141–151 (2005).
25. Green, J. D., Maxwell, D. S., Schindler, W. J. & Stumpf, C. Rabbit EEG 'theta' rhythm: its anatomical source and relation to activity in single neurons. *J. Neurophysiol.* **23,** 403–420 (1960).
26. Petsche, H. & Stumpf, C. Topographic and toposcopic study of origin and spread of the regular synchronized arousal pattern in the rabbit. *Electroencephalogr. Clin. Neurophysiol.* **12,** 589–600 (1960).
27. Winson, J. Patterns of hippocampal theta rhythm in the freely moving rat. *Electroencephalogr. Clin. Neurophysiol.* **36,** 291–301 (1974).
28. Bland, B. H., Anderson, P. & Ganes, T. Two generators of hippocampal theta activity in rabbits. *Brain Res.* **94,** 199–218 (1975).
29. Leung, L. W. Model of gradual phase shift of theta rhythm in the rat. *J. Neurophysiol.* **52,** 1051–1065 (1984).
30. Buzsaki, G., Czopf, J., Kondakor, I. & Kellenyi, L. Laminar distribution of hippocampal rhythmic slow activity (RSA) in the behaving rat: current-source density analysis, effects of urethane and atropine. *Brain Res.* **365,** 125–137 (1986).
31. Kocsis, B., Bragin, A. & Buzsaki, G. Interdependence of multiple theta generators in the hippocampus: a partial coherence analysis. *J. Neurosci.* **19,** 6200–6212 (1999).
32. Buzsaki, G. Theta oscillations in the hippocampus. *Neuron* **33,** 325–340 (2002).
33. Bullock, T. H., Buzsaki, G. & McClune, M. C. Coherence of compound field potentials reveals discontinuities in the CA1-subiculum of the hippocampus in freely-moving rats. *Neuroscience* **38,** 609–619 (1990).
34. Bragin, A. *et al.* Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* **15,** 47–60 (1995).
35. Amaral, D. G. & Witter, M. P. in *The Rat Nervous System* (ed. Paxinos G.) Ch. 21, 443–493 (Academic, 2004).

36. van Groen, T. & Wyss, J. M. Extrinsic projections from area CA1 of the rat hippocampus: olfactory, cortical, subcortical, and bilateral hippocampal formation projections. *J. Comp. Neurol.* **302,** 515–528 (1990).
37. Cenquizca, L. A. & Swanson, L. W. Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. *Brain Res. Rev.* **56,** 1–26 (2007).
38. Jung, M. W., Wiener, S. I. & McNaughton, B. L. Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J. Neurosci.* **14,** 7347–7356 (1994).
39. Maurer, A. P., Vanrhoads, S. R., Sutherland, G. R., Lipa, P. & McNaughton, B. L. Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. *Hippocampus* **15,** 841–852 (2005).
40. Kjelstrup, K. B. *et al.* Finite scale of spatial representation in the hippocampus. *Science* **321,** 140–143 (2008).
41. Huxter, J. R., Senior, T. J., Allen, K. & Csicsvari, J. Theta phase-specific codes for two-dimensional position, trajectory and heading in the hippocampus. *Nature Neurosci.* **11,** 587–594 (2008).
42. Ermentrout, G. B. & Kleinfeld, D. Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron* **29,** 33–44 (2001).
43. Miles, R., Traub, R. D. & Wong, R. K. Spread of synchronous firing in longitudinal slices from the CA3 region of the hippocampus. *J. Neurophysiol.* **60,** 1481–1496 (1988).
44. Traub, R. D. & Miles, R. *Neuronal Networks of the Hippocampus* Ch. 6, 119–156 (Cambridge Univ. Press, 1991).
45. Alonso, A. & Llinas, R. R. Subthreshold Na^+-dependent theta-like rhythmicity in stellate cells of entorhinal cortex layer II. *Nature* **342,** 175–177 (1989).
46. Haas, J. S. & White, J. A. Frequency selectivity of layer II stellate cells in the medial entorhinal cortex. *J. Neurophysiol.* **88,** 2422–2429 (2002).
47. Giocomo, L. M., Zilli, E. A., Fransen, E. & Hasselmo, M. E. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science* **315,** 1719–1722 (2007).
48. Leung, L. S. & Yu, H.-W. Theta-frequency resonance in hippocampal CA1 neurons *in vitro* demonstrated by sinusoidal current injection. *J. Neurophysiol.* **79,** 1592–1596 (1998).
49. Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O. & Somogyi, P. Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* **378,** 75–78 (1995).
50. Miller, R. *Cortico-Hippocampal Interplay and the Representation of Contexts in the Brain* Ch. 9, 159–188 (Springer-Verlag, 1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Wierzynski, M. Gu, G. Laurent, E. Schuman and A. Tolias for critical discussions and comments on the manuscript. We also thank the Caltech Brain Imaging Center and D. Procissi for the MRI brain scans. This work was supported by the Caltech Information Science and Technology Center for Biological Circuit Design, a 21st Century McDonnell Foundation Award, the Bren Foundation, and the McKnight Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.G.S. (thanos@caltech.edu) or E.V.L. (lubenov@caltech.edu).

Vol 459|28 May 2009|doi:10.1038/nature08007

LETTERS

Broad line emission from iron K- and L-shell transitions in the active galaxy 1H 0707-495

A. C. Fabian[1], A. Zoghbi[1], R. R. Ross[2], P. Uttley[3], L. C. Gallo[4], W. N. Brandt[5], A. J. Blustin[1], T. Boller[6], M. D. Caballero-Garcia[1], J. Larsson[1], J. M. Miller[7], G. Miniutti[8], G. Ponti[9], R. C. Reis[1], C. S. Reynolds[10], Y. Tanaka[6] & A. J. Young[11]

Since the 1995 discovery of the broad iron K-line emission from the Seyfert galaxy MCG–6-30-15 (ref. 1), broad iron K lines have been found in emission from several other Seyfert galaxies[2], from accreting stellar-mass black holes[3] and even from accreting neutron stars[4]. The iron K line is prominent in the reflection spectrum[5,6] created by the hard-X-ray continuum irradiating dense accreting matter. Relativistic distortion[7] of the line makes it sensitive to the strong gravity and spin of the black hole[8]. The accompanying iron L-line emission should be detectable when the iron abundance is high. Here we report the presence of both iron K and iron L emission in the spectrum of the narrow-line Seyfert 1 galaxy[9] 1H 0707-495. The bright iron L emission has enabled us to detect a reverberation lag of about 30 s between the direct X-ray continuum and its reflection from matter falling into the black hole. The observed reverberation timescale is comparable to the light-crossing time of the innermost radii around a supermassive black hole. The combination of spectral and timing data on 1H 0707-495 provides strong evidence that we are witnessing emission from matter within a gravitational radius, or a fraction of a light minute, from the event horizon of a rapidly spinning, massive black hole.

The galaxy 1H 0707-495 has been observed several times by the European Space Agency's X-ray Multi-Mirror Mission (XMM-Newton)[10–12]. The first observation revealed a sharp and deep spectral drop at 7 keV in the rest frame (the source redshift is 0.041) but no narrow emission features. This led to two main interpretations[10]: either the source is partly obscured by a large column of iron-rich material or it has very strong X-ray reflection[11] in its innermost regions where relativistic effects modify the observed spectrum. In other words, the sharp drop is due either to a photoelectric absorption edge or to the blue wing (higher-energy side) of a line partly shaped by relativistic Doppler shifts. The absorption origin requires that the iron abundance be about 30 times the solar value (unless the spectrum is e-folding below 10 keV (ref. 13), in which case the value is reduced), whereas reflection requires that this factor be between 5 and 10. The extreme variability of the source moreover appears to be due to changes mostly in the intensity of the power-law continuum, above 1 keV, which does not make sense in a partial-covering model.

We have analysed the spectral variability of new observations of the source made with XMM-Newton in January 2008. Variations from 1 to 12 counts per second are seen over four XMM-Newton orbits. The difference spectrum between low- and high-flux states is well fitted by a power-law continuum with photon index of 3, with an excess at lower energies (below 1.1 keV). Strong skewed and broad residuals are seen above such a power-law fit to the full source spectrum, peaking around 0.9 and 6.7 keV. The data are well described by a simple phenomenological model consisting of a power-law continuum, a soft black body, two relativistically broad (Laor[14]) lines and Galactic absorption (corresponding to a hydrogen column density of $N_{\rm H} = 5 \times 10^{20}\,{\rm cm}^{-2}$). We show the ratio of the spectrum and the continuum in that model in Fig. 1. This ratio spectrum is clearly dominated by two strong broad emission lines. They are characterized by energies of 0.89 and 6.41 keV (in our frame), an innermost radius of $1.3r_{\rm g}$, an outermost radius of $400r_{\rm g}$, an emissivity index of 4 and an inclination of 55.7°. (Here $1r_{\rm g} = GM/c^2$, where G is the gravitational constant, M is the mass of the black hole and c is the speed of light.) The normalizations of the lines (in photon spectra) are in the ratio of 20:1, and their rest energies correspond well to ionized iron L and K emission, respectively. We have tried using a

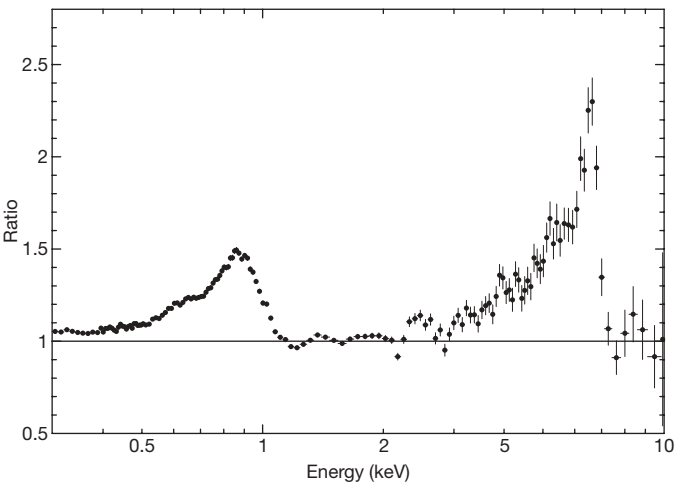


Figure 1 | Ratio of the observed spectrum and a simple phenomenological model. The model consists of a power law, a black body and two broad emission lines. The data are from all four XMM-Newton orbits and the normalizations of the broad lines have been set to zero to make the plot. Ionized iron L and K emissions peak in the rest frame around 0.9 keV and 6.5–6.7 keV with equivalent widths of 180 and 970 eV, respectively. Error bars, 1σ.

[1]Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK. [2]Physics Department, College of the Holy Cross, Worcester, Massachusetts 01610, USA. [3]School of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK. [4]Department of Astronomy and Physics, Saint Mary's University, Halifax, Nova Scotia B3H 3C3, Canada. [5]Department of Astronomy and Astrophysics, The Pennsylvania State University, 525 Davey Lab, University Park, Pennsylvania 16802, USA. [6]Max-Planck-Institut fur Extraterrestrische Physik, Postfach 1312, 85741 Garching, Germany. [7]Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA. [8]LAEX, Centro de Astrobiologia (CSIC–INTA), LAEFF, PO Box 78, Villanueva de la Cañada, Madrid E-28691, Spain. [9]Laboratoire APC, UMR 7164, 10 rue A. Domon et L. Duquet, 75205 Paris, France. [10]Department of Astronomy and the Center for Theory and Computation, University of Maryland, College Park, Maryland 20742, USA. [11]H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK.

variety of continuum models and always obtain similar results. The data have also been fitted with a self-consistent model reflection spectrum[15] (Fig. 2; see Supplementary Information for more details), which reproduces the correct relative fluxes for the lines. Iron is almost nine times more abundant than in the Sun, and the other elements have the solar abundance values. Perhaps a dense nuclear star cluster has led to the formation of massive white dwarf binaries that have enriched the nucleus with type Ia supernova ejecta rich in iron (such an explanation has been proposed for globular star clusters[16]).

The presence of broadened and skewed lines other than the iron K line is a prediction of the disk reflection model that we have now confirmed. Furthermore, we have shown that the relative strengths of the observed iron L and K lines agree well with predictions based on atomic physics. Whereas it is possible to construct a partial-covering, absorption-dominated model[10,12] for the prominent K-shell iron feature, the L-shell absorption edge of ionized iron required around 1 keV is accompanied by an unacceptably strong absorption feature around 0.75 keV from an unresolved transition array of iron ions Fe IX–Fe XI. Previous work[10,12] on absorption models for 1H 0707-495 have been unable to account for the spectral structure around 1 keV without invoking emission. The strong variability, and shape, of this emission component points to the inner regions of the flow and, thus, to a reflection solution.

The source fractional root-mean-square variability is roughly constant below 1 keV, increases abruptly at 1 keV and then drops back to the soft level above 4 keV, in agreement with the expectation of the two-component model (reflection plus power law) used to fit the data. Both components vary in amplitude, but the power law does so by nearly twice as much as the reflection. The nonlinear behaviour of the variability in accreting black holes[17] is considered to be due to the cumulative random effects of the orbital variations from many radii effectively multiplied together. On the shortest timescales, light-crossing effects will become important, as the light path for the direct primary radiation is shorter than that for the reflected radiation. This effect is seen for the first time in the frequency-dependent lags (Fig. 3). The large positive lag for variations lower in frequency than 0.6 mHz (timescales greater than 30 min) is probably due to the inward drift of accretion fluctuations through the emitting region[18], causing the density and, thus, the ionization state of the irradiated disc to respond first. Variations higher in frequency than 0.6 mHz show a negative lag in the sense that the soft reflection-dominated band follows the hard power-law-dominated band by about 30 s. This is in the opposite sense to a Comptonization lag produced by up-scattering of photons (or to any model where the spectral drop at 1 keV is instead produced by absorption that is responding to the changing continuum), and is explained by reverberation. If the lag time corresponds to the natural length of about hr_g, where we expect h to be between two and five, then we deduce a mass for the black hole of about $7 \times 10^6 h^{-1} M_\odot$ ($M_\odot$, solar mass), which is reasonable for this source (no definitive mass is known; see, for example, ref. 19) and implies that the accretion rate is only just below the Eddington limit. The breadth of the iron lines implies (using the methods of ref. 8) that the black hole has a high spin, of dimensionless spin parameter $a = cJ/GM^2 > 0.98$, so much of the emission should originate from within a few gravitational radii. Iron L-line emission should be detectable in similar sources[20] with high iron abundances, thereby enabling reverberation studies to be made.

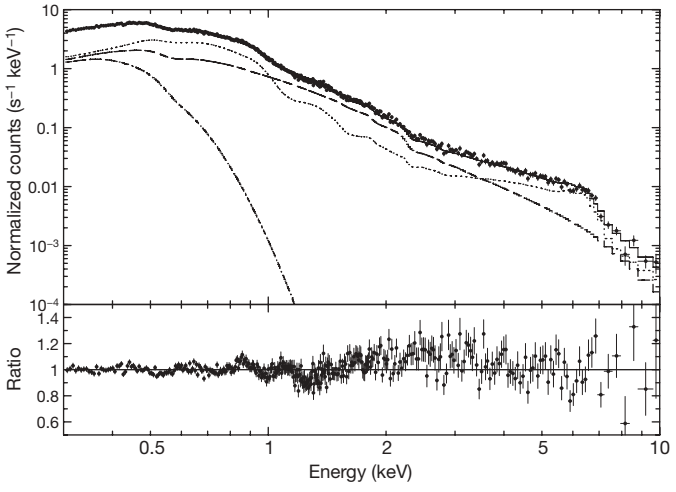


Figure 2 | Spectrum of the first orbit showing the best-fitting self-consistent, relativistically blurred reflection model. The ratio of the data and the model is shown in the lower panel. The total exposure time for this spectrum is 102 ks, with the source changing flux continuously and by more than a factor of two every few kiloseconds. The offset in the 1.5–4-keV band is due to the simplicity of using only two components in the 1–10-keV band despite the high variability of the source. The addition of a further reflection component with higher ionization parameter considerably reduces this offset. This is to be expected if the ionization changes with time or flux. The contributions of the power-law, reflection and black-body components are indicated by the dashed, dotted and dash–dot lines, respectively. Error bars, 1σ.

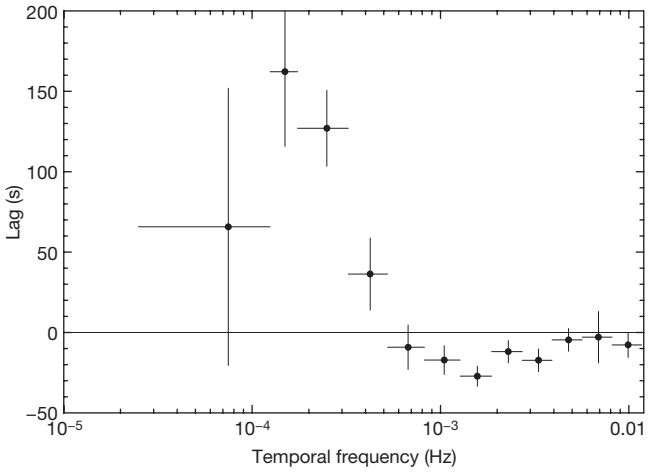


Figure 3 | Frequency-dependent lags between the 0.3–1-keV and 1–4-keV bands. A negative lag, such as that found above frequencies of 6×10^{-4} Hz or timescales shorter than 30 min, indicates that the harder flux, which is dominated by the power-law continuum, changes before the softer flux, which is dominated by reflection (particularly the iron L line). Error bars, 1σ.

Received 9 December 2008; accepted 13 March 2009.

1. Tanaka, Y. *et al.* Gravitationally redshifted emission implying an accretion disk and massive black hole in the active galaxy MCG–6–30–15. *Nature* **375,** 659–661 (1995).
2. Nandra, K. *et al.* An XMM-Newton survey of broad iron lines in Seyfert galaxies. *Mon. Not. R. Astron. Soc.* **382,** 194–228 (2008).
3. Miller, J. M. Relativistic X-Ray lines from the inner accretion disks around black holes. *Annu. Rev. Astron. Astrophys.* **45,** 441–479 (2007).
4. Cackett, E. M. *et al.* Relativistic iron emission lines in neutron star low-mass X-ray binaries as probes of neutron star radii. *Astrophys. J.* **674,** 415–420 (2008).
5. Guilbert, P. W. & Rees, M. J. 'Cold' material in non-thermal sources. *Mon. Not. R. Astron. Soc.* **233,** 475–484 (1988).
6. Ross, R. R. & Fabian, A. C. The effects of photoionization on X-ray reflection spectra in active galactic nuclei. *Mon. Not. R. Astron. Soc.* **261,** 74–82 (1993).
7. Fabian, A. C. *et al.* X-ray fluorescence from the inner disc in Cygnus X-1. *Mon. Not. R. Astron. Soc.* **238,** 729–736 (1989).
8. Brenneman, L. W. & Reynolds, C. S. Constraining black hole spin via X-ray spectroscopy. *Astrophys. J.* **652,** 1028–1043 (2006).
9. Osterbrock, D. E. & Pogge, R. W. The spectra of narrow-line Seyfert 1 galaxies. *Astrophys. J.* **297,** 166–176 (1985).
10. Boller, Th *et al.* XMM-Newton discovery of a sharp spectral feature at 7 keV in the narrow-line Seyfert 1 galaxy 1H 0707–49. *Mon. Not. R. Astron. Soc.* **329,** L1–L5 (2002).
11. Fabian, A. C. *et al.* X-ray reflection in the narrow-line Seyfert 1 galaxy 1H 0707–495. *Mon. Not. R. Astron. Soc.* **353,** 1071–1077 (2004).
12. Gallo, L. C. *et al.* Long-term spectral changes in the partial-covering candidate narrow-line Seyfert 1 galaxy 1H 0707–495. *Mon. Not. R. Astron. Soc.* **353,** 1064–1070 (2004).
13. Tanaka, Y. *et al.* Partial covering interpretation of the X-ray spectrum of the NLS1 1H 0707–495. *Publ. Astron. Soc. Jpn* **56,** L9–L13 (2004).

14. Laor, A. Line profiles from a disk around a rotating black hole. *Astrophys. J.* **376,** 90–94 (1991).
15. Ross, R. R. & Fabian, A. C. A comprehensive range of X-ray ionized-reflection models. *Mon. Not. R. Astron. Soc.* **358,** 211–216 (2005).
16. Shara, M. M. & Hurley, J. R. Star clusters as type Ia supernova factories. *Astrophys. J.* **571,** 830–842 (2002).
17. Uttley, P., McHardy, I. M. & Vaughan, S. Non-linear X-ray variability in X-ray binaries and active galaxies. *Mon. Not. R. Astron. Soc.* **359,** 345–362 (2005).
18. Arevalo, P. & Uttley, P. Investigating a fluctuating-accretion model for the spectral-timing properties of accreting black hole systems. *Mon. Not. R. Astron. Soc.* **367,** 801–814 (2006).
19. Bian, W. & Zhao, Y. On X-ray variability in narrow line and broad line active galactic nuclei. *Mon. Not. R. Astron. Soc.* **343,** 164–168 (2003).
20. Gallo, L. Investigating the nature of narrow-line Seyfert 1 galaxies with high-energy spectral complexity. *Mon. Not. R. Astron. Soc.* **368,** 479–486 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.C.F. thanks the Royal Society for support. This work is based on observations made with XMM-Newton, a European Space Agency (ESA) science mission with instruments and contributions directly funded by ESA member states and the USA (NASA). A.Z. acknowledges the support of the Algerian Higher Education Ministry and the UK Science and Technology Facilities Council (STFC). C.S.R., W.N.B., G.M., G.P. and R.C.R., and A.J.B. acknowledge support from the US National Science Foundation, NASA, the Spanish Ministerio de Ciencia e Innovacion, the French National Research Agency and the UK STFC for support, respectively.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.C.F. (acf@ast.cam.ac.uk).

LETTERS

The changing phases of extrasolar planet CoRoT-1b

Ignas A. G. Snellen[1], Ernst J. W. de Mooij[1] & Simon Albrecht[1]

Hot Jupiters are a class of extrasolar planet that orbit their parent stars at very short distances. They are expected to be tidally locked, which can lead to a large temperature difference between their daysides and nightsides. Infrared observations of eclipsing systems have yielded dayside temperatures for a number of transiting planets[1–5]. The day–night contrast of the transiting extrasolar planet HD 189733b was 'mapped' using infrared observations[6,7]. It is expected that the contrast between the daysides and nightsides of hot Jupiters is much higher at visual wavelengths, shorter than that of the peak emission, and could be further enhanced by reflected stellar light. Here we report the analysis of optical photometric data[8] obtained over 36 planetary orbits of the transiting hot Jupiter CoRoT-1b. The data are consistent with the nightside hemisphere of the planet being entirely black, with the dayside flux dominating the optical phase curve. This means that at optical wavelengths the planet's phase variation is just as we see it for the interior planets in the Solar System. The data allow for only a small fraction of reflected light, corresponding to a geometric albedo of <0.20.

The Convection, Rotation and Planetary Transit (CoRoT) satellite monitored the extrasolar planet CoRoT-1b nearly continuously for 55 days, among ~12,000 other stars in its fields of view[9]. The time sampling was 512 s during the first 30 days and 32 s for the remainder of the observations, providing a light curve with nearly 69,000 data points covering 36 planetary orbital periods ($P = 1.509$ d). A prism in front of the extrasolar-planet CCDs (charge-coupled devices) produces small spectra for each star, on which aperture photometry is performed in three bands (red, green and blue)[8]. This is done on board to comply with the available telemetry volume. The transmission curves of the three bands are different for each targeted star. They depend on the template chosen for the on-board aperture photometry, which is based on the effective temperature of the star and its position on the CCD. The Earth has a significant influence on the photometric performance of the satellite and introduces relevant perturbations on timescales of the satellite orbital period (103 min) and the 24-h day. Most of these effects were corrected for before the data release for the general astronomy community[10].

For a detailed description of the data analysis, we refer the reader to the Supplementary Information. Concentrating on the data from the red-channel passband, we rejected outlier data points and removed residual instrumental effects on the timescale of the orbital period of the satellite and the 24-h day. The final, corrected and partly resampled light curve contains 7,883 data points with a relative standard deviation of 1.0×10^{-3} and correlated noise estimated at a level of $\sim 1.2 \times 10^{-4}$, which decreases significantly when the signal is averaged over 34 transits. In strong contrast to the red-channel data, the light curves from the green and blue channels unfortunately exhibit ramps, sudden jumps and high levels of correlated noise, which make them unusable for the analysis carried out here. The short-wavelength cut-off of the red-channel passband was determined from the overall transmission curve for the telescope–CCD combination[8] multiplied with a Kurucz model spectrum[11] of the host star, which was compared with the fraction of photons collected in the red channel. This results in a wavelength cut-off of 560 nm and an effective wavelength of 710 nm.

The final, corrected light curve (Fig. 1), folded over the orbital period of the planet and binned in phase over intervals of 0.05, shows a distinct rise in flux over the first half of the orbit, followed by a dip at an orbital phase of 0.5 and a significant decrease during the second half of the orbit. This is entirely consistent with the dayside hemisphere rotating into view, being eclipsed by the star and rotating out

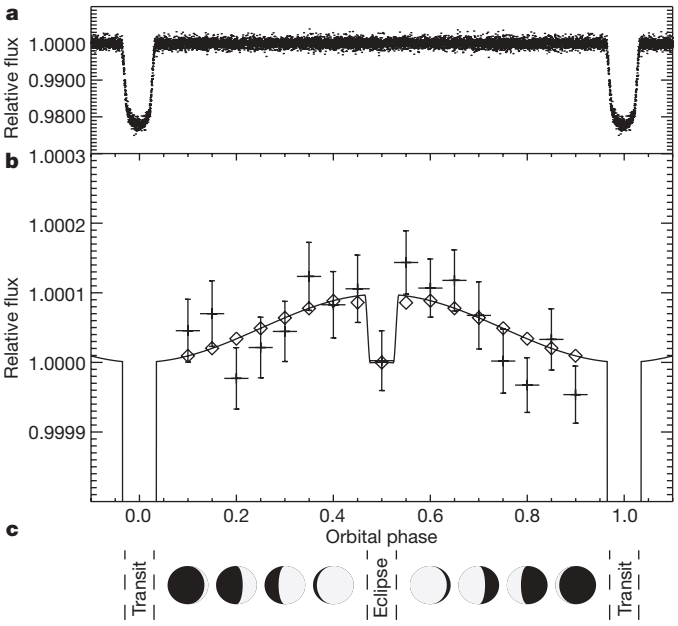


Figure 1 | Optical phase variation for CoRoT-1b centred on the planetary eclipse. Background-subtracted photometry is shown from 55 days of CoRoT monitoring in its red channel, after the rejection of $>3\sigma$ outliers and corrections for perturbations on the 103-min orbital timescale of the satellite and the 24-h day. **a**, **b**, The same data is shown in each panel, phase-folded over the planetary orbital period, $P = 1.5089557$ d. In **b**, the data are binned in phase over intervals of 0.05 (showing the 1σ error bars as determined from the scatter in the individual data points), and the scale of the y axis is magnified by a factor of ~200. The data are consistent with the dayside hemisphere rotating into view, being eclipsed by the star and rotating out of view again (**c**). The unbinned data are fitted with a model assuming uniform surface brightnesses for the dayside and nightside hemispheres; this is indicated by the solid curve. The ratio of integrated flux from the dayside hemisphere to stellar flux, as determined from the planetary eclipse depth and the phase variation, is found to be $(1.26 \pm 0.33) \times 10^{-4}$. The ratio of the flux from the nightside hemisphere to stellar flux, as determined from the difference between the secondary-eclipse depth and the amplitude of the phase variation, is found to be $<3.0 \times 10^{-5}$ at 1σ ($<5.9 \times 10^{-5}$ at 2σ) and is consistent with the hemisphere being entirely black. This means that the phase variation is just as we see it for the interior planets in the Solar System. **c**, Sketch of the phase variation.

[1]Leiden Observatory, Leiden University, Postbus 9513, 2300 RA Leiden, The Netherlands.

of view again. The amount of light received from the system at the moment the planet is eclipsed is similar to that when the nightside of the planet is in full view. This means that to within the observational uncertainties we receive no light from the nightside of the planet.

We fitted the light curve with a three-parameter model using a chi-squared analysis, assuming a homogeneous surface brightness for each of the hemispheres. The first parameter, R_{Day}, is the contrast ratio between the planet dayside flux and the stellar flux. The second parameter, $F_{N/D}$, denotes the ratio of nightside to dayside flux. A third parameter represents the flux of the star. We determine the two relevant parameters to be $R_{Day} = (1.26 \pm 0.33) \times 10^{-4}$ (with a null detection rejected at the $\sim 4\sigma$ confidence level) and $F_{N/D} = 0$ (a best-fit value), with upper limits of $F_{N/D} < 0.24$ at 1σ and $F_{N/D} < 0.47$ at 2σ, meaning that the integrated light from the nightside hemisphere is $<24\%$ (1σ) or $<47\%$ (2σ) of that from the dayside hemisphere. The influence of correlated noise was assessed by also fitting the model to a light curve with 1-h bins, for which the error of each binned point was calculated from the variation of the points at the same orbital phase over all periods. In this way, R_{Day} was determined to be slightly higher, $(1.47 \pm 0.40) \times 10^{-4}$, at a lower significance. In addition, we performed a Markov-chain Monte Carlo simulation, which yielded $R_{Day} = (1.40 \pm 0.33) \times 10^{-4}$ (Supplementary Information). Both R_{Day} and $F_{N/D}$ are influenced by the phase variation and the eclipse depth, but in different ways. Whereas the eclipse depth is a direct measure of R_{Day}, and the phase variation places a lower limit on it, it is the ratio of phase variation over the eclipse depth that governs $F_{N/D}$. The entire phase curve contributes to the estimates of R_{Day} and $F_{N/D}$, including the points near the transit.

The high level of irradiation from the nearby host star is a major factor in determining the atmospheric properties of a hot Jupiter. There is mounting evidence, both theoretical and observational, that differences in incident star flux between planets lead to at least two distinct classes of hot-Jupiter atmospheres, depending on whether or not their atmospheres contain highly absorbing substances such as gaseous titanium oxide and vanadium oxide (refs 12–19). These hot Jupiters are called 'pM-' and 'pL-class' planets[12], analogous to the M- and L-type stellar dwarfs, and the two classes are expected to have very different spectra and dayside-to-nightside circulations.

The pM-class planets, with high levels of irradiation, are thought to be warm enough to prevent condensation of titanium- and vanadium-bearing compounds. This leads to absorption of incident flux by titanium oxide and vanadium oxide at low pressure and, subsequently, a temperature inversion in the planets' stratospheres. These planets are expected to appear anomalously bright in the infrared, and to exhibit molecular bands in emission rather than in absorption[12–15]. Broadband infrared secondary-eclipse measurements for HD 209458b, a pM-class planet thought to be just above the pM–pL boundary, are indeed best explained by the presence of thermal inversion and water emission bands[14,19]. In addition, the broadband infrared spectrum of the even warmer planet TrES-4b is also best fitted with models assuming a temperature inversion in its atmosphere[17].

In the atmospheres of planets that receive less stellar flux (pL-class planets such as HD 189733b), titanium and vanadium are expected to be condensed out; the atmospheres therefore should not exhibit thermal inversion. The recent infrared spectrum of HD 189733b indeed shows strong water absorption and is best matched with models that do not include an atmospheric temperature inversion[16].

Whether or not the incident stellar flux is absorbed in a planet's stratosphere strongly influences the ratio of radiative timescales to expected dynamical timescales, and determines to what extent the absorbed energy is redistributed to the planet's nightside[12–14]. The cooler, pL-class, planets absorb incident flux deep in the atmosphere where the atmospheric dynamics are more likely to redistribute absorbed energy, leading to cooler daysides, warmer nightsides and strong jet flows resulting in significant phase shifts in their thermal emission light curves. Indeed, the pL-class planet HD 189733b is observed to have dayside–nightside temperature differences of ~ 240 K at both 8 and 24 μm, accompanied by a phase shift of 20–30° (refs 6, 7). For pM-class planets, the absorbed energy is reradiated before it can be transported to the nightside, resulting in large dayside–nightside temperature contrasts and negligible phase shifts in their thermal emission light curves. Large day–night contrasts have been found for two pM-class planets, υ Andromedae b[20] and HD 179949b[18], although a detailed interpretation is hampered because both systems are non-transiting and their orbital inclinations and planet radii are unknown.

CoRoT-1b is a strongly irradiated planet and should therefore fall well within the pM class. The measured difference between the dayside and nightside fluxes, each relative to the stellar flux, is $(1.26 \pm 0.36) \times 10^{-4}$, meaning that if there is no reflective component in the red-channel light curve, the redistribution fraction, P_n, which is the fraction of absorbed stellar radiation that is transported to the nightside of the planet[21], has an upper limit of $P_n < 0.22$ at 2σ ($P_n < 0.39$ at 3σ), in line with the low distribution fractions expected for pM-class planets. In addition, there is no evidence for a phase shift in the light curve. Therefore, if CoRoT-1b has a low albedo, it exhibits all the characteristics expected of a pM-class planet. We measure its hemisphere-averaged dayside brightness temperature to be 2390 ± 90 K. Assuming a uniform hemispheric emission, the maximum possible brightness temperature is slightly lower, at $\sim 2{,}260$ K

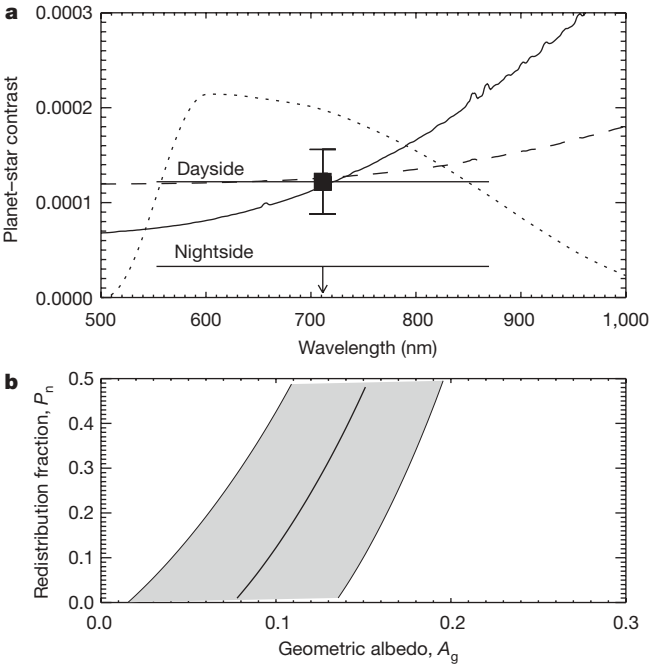


Figure 2 | The optical planet–star contrast compared with models. a, The planet–star contrast as determined from the phase variation and the planetary eclipse for the dayside and nightside of CoRoT-1b (with 1σ error bars), together with the arbitrarily scaled passband of the red-channel data (dotted line). **b**, The allowed ranges for A_g, the geometric albedo for reflected light, and for P_n, the redistribution fraction that indicates how much of absorbed stellar radiation is transported to the nightside of the planet. Here we assume wavelength-independent Lambert scattering. To obtain the planet–star contrast expected for a given combination of A_g and P_n, first the Planck curve for the planet and the Kurucz model spectrum for the primary star were separately multiplied by the transmission function of the red-channel passband and integrated over wavelength. These integrated fluxes were multiplied by their respective surface areas and the ratio was taken. This was subsequently added to a possible reflective component. The solid and dashed lines in **a** show the two most extreme cases that perfectly fit the measured planet–star contrast: $P_n = 0.0$, $A_g = 0.08$ and $P_n = 0.5$, $A_g = 0.15$. Within the 1σ uncertainty in the dayside planet–star contrast, the geometric albedo is constrained to within the range $0.02 < A_g < 0.20$. Assuming that there is no reflected light component, the 3σ limit on the redistribution fraction is $P_n < 0.39$.

for a non-reflective planet with a redistribution fraction of $P_n = 0$. If instantaneous re-emission of absorbed radiation is assumed without advection, a maximum dayside brightness temperature of ~2,430 K is obtained, within the 1σ uncertainty limits of the measured temperature. Because this indicates that any reflective component in the planet's light curve is probably small, it means that the planet has a very low geometric albedo in CoRoT's red channel.

Both theoretical modelling and observed upper limits do imply very low reflectivity for hot Jupiters. Ground-based spectroscopy that exploits the Doppler effect to separate the spectral lines of a hot Jupiter from the lines of its parent star has yielded stringent upper limits for geometric albedos (for example $A_g < 0.12$ at 3σ for HD 75289b)[22–24]. In addition, analysis of data from the MOST satellite has yielded an upper limit of $A_g < 0.17$ at 3σ for HD 209458b[25]. In comparison, the geometric albedos of Solar System gas giant planets range from 0.41 to 0.52 (ref. 26). Theoretical modelling of extrasolar-planet atmospheres shows that many parameters can cause the low albedo of hot Jupiters, in particular the strong absorption of the alkali metals sodium and potassium (and/or the aforementioned absorption of titanium oxide and vanadium oxide), and the sizes and types of condensates in the atmospheres. In the absence of clouds, the low albedo could be due to atomic or molecular absorption[27–29].

The red-channel light curve of CoRoT-1b can also be fitted with the albedo and redistribution fraction allowed to vary freely. For this, we assumed that the geometric albedo is independent of wavelength (λ) and is related to the Bond albedo by $A_g(\lambda) = 2A_B/3$, as for a diffusely scattering (Lambert) sphere. We found that in the context of this model the variation in planet–star contrast can be explained by a geometric albedo of 0.02 to 0.2 for the full range of possible redistribution fractions (Fig. 2). However, assuming that the planet's albedo is at the low end of this range, in line with both theoretical modelling and observations of other hot Jupiters, the day–night temperature contrast of CoRoT-1b is high and the redistribution fraction is low, as expected for a highly irradiated planet.

This year we celebrate the 400th anniversary of the first published astronomical observations with a telescope, by Galileo Galilei. Galilei used his telescope to observe the changing phases of Venus and reveal the true configuration of the Solar System. Now, exactly four centuries later, CoRoT observations have shown the changing phases of an extrasolar planet for the first time in optical light.

Received 28 January; accepted 6 April 2009.

1. Charbonneau, D. *et al.* Detection of thermal emission from an extrasolar planet. *Astrophys. J.* **626,** 523–529 (2005).
2. Deming, D., Seager, S., Richardson, L. J. & Harrington, J. Infrared radiation from an extrasolar planet. *Nature* **434,** 740–743 (2005).
3. Harrington, J., Luszcz, S., Seager, S., Deming, D. & Richardson, L. The hottest planet. *Nature* **447,** 691–693 (2007).
4. Sing, D. K. & Lopez-Morales, M. Ground-based secondary eclipse detection of the very-hot Jupiter OGLE-TR-56b. *Astron. Astrophys.* **493,** L31–L34 (2009).
5. de Mooij, E. J. W. & Snellen, I. A. G. Ground-based K-band detection of thermal emission from the exoplanet TrES-3b. *Astron. Astrophys.* **493,** L35–L38 (2009).
6. Knutson, H. A. *et al.* A map of the day–night contrast of the extrasolar planet HD 189733b. *Nature* **447,** 183–186 (2007).
7. Knutson, H. *et al.* Multiwavelength constraints on the day-night circulation patterns of HD 189733b. *Astrophys. J.* **690,** 822–836 (2009).
8. Auvergne, M. Bodin, P. Boisnard, L., Buey, J., Chaintreuil, S. & CoRoT team. The CoRoT satellite in flight: description and performance. *Astron. Astrophys*. (in the press); preprint at ⟨http://arxiv.org/abs/0901.2206⟩ (2009).
9. Barge, P. *et al.* Transiting exoplanets from the CoRoT space mission. I. CoRoT-Exo-1b: a low-density short-period planet around a G0V star. *Astron. Astrophys.* **482,** L17–L20 (2008).
10. Samadi, R. *et al.* Extraction of the photometric information: corrections. Preprint at ⟨http://arXiv.org/astro-ph/0703354⟩ (2007).
11. Kurucz, R. *ATLAS9 Stellar Atmosphere Programs and 2 km/s Grid.* (Kurucz CD-ROM No. 13, Smithsonian Astrophysical Observatory, 1993).
12. Fortney, J., Lodders, K., Marley, M. & Freedman, R. A unified theory for the atmospheres of the hot and very hot Jupiters: two classes of irradiated atmospheres. *Astrophys. J.* **678,** 1419–1435 (2008).
13. Cooper, C. & Showman, A. Dynamic meteorology at the photosphere of HD 209458b. *Astrophys. J.* **629,** L45–L48 (2005).
14. Showman, A. *et al.* Atmospheric circulation of hot Jupiters: coupled radiative-dynamical general circulation model simulations of HD 189733b and HD 209458b. *Astrophys. J.* (submitted); preprint at ⟨http://arXiv.org/abs/0809.2089⟩ (2008).
15. Burrows, A., Hubeny, I., Budaj, J., Knutson, H. & Charbonneau, D. Theoretical spectral models of the planet HD 209458b with a thermal inversion and water emission bands. *Astrophys. J.* **668,** L171–L174 (2007).
16. Grillmair, C. *et al.* Strong water absorption in the dayside emission spectrum of the planet HD 189733b. *Nature* **456,** 767–769 (2008).
17. Knutson, H., Charbonneau, D., Burrows, A., O'Donovan, F. & Mandushev, G. Detection of a temperature inversion in the broadband infrared emission spectrum of TrES-4. *Astrophys. J.* **691,** 866–874 (2009).
18. Cowan, N., Agol, E. & Charbonneau, D. Hot nights on extrasolar planets: mid-infrared phase variations of hot Jupiters. *Mon. Not. R. Astron. Soc.* **379,** 641–646 (2007).
19. Knutson, H., Charbonneau, D. & Allen, L. The 3.6–8.0 μm broadband emission spectrum of HD 209458b: evidence for an atmospheric temperature inversion. *Astrophys. J.* **673,** 526–531 (2008).
20. Harrington, J. *et al.* The phase-dependent infrared brightness of the extrasolar planet υ Andromedae b. *Science* **314,** 623–626 (2006).
21. Burrows, A., Budaj, J. & Hubeny, I. Theoretical spectra and light curves of close-in extrasolar giant planets and comparison with data. *Astrophys. J.* **678,** 1436–1457 (2008).
22. Collier Cameron, A., Horne, K., Penny, A. & Leigh, C. A search for starlight reflected from υ And's innermost planet. *Mon. Not. R. Astron. Soc.* **330,** 187–204 (2002).
23. Leigh, C. *et al.* A search for starlight reflected from HD 75289b. *Mon. Not. R. Astron. Soc.* **346,** L16–L20 (2003).
24. Leigh, C., Collier Cameron, A., Horne, K., Penny, A., &. James, D. A new upper limit on the reflected starlight from τ Bootis b. *Mon. Not. R. Astron. Soc.* **344,** 1271–1282 (2003).
25. Rowe, J. F. *et al.* The very low albedo of an extrasolar planet: MOST space-based photometry of HD 209458. *Astrophys. J.* **689,** 1345–1353 (2008).
26. Cox, A. N. in *Allen's Astrophysical Quantities* 4th edn (ed. Cox, A. N.) Ch. 1 (Athlone, 2000).
27. Seager, S., Whitney, B. A. & Sasselov, D. D. Photometric light curves and polarization of close-in extrasolar giant planets. *Astrophys. J.* **540,** 504–520 (2000).
28. Green, D., Matthews, J., Seager, S. & Kuschnig, R. Scattered light from close-in extrasolar planets: prospects of detection with the MOST satellite. *Astrophys. J.* **597,** 590–601 (2003).
29. Hood, B., Wood, K., Seager, S. & Collier Cameron, A. Reflected light from three dimensional exoplanetary atmospheres and simulations of HD 209458b. *Mon. Not. R. Astron. Soc.* **389,** 257–269 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the CoRoT team for making the CoRoT data, which forms the basis of this study, publicly available in a high-quality and comprehensible way. The CoRoT space mission, launched on 27 December 2006, was developed and is operated by the Centre National D'Études Spatial, with participation of the science programmes of the European Space Agency, the European Space Research and Technology Centre and the Research and Scientific Support Department, Austria, Belgium, Brazil, Germany and Spain. We thank R. Le Poole for discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.A.G.S. (snellen@strw.leidenuniv.nl).

Vol 459|28 May 2009|doi:10.1038/nature08005

LETTERS

Synthesizing arbitrary quantum states in a superconducting resonator

Max Hofheinz[1], H. Wang[1], M. Ansmann[1], Radoslaw C. Bialczak[1], Erik Lucero[1], M. Neeley[1], A. D. O'Connell[1], D. Sank[1], J. Wenner[1], John M. Martinis[1] & A. N. Cleland[1]

The superposition principle is a fundamental tenet of quantum mechanics. It allows a quantum system to be 'in two places at the same time', because the quantum state of a physical system can simultaneously include measurably different physical states. The preparation and use of such superposed states forms the basis of quantum computation and simulation[1]. The creation of complex superpositions in harmonic systems (such as the motional state of trapped ions[2], microwave resonators[3–5] or optical cavities[6]) has presented a significant challenge because it cannot be achieved with classical control signals. Here we demonstrate the preparation and measurement of arbitrary quantum states in an electromagnetic resonator, superposing states with different numbers of photons in a completely controlled and deterministic manner. We synthesize the states using a superconducting phase qubit to phase-coherently pump photons into the resonator, making use of an algorithm[7] that generalizes a previously demonstrated method of generating photon number (Fock) states in a resonator[8]. We completely characterize the resonator quantum state using Wigner tomography, which is equivalent to measuring the resonator's full density matrix.

The quantum state of a resonator is extraordinarily rich, with infinitely many energy levels, of which each can have a non-zero amplitude. However, this richness is difficult to access when driving a resonator with a classical signal, as the two adjustable parameters of an on-resonant drive, the amplitude and the phase, give very limited control. Creating an arbitrary quantum state instead requires a nonlinear element and a control scheme with many parameters. Here we demonstrate quantum state generation in a resonator by interposing a highly nonlinear Josephson phase qubit[9] between a superconducting resonator and a classical signal. A qubit[4,5,10–14] has two quantum degrees of freedom, the relative amplitude and phase of its ground $|g\rangle$ and excited $|e\rangle$ energy eigenstates. This simplicity allows full quantum control of a qubit with a classical signal[15]. By following a sequence of steps developed for trapped ions[2,7] (and later adapted to charge qubits[16]), where each step involves creating a particular qubit state and then having the qubit interact with the resonator for a controlled time, we synthesize arbitrary states in the resonator. The preparation is deterministic, unlike methods involving probabilistic projective measurements[17]. After the preparation, we analyse the resonator state using Wigner tomography[18–22], mapping out the Wigner quasi-probability distribution[23,24], from which we extract the resonator's full density matrix.

The quantum circuit we used is shown in Fig. 1a. The phase qubit is capacitively coupled to a superconducting coplanar waveguide resonator, and the circuit includes control lines for the qubit and resonator, and a qubit measurement circuit described elsewhere[25]. This circuit is similar to that used previously to generate Fock states in a resonator[8]; here, however, most of the superconducting wiring is made of rhenium in place of aluminium, and we removed unnecessary dielectric, reducing dissipative elements in the circuit.

The qubit frequency $\omega_q/2\pi$ can be externally adjusted, whereas the resonator frequency $\omega_r/2\pi = 6.570$ GHz is fixed. This allows us to

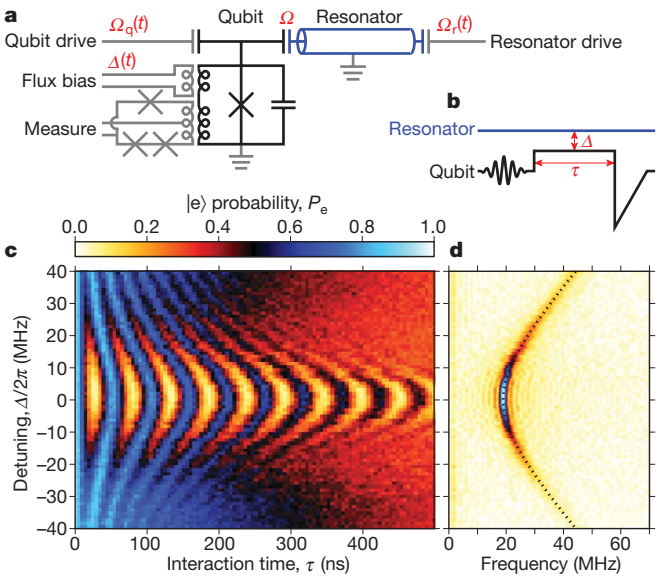


Figure 1 | Circuit diagram and one photon Rabi-swap oscillations between qubit and resonator. a, The qubit (black) is made from a Josephson junction (cross) and a capacitor, biased through a shunting inductor. The qubit detuning Δ is adjusted through a flux bias coil, and the qubit state is read out by a three-Josephson-junction superconducting quantum interference device (SQUID). The coplanar waveguide resonator (blue) has fixed capacitive coupling Ω to the qubit, and small capacitors couple external microwave signals Ω_q and Ω_r to the qubit and resonator. The device was measured in a dilution refrigerator at 25 mK. The qubit relaxation and dephasing times were respectively $T_{1,q} \approx 650$ ns and $T_{2,q} \approx 150$ ns, and the resonator relaxation time was $T_{1,r} \approx 3.5$ μs with no measurable dephasing. **b**, Schematic of Rabi-swap pulse sequence. The qubit starts in its ground state, detuned at its typical off-resonance point by $\Delta_{off}/2\pi = -463\,\text{MHz} \approx -25\Omega/2\pi$ from the resonator. A resonant qubit microwave π-pulse brings the qubit to its excited state $|e\rangle$, injecting one quantum of energy into the system. A flux bias pulse reduces the qubit detuning Δ from the resonator for a controlled time τ, and the qubit state is then measured with a current pulse. **c**, Excited state probability P_e versus detuning Δ and interaction time τ. P_e is obtained by averaging 600 repetitions. **d**, Fourier transform of data in **c**, showing the expected hyperbolic relation between detuning Δ and swap frequency $\sqrt{\Omega^2+\Delta^2}/2\pi$ (dotted line), and the expected fall-off in probability (colour scale). Resonance $\Delta = 0$ corresponds to lowest swap frequency and maximum probability amplitude.

[1]Department of Physics, University of California, Santa Barbara, California 93106, USA.

Table 1 | Sequence to generate the resonator state $|\psi\rangle = |1\rangle + i|3\rangle$

Sequence of states, operations	Operational parameter	System state, parameter value
$\|\psi\rangle$		$\|g\rangle(0.707\|1\rangle+0.707i\|3\rangle)$
S_3	$\tau_3\Omega$	1.81
Q_3	q_3	3.14
$\|\psi_2\rangle$		$\|g\rangle(-0.557i\|0\rangle+0.707\|2\rangle)+0.436\|e\rangle\|1\rangle$
Z_2	$t_2\Delta$	4.71
S_2	$\tau_2\Omega$	1.44
Q_2	q_2	$-2.09-2.34i$
$\|\psi_1\rangle$		$(0.553-0.62i)\|g\rangle\|1\rangle-(0.371+0.416i)\|e\rangle\|0\rangle$
Z_1	$t_1\Delta$	3.26
S_1	$\tau_1\Omega$	1.96
Q_1	q_1	$-2.71-1.59i$
$\|\psi_0\rangle$		$(0.197-0.98i)\|g\rangle\|0\rangle$

This resonator state is used for the measurements described in Fig. 2. The sequence is computed top to bottom, but applied bottom to top. The area and phase for the nth qubit drive Q_n is $q_n = \int \Omega_q(t)e^{i\Delta_{off}t}dt$ ($t = 0$ being the time when the qubit is tuned into resonance directly after the step Q_n), the time on-resonance for the qubit–resonator swap operation S_n is τ_n, and the time off-resonance (mod $2\pi/\Delta$) for the phase rotation Z_n is t_n. We note that the initial state $|\psi_0\rangle$ differs by an overall phase factor from the ground state $|g\rangle|0\rangle$, but this is not detectable. State descriptions are shown bold; operations are not in bold.

describe the system with a Hamiltonian in the resonator rotating frame, so that the resonator states have zero frequency:

$$\frac{H}{\hbar} = \Delta(t)\sigma_+\sigma_- + \left(\frac{\Omega}{2}\sigma_+ a + \frac{\Omega_q(t)}{2}\sigma_+ + \frac{\Omega_r(t)}{2}a^\dagger\right) + \text{h.c.} \quad (1)$$

Here σ_+ and σ_- ($a^\dagger$ and a) are the qubit (resonator) raising and lowering operators, and h.c. is the Hermitian conjugate of the terms in parentheses. The first term is the qubit energy, which appears as the qubit–resonator detuning $\Delta(t) = \omega_q(t) - \omega_r$. The first term in the parentheses gives the qubit–resonator interaction, proportional to the fixed interaction strength $\Omega/2\pi = 19$ MHz, while the second and third terms give the effect of the external microwave drive signals applied to the qubit and resonator; these parameters $\Omega_q(t)$ and $\Omega_r(t)$ are complex to account for amplitude and phase. All control signals in equation (1) vary on a ~2 ns timescale, long compared to $2\pi/\omega_r$, so counter-rotating terms in equation (1) are neglected.

Although the coupling Ω is fixed, we control the qubit–resonator interaction by adjusting the qubit frequency between two operating points, one with qubit and resonator exactly on resonance ($\Delta_{on} = 0$), the other with the qubit well off-resonance ($|\Delta_{off}| \gg \Omega$). On resonance, the coupling will produce an oscillation where a single photon transfers between qubit and resonator with unit probability, alternating between states with, say, the qubit in its ground state with n photons in the resonator, $|g\rangle \otimes |n\rangle = |g, n\rangle$, and the qubit in its excited state with $n-1$ photons in the resonator, $|e, n-1\rangle$; this occurs at the n-photon 'Rabi-swap' frequency $\sqrt{n}\Omega$. Off resonance, the system oscillates at a higher frequency $\sqrt{n\Omega^2+\Delta^2}$ but with reduced $|e, n-1\rangle$ probability $n\Omega^2/(n\Omega^2+\Delta^2) < 1$. This detuning dependence is shown in Fig. 1 for $n = 1$ photon and small detunings $|\Delta| \lesssim \Omega$. At our typical off-resonance operating point $\Delta_{off} \approx -25\Omega$, the photon transfer probability is only 0.0016 n, so the coupling is essentially turned off.

We determine from Fig. 1 the flux bias for on-resonance tuning ($\Delta = 0$) and the one-photon swap time. Using these parameters, we can pump photons one at a time into the resonator by repeatedly exciting the detuned qubit from $|g\rangle$ to $|e\rangle$ using a qubit microwave π-pulse, followed by a controlled-time, on-resonance photon swap[8], where we scale the swap time for the nth photon by $1/\sqrt{n}$. Precise scaling of this swap time is crucial for proper control, and was verified for up to 15 photons (see Supplementary Information).

Our goal is to synthesize arbitrary N-photon states in the resonator with the qubit in its ground state, disentangled from the resonator. Our target state for the coupled system is

$$|\psi\rangle = |g\rangle \otimes \sum_{n=0}^{N} c_n|n\rangle \quad (2)$$

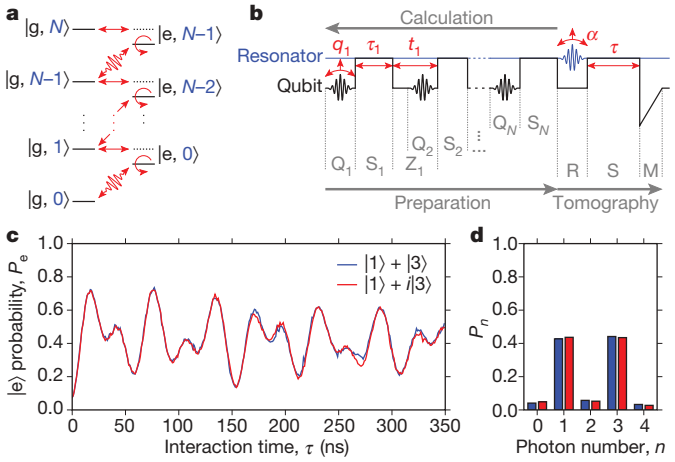


Figure 2 | Sequence to synthesize an arbitrary resonator state. **a**, Qubit–resonator energy ladder. Levels are depicted by dotted and solid lines when tuned ($\Delta = 0$) and detuned, respectively; qubit states are in black, resonator states are in blue. Three types of operations (in red) are used in state preparation: qubit drive operations Q_n, indicated by undulating lines; qubit–resonator swap operations S_n, indicated by straight horizontal lines; and phase rotations of the qubit state Z_n, indicated by circles. Each operation affects all the levels in the diagram. **b**, Microwave pulse sequence. The qubit and resonator are traced in black and blue, respectively, with qubit operations in red. The sequence is computed in reverse order by emptying energy levels from top to bottom. To descend the first step of the ladder in **a**, a swap operation S_N transfers the highest occupied resonator state to the qubit, $|g, N\rangle \rightarrow |e, N-1\rangle$. This operation also performs incomplete transfers on all the lower-lying states, as do the succeeding steps. A qubit microwave drive Q_N then transfers all the population of $|e, N-1\rangle$ to $|g, N-1\rangle$ (in general this step is not a π-pulse as $|g, N-1\rangle$ is not completely emptied by pulse S_N). For the second step down the ladder, a rotation Z_{N-1} first adjusts the phase of the qubit excited state $|e\rangle$ relative to the ground state $|g\rangle$. The succeeding swap pulse S_{N-1} can then move the entire population of $|g, N-1\rangle$ to $|e, N-2\rangle$. This sequence is repeated N times until the ground state $|g, 0\rangle$ is reached. Steps Q_n are performed with resonant qubit microwave pulses of amplitude q_n, swaps S_n achieved by bringing the qubit and resonator on resonance for time τ_n, and phase rotations Z_n completed by adjusting the detuning time t_n; see Table 1 for a detailed example. After state preparation, tomographic readout is performed: a displacement $D(-\alpha)$ of the resonator is performed by a microwave pulse R to the resonator, then the resonator state is probed by a qubit–resonator swap S for a variable interaction time τ, and finally the qubit state measured by the measurement pulse M. **c**, Plot of the qubit excited state probability P_e versus interaction time τ for the resonator states $|\psi_a\rangle = |1\rangle + |3\rangle$ (blue) and $|\psi_b\rangle = |1\rangle + i|3\rangle$ (red), taken with $\alpha = 0$. We clearly observe oscillations at the $|1\rangle$ and $|3\rangle$ Fock state frequencies. Nearly identical traces for $|\psi_a\rangle$ and $|\psi_b\rangle$ indicate the same photon number probability distribution, as expected. **d**, Photon number distributions for $|\psi_a\rangle$ (blue) and $|\psi_b\rangle$ (red). Both states are equal superpositions of $|1\rangle$ and $|3\rangle$ but the phase information that distinguishes the two state is lost.

with complex amplitude c_n for the nth Fock state. Law and Eberly[7] showed that these states can be generated by sequentially exciting the qubit into the proper superposition of $|g\rangle$ and $|e\rangle$, and then performing a partial transfer to the resonator. As illustrated in Fig. 2, and detailed in Table 1, a sequence generating the desired state can be found by solving the time-reversed problem: starting with the desired final state, we first transfer the amplitude of the highest occupied resonator Fock state to the qubit, then remove the excitation from the subsequently detuned qubit using a classical microwave signal, and repeat until the ground state $|g, 0\rangle$ is reached. The actual control signals are sequenced in the normal (un-reversed) order to generate the desired final state from the initial ground state. We note that the Law and Eberly protocol[7] assumes an adjustable phase for the qubit–resonator coupling Ω, which equation (1) does not allow; instead, we correct the relative phases of $|g, n\rangle$ and $|e, n-1\rangle$ by adjusting the time t_n over which the qubit and resonator are detuned.

To calibrate the actual microwave signals needed to implement this sequence, it is impractical to individually tune each sequence step,

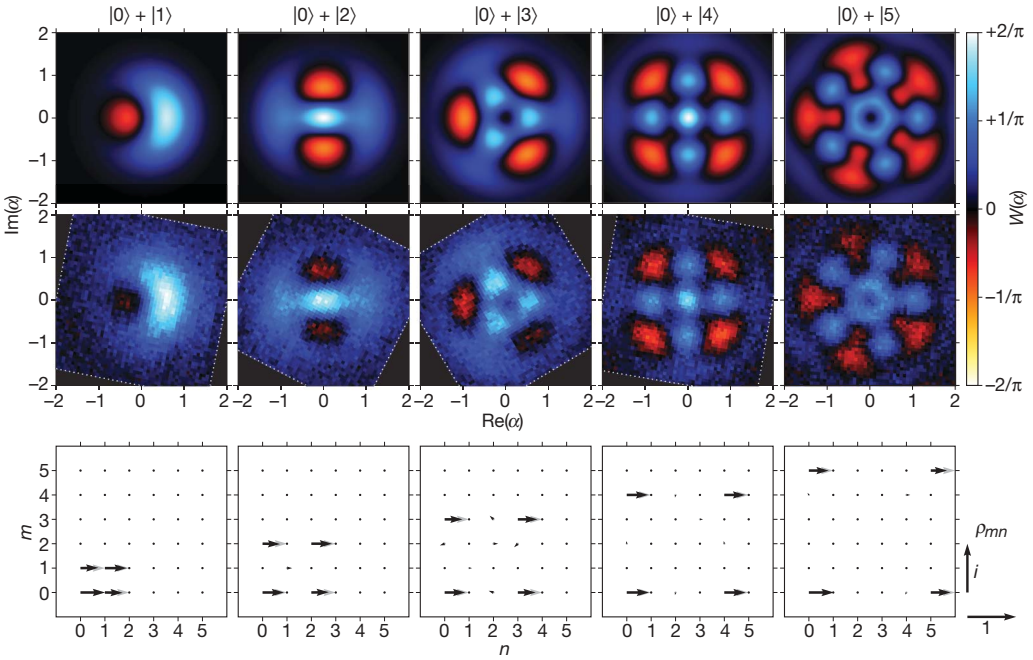


Figure 3 | Wigner tomography of superpositions of resonator Fock states $|0\rangle + |N\rangle$. The top row displays the theoretical form of the Wigner function $W(\alpha)$ as a function of the complex resonator amplitude α in photon number units, for states $N = 1$ to 5. The measured Wigner functions are shown in the middle row, with the colour scale bar on the far right. Negative quasi-probabilities are clearly measured. The experimental Wigner functions have been rotated to match theory, compensating for a phase delay between the qubit and resonator microwave lines; the measured area is bounded by a dotted white line. The bottom row displays the calculated (grey) and measured (black) values for the resonator density matrix ρ, projected onto the number states $\rho_{mn} = \langle m|\rho|n\rangle$. The magnitude and phase of ρ_{mn} is represented by the length and direction of an arrow in the complex plane (for scale, see key on right). The fidelities $F = \sqrt{\langle\psi|\rho|\psi\rangle}$ between the desired states $|\psi\rangle$ and the measured density matrices ρ are, from left to right, $F = 0.92$, 0.89, 0.88, 0.94 and 0.91. Each of the 51 by 51 pixels (61 by 61 for $N = 5$) in the Wigner function represents a local measurement. The value of $W(\alpha)$ is calculated at each pixel from 50 (41 for $N = 4$ and 5) interaction times τ, each repeated 900 times to give $P_e(\tau)$. This direct mapping of the Wigner function takes $\sim 10^8$ measurements or ~ 5 h.

because the intermediate states are quite complex and measuring them is time-consuming. Instead we perform careful calibrations of the experimental system independent of the particular state preparation (see Supplementary Information).

An initial check of the outcome of the preparation is to determine if the qubit ends up in the ground state $|g\rangle$, as desired. We find that this holds with a probability typically greater than 80%, the remaining 20% being compatible with decoherence during the preparation sequence (see Supplementary Information).

With the qubit near its ground state and not entangled with the resonator, we can use the qubit to measure the resonator state. By bringing the qubit and resonator into resonance for a variable time τ and subsequently measuring the probability P_e for the qubit excited state, we can determine[8] the n-photon probabilities $P_n = |c_n|^2$, correcting for measurement fidelity and initial qubit state probability (see Supplementary Information). In Fig. 2c we compare $P_e(\tau)$ for the experimentally prepared states $|\psi_a\rangle = |1\rangle + |3\rangle$ and $|\psi_b\rangle = |1\rangle + i|3\rangle$, showing the expected superposed oscillations corresponding to the $|1\rangle$ and $|3\rangle$ Fock states. This measurement however only yields the probabilities P_n: the relative phases of the Fock states are lost, so the states $|\psi_a\rangle$ and $|\psi_b\rangle$ cannot be distinguished.

To measure the complex amplitudes c_n, we need to probe the interference between the superposed Fock states. This may be done using Wigner tomography[19,21,24], which maps out the Wigner quasi-probability distribution $W(\alpha)$ as a function of the phase space amplitude α of the resonator (see Supplementary Information). Wigner tomography is performed by following the functional definition:

$$W(\alpha) = \frac{2}{\pi}\langle\psi|D^{\dagger}(-\alpha)\Pi\ D(-\alpha)|\psi\rangle \qquad (3)$$

The resonator state $|\psi\rangle$ is first displaced by the operator $D(-\alpha)$, implemented with a microwave drive pulse $-\alpha = (1/2)\int \Omega_r(t)\mathrm{d}t$. The photon number probabilities P_n are then measured and finally the parity $\langle\Pi\rangle = \sum_n (-1)^n P_n$ evaluated. The corresponding pulse sequence is depicted in Fig. 2b.

Calculated and measured Wigner functions are shown in Fig. 3 top and middle rows, respectively, for the resonator states $|0\rangle + |N\rangle$, with $N = 1$ to 5. The structures of the Wigner functions match well, including fine details, indicating that the superposed states are created and measured accurately. The density matrices for each state are also calculated (Fig. 3 bottom row; see Supplementary Information) and are as expected. The Wigner function of non-classical states has been measured previously, either calculated via an inverse Radon transform[18,26,27], or measured at enough points to fit the density matrix[3,28], from which the Wigner function is reconstructed. The high resolution direct mapping of the Wigner function used here is an important verification of our state preparation. The good agreement in shape shows that very pure superpositions of $|0\rangle$ and $|N\rangle$ have been created. Slight deviations in amplitude can be due to small errors in the read-out process, the relative amplitudes of the $|0\rangle$ and $|N\rangle$ states, or statistical mixtures with other Fock states.

The data in Fig. 3 do not demonstrate phase control between Fock states, as a change in the relative phase of a two-state superposition only rotates the Wigner function. The phase accuracy may be robustly demonstrated by preparing states with a superposition of three Fock states, as changing the phase of one state then changes the shape of the Wigner function. Figure 4 shows Wigner tomography for a superposition of the $|0\rangle$, $|3\rangle$ and $|6\rangle$ Fock states, where the phase of the $|3\rangle$ state has been changed in each of the five panels. The shape of the calculated and measured Wigner functions (Fig. 4 top and middle rows, respectively) again agree, including small details, indicating that precise phase control has been achieved. The calculated and measured density matrices (Fig. 4 bottom row) also match well.

In conclusion, we have generated and measured arbitrary superpositions of resonator quantum states. State preparation is deterministic and 'on-demand', requiring no projective measurements, and limited

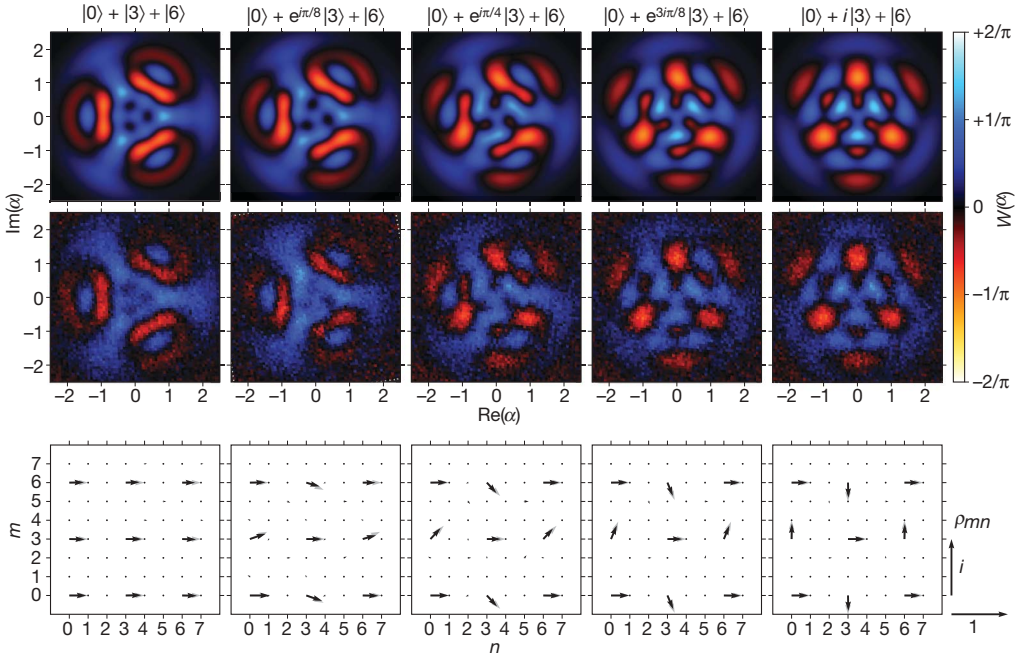


Figure 4 | Wigner tomography of the states $|0\rangle + e^{ik\pi/8}|3\rangle + |6\rangle$ for five values of phase $k = 0$ to 4. The top row is calculated, whereas the middle row shows measurements. The bottom row displays the calculated (grey) and measured (black) values for the density matrix obtained from the Wigner functions, displayed as for Fig. 3. The fidelities between the expected states and the measured density matrices are, from left to right, $F = 0.89$, 0.91, 0.91, 0.91 and 0.91.

to about ten photons, mainly by decoherence[29]. The accuracy of the prepared states demonstrates that a qubit, when controlled with high fidelity, is ideally suited for synthesizing and measuring arbitrary quantum states of light.

Received 15 January; accepted 19 March 2009.

1. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
2. Ben-Kish, A. *et al.* Experimental demonstration of a technique to generate arbitrary quantum superposition states of a harmonically bound spin-1/2 particle. *Phys. Rev. Lett.* **90,** 037902 (2003).
3. Deléglise, S. *et al.* Reconstruction of non-classical cavity field states with snapshots of their decoherence. *Nature* **455,** 510–514 (2008).
4. Houck, A. A. *et al.* Generating single microwave photons in a circuit. *Nature* **449,** 328–331 (2007).
5. Sillanpää, M. A., Park, J. I. & Simmonds, R. W. Coherent quantum state storage and transfer between two phase qubits via a resonant cavity. *Nature* **449,** 438–442 (2007).
6. Boozer, A. D., Boca, A., Miller, R., Northup, T. E. & Kimble, H. J. Reversible state transfer between light and a single trapped atom. *Phys. Rev. Lett.* **98,** 193601 (2007).
7. Law, C. K. & Eberly, J. H. Arbitrary control of a quantum electromagnetic field. *Phys. Rev. Lett.* **76,** 1055–1058 (1996).
8. Hofheinz, M. *et al.* Generation of Fock states in a superconducting quantum circuit. *Nature* **454,** 310–314 (2008).
9. Martinis, J. M., Devoret, M. H. & Clarke, J. Energy-level quantization in the zero-voltage state of a current-biased Josephson junction. *Phys. Rev. Lett.* **55,** 1543–1546 (1985).
10. Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453,** 1031–1042 (2008).
11. Vion, D. *et al.* Manipulating the quantum state of an electrical circuit. *Science* **296,** 886–889 (2002).
12. Niskanen, A. O. *et al.* Quantum coherent tunable coupling of superconducting qubits. *Science* **316,** 723–726 (2007).
13. Plantenberg, J. H., de Groot, P. C., Harmans, C. J. P. M. & Mooij, J. E. Demonstration of controlled-NOT quantum gates on a pair of superconducting quantum bits. *Nature* **447,** 836–839 (2007).
14. Fink, J. M. *et al.* Climbing the Jaynes-Cummings ladder and observing its $\sqrt{n}$ nonlinearity in a cavity QED system. *Nature* **454,** 315–318 (2008).
15. Steffen, M. *et al.* State tomography of capacitively shunted phase qubits with high fidelity. *Phys. Rev. Lett.* **97,** 050502 (2006).
16. Liu, Y. X. *et al.* Generation of non-classical photon states using a superconducting qubit in a quantum electrodynamic microcavity. *Europhys. Lett.* **67,** 941–947 (2004).
17. Vogel, K., Akulin, V. M. & Schleich, W. P. Quantum state engineering of the radiation field. *Phys. Rev. Lett.* **71,** 1816–1819 (1993).
18. Smithey, D. T., Beck, M., Raymer, M. G. & Faridani, A. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Phys. Rev. Lett.* **70,** 1244–1247 (1993).
19. Banaszek, K. & Wódkiewicz, K. Direct probing of quantum phase space by photon counting. *Phys. Rev. Lett.* **76,** 4344–4347 (1996).
20. Lutterbach, L. G. & Davidovich, L. Method for direct measurement of the Wigner function in cavity QED and ion traps. *Phys. Rev. Lett.* **78,** 2547–2550 (1997).
21. Banaszek, K., Radzewicz, C., Wódkiewicz, K. & Krasiński, J. S. Direct measurement of the Wigner function by photon counting. *Phys. Rev. A* **60,** 674–677 (1999).
22. Bertet, P. *et al.* Direct measurement of the Wigner function of a one-photon Fock state in a cavity. *Phys. Rev. Lett.* **89,** 200402 (2002).
23. Wigner, E. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.* **40,** 749–759 (1932).
24. Haroche, S. & Raimond, J.-M. *Exploring the Quantum — Atoms, Cavities and Photons* (Oxford Univ. Press, 2006).
25. Neeley, M. *et al.* Transformed dissipation in superconducting quantum circuits. *Phys. Rev. B* **77,** 180508 (2008).
26. Breitenbach, G., Schiller, S. & Mlynek, J. Measurement of the quantum states of squeezed light. *Nature* **387,** 471–475 (1997).
27. Lvovsky, A. I. & Babichev, S. A. Synthesis and tomographic characterization of the displaced Fock state of light. *Phys. Rev. A* **66,** 011801 (2002).
28. Leibfried, D. *et al.* Experimental determination of the motional quantum state of a trapped atom. *Phys. Rev. Lett.* **77,** 4281–4285 (1996).
29. Wang, H. *et al.* Measurement of the decay of Fock states in a superconducting quantum circuit. *Phys. Rev. Lett.* **101,** 240401 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Devices were made at the UCSB Nanofabrication Facility, a part of the NSF-funded National Nanotechnology Infrastructure Network. We thank M. Geller for discussions. This work was supported by IARPA (grant W911NF-04-1-0204) and by the NSF (grant CCF-0507227).

Author Contributions M.H. performed the experiments and analysed the data. H.W. improved the resonator design and fabricated the sample. J.M.M. and E.L. designed the custom electronics and M.H. developed the calibrations for it. M.A. and M.N. provided software infrastructure. All authors contributed to the fabrication process, qubit design or experimental set-up. M.H., J.M.M. and A.N.C. conceived the experiment and co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.N.C. (Cleland@physics.ucsb.edu).

Vol 459|28 May 2009|doi:10.1038/nature08061

LETTERS

A picogram- and nanometre-scale photonic-crystal optomechanical cavity

Matt Eichenfield[1], Ryan Camacho[1], Jasper Chan[1], Kerry J. Vahala[1] & Oskar Painter[1]

The dynamic back-action caused by electromagnetic forces (radiation pressure) in optical[1–6] and microwave[7] cavities is of growing interest[8]. Back-action cooling, for example, is being pursued as a means of achieving the quantum ground state of macroscopic mechanical oscillators. Work in the optical domain has revolved around millimetre- or micrometre-scale structures using the radiation pressure force. By comparison, in microwave devices, low-loss superconducting structures have been used for gradient-force-mediated coupling to a nanomechanical oscillator of picogram mass[7]. Here we describe measurements of an optical system consisting of a pair of specially patterned nanoscale beams in which optical and mechanical energies are simultaneously localized to a cubic-micron-scale volume, and for which large per-photon optical gradient forces are realized. The resulting scale of the per-photon force and the mass of the structure enable the exploration of cavity optomechanical regimes in which, for example, the mechanical rigidity of the structure is dominantly provided by the internal light field itself. In addition to precision measurement and sensitive force detection[9], nano-optomechanics may find application in reconfigurable and tunable photonic systems[10], light-based radio-frequency communication[11] and the generation of giant optical nonlinearities for wavelength conversion and optical buffering[12].

Optical forces arising from near-field effects in guided-wave structures have been proposed[13], and recently demonstrated[14,15], as a means of providing large optomechanical coupling between the field being guided and the dielectric mechanical structure doing the guiding. The resulting optical force can be viewed as an intensity gradient force much like that used to tweeze dielectric particles or to trap cold gases of atoms[16]. In the devices studied in this work, doubly clamped silicon nitride nanobeams are converted into optical resonant cavities through the patterning of a linear array of etched holes (Fig. 1b). Bringing two such cavities into each other's near-field forms a supercavity supporting even and odd superpositions of the individual beam modes. This 'zipper' cavity, so-named because of its resemblance to the mechanical fastener, allows for sensitive probing and actuation of the differential motion of the beams through the internal optical cavity field.

A figure of merit for cavity optomechanical systems is the coupling constant $g_{OM} \equiv d\omega_c/dx$, which represents the differential frequency shift of the cavity resonance (ω_c) with mechanical displacement of the beams (x). For the commonly studied Fabry–Pérot cavity structure (Fig. 1a), momentum transfer between the circulating light field and the mechanically compliant end mirror(s) occurs at a rate of $2\hbar\mathbf{k}_{ph}$ per round trip time (where $\mathbf{k}_{ph}$ is the optical wavevector and $\hbar$ is Planck's constant divided by 2π), resulting in an optomechanical coupling constant that scales with the inverse of the cavity length (L_c): $g_{OM} = \omega_c/L_c$. Similarly for whispering-gallery-mode structures, such as the recently studied microtoroid[17], g_{OM} scales with the perimeter length through the radius of the cavity (R): $g_{OM} = \omega_c/R$. In the case of the zipper cavity, the optomechanical coupling is exponentially proportional to the slot gap (s) between the beams: $g_{OM} = \omega_c/L_{OM}$ with $L_{OM} \approx w_o e^{\alpha s}$. In-plane differential motion of the nanobeams is thus strongly coupled to the optical cavity field, whereas common motion is nearly decoupled.

The minimum value of L_{OM} is determined by the constant w_o, which is approximately equal to the beam width, and the decay constant (α), which is set by the wavelength of light (λ) and the refractive-index contrast of the nanobeam system. Thus, for beam widths on the order of the wavelength of light and for a subwavelength slot gap, we have $L_{OM} \approx \lambda$, independent of the length of the nanobeams (Fig. 2a, b). This yields an optomechanical coupling more than an order of magnitude larger than that which can be accomplished in high-finesse Fabry–Pérot cavities or glass microtoroid structures[17]. In addition, this large optomechanical coupling is realized in a versatile geometry in which motional mass and mechanical stiffness can be greatly varied, and for which the mechanical displacement energy density and optical energy density can be efficiently co-localized at optical wavelengths in the visible–near-infrared range and for mechanical frequencies in the megahertz–gigahertz range.

For the devices studied in this work, optically thin (thickness, $t = 400$ nm) stoichiometric silicon nitride (Si_3N_4) is deposited using low-pressure chemical vapour deposition on a silicon wafer to form the optical guiding layer and the mechanical beams. Electron beam lithography is used to pattern the zipper cavity, which consists of beams of length $l = 25$–40 μm and width $w = 0.6$–1.4 μm, with a slot gap of $s = 60$–250 nm (Fig. 1f). The optical cavity is created in the nanobeams by patterning holes to form a quasi-one-dimensional photonic bandgap for light (Figs 1e–g). A C_4F_8–SF_6-based plasma etch is then used to transfer the nanobeam and photonic-crystal pattern into the Si_3N_4. This is followed by a wet chemical etch of KOH that selectively etches the underlying silicon substrate and releases the patterned beams.

As shown in Fig. 1g, optical excitation and probing of the zipper cavity is performed using a high-efficiency optical-fibre-taper coupler[18] in conjunction with a bank of tunable external-cavity diode lasers. A fibre-polarization controller is used to adjust the polarization to selectively excite the transverse electric (TE) polarization modes of the zipper cavity. The design of the zipper optical cavity is based on a graded-lattice concept[18–20] in which the lattice period is varied harmonically from the centre to the ends of the nanobeam. This results in an optical potential for photons that increases harmonically towards the cavity centre. Localized modes form from photonic bands near the zone boundary that have negative group-velocity dispersion, with the fundamental mode of the cavity having the highest frequency and higher-order cavity modes decreasing in frequency (Fig. 2c, inset). Owing to the strong optical coupling between the pair of nanobeams, the photonic bands in the zipper cavity break into pairs of positive- and

[1]Thomas J. Watson, Sr. Laboratory of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA.

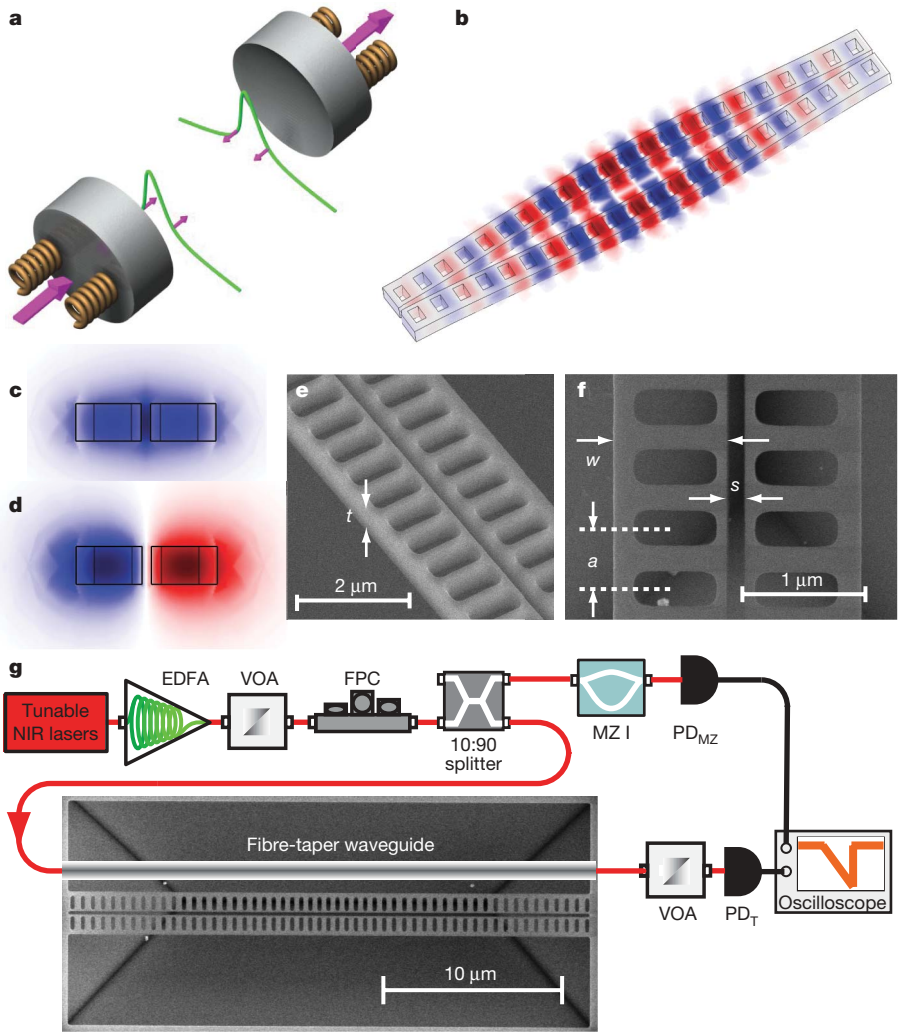


Figure 1 | Zipper cavity optomechanical system and experimental set-up. **a**, **b**, Fabry–Pérot (**a**) and photonic-crystal (**b**) optomechanical systems. **c**, **d**, Finite-element-method (FEM) simulated bonded (**c**) and antibonded (**d**) supermodes of the zipper optical cavity, shown in cross-section. **e**, **f**, Scanning electron microscope images of a typical zipper cavity, indicating the beam thickness (t), the slot width (s), the beam width (w) and the photonic-crystal lattice constant (a). **g**, Experimental set-up used to probe the optical and mechanical properties of the zipper cavity. NIR, near infrared; EDFA, erbium-doped fibre amplifier; VOA, variable optical attenuator; FPC, fibre-polarization controller; MZI, fibre Mach–Zehnder interferometer, PD_{MZ}, photodetected Mach–Zehnder transmission; PD_T, zipper cavity transmission.

negative-parity supermode bands. The positive-parity superposition, designated TE_+, corresponds to a manifold of modes that have an even-parity mode profile for the transverse electric field polarization and a peak electric field intensity in the centre of the slot gap between the beams (Fig. 1c). These we term bonded modes[13]. The negative-parity (TE_-) manifold of modes (the antibonded modes) have an odd-parity mode profile and a node at the centre of the slot gap (Fig. 1d).

By systematically varying the lattice constant of the devices, and measuring the parity of the cavity modes using the fibre taper as a near-field probe[18], it is possible to identify the various zipper cavity modes. For example, for a zipper cavity with photonic-crystal lattice constant $a = 640$ nm, beam width $w = 650$ nm and slot gap $s = 120$ nm, the measured transmission scan across the wavelength range $\lambda = 1{,}420$–$1{,}625$ nm is shown in Fig. 2c. From shortest to longest wavelength, the resonance peaks all have an even mode profile and are associated with the $TE_{+,0}$ to $TE_{+,4}$ modes of the manifold of bonded modes. Wavelength scans of a different zipper cavity, with larger beam width, $w = 1.4$ μm, and slot gap, $s = 250$ nm, show a spectrum in which the manifolds of bonded and antibonded modes overlap (Fig. 2d). In Fig. 2e, f, we show the measured on-resonance transmission contrast as a function of lateral taper position for each of the modes, indicating their even (bonded) and odd (antibonded) mode character. The optical Q-factor of the zipper cavity mode $TE_{+,0}$ can theoretically reach a value much greater than 10^6 even with the modest refractive index afforded by the silicon nitride[19,21]. We have experimentally measured zipper cavity modes with Q-factors in the range of $Q = 10^4$–10^5 (finesse $\mathcal{F} \approx 10^4$–10^5). For devices at the high end of the Q range ($Q \approx 3 \times 10^5$), we find a significant contribution to optical loss from absorption (Methods and Supplementary Information).

Mechanical motion of the zipper cavity nanobeams is imprinted on the transmitted optical intensity through the phase modulation of the internal cavity field[17]. Figure 3a shows the high-temporal-resolution (blue curve) and low-pass-filtered (red curve) transmitted signals as the input laser frequency (ω_l) is swept at low optical input power ($P_i = 12$ μW) across the zipper cavity mode $TE_{+,1}$ shown in Fig. 2c. The magnified temporal response of the transmitted intensity for a detuning on the side of the Lorentzian line shape (Fig. 3b) shows an oscillating signal of frequency ~8 MHz and peak-to-peak amplitude of roughly a third of the transmission contrast of the resonance. FEM simulations (Fig. 3c, d) indicate that the lowest-order in-plane common (h_{1c}) and differential (h_{1d}) mechanical modes of the pair of coupled nanobeams have frequencies of 8.19 and 8.16 MHz (mass, $m_x \approx 43$ pg; spring constant, $k_{h1} \approx 110\ \mathrm{N\,m^{-1}}$), respectively, when accounting for ~1 GPa of internal tensile stress in the nitride film[19,22].

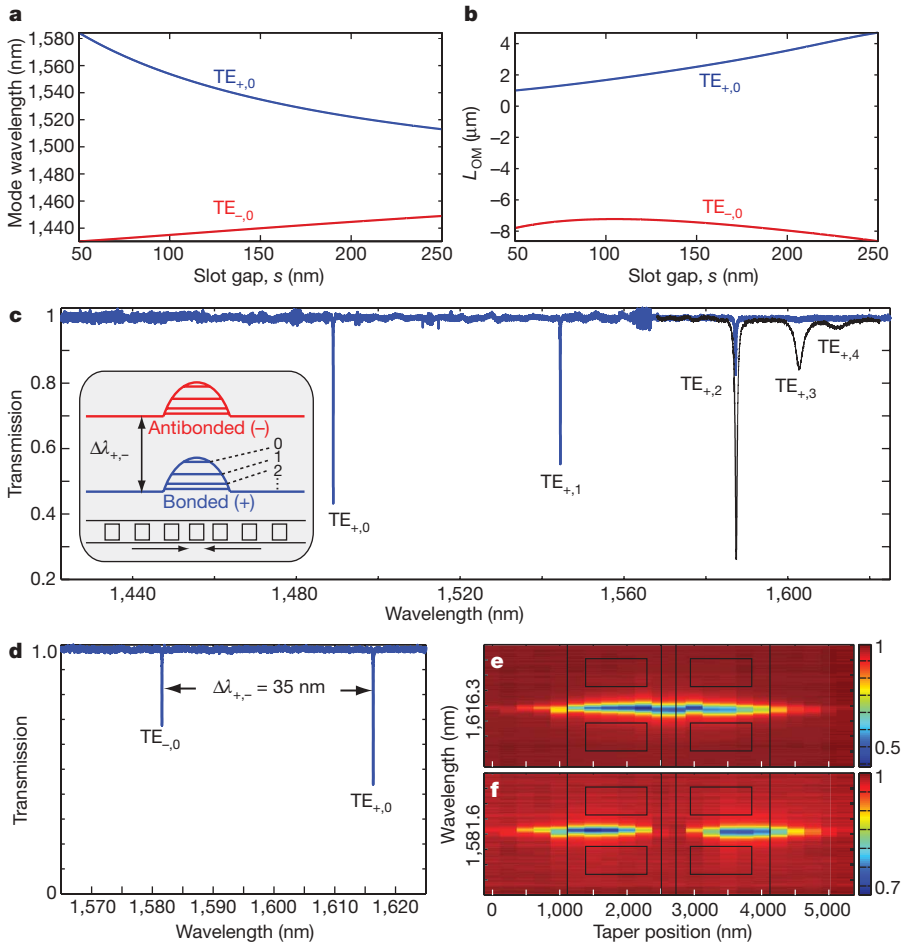


Figure 2 | Optical spectroscopy of the zipper cavity. a, **b**, FEM simulation of the wavelength tuning (**a**) and the effective optomechanical-coupling length parameter (L_{OM}) (**b**) as functions of nanobeam slot gap for the bonded and antibonded fundamental zipper cavity modes ($w = 650$ nm). **c**, Measured optical transmission of a zipper cavity with $w = 650$ nm and $s = 120$ nm, showing four orders of the bonded (TE_+) resonant modes. Inset, schematic of the graded photonic-crystal design and the resulting bonded and antibonded resonance manifolds. **d**, Measured optical transmission of a zipper cavity with a larger beam width and gap ($w = 1{,}400$ nm, $s = 250$ nm). **e**, **f**, Optical transmission (colour scale) versus fibre-taper lateral position for each of the bonded and antibonded resonant modes in **d**, indicating the even and odd parities of the modes. The zipper cavity is shown outlined in black.

The corresponding mechanical amplitude of oscillation is calibrated by fitting g_{OM} using the optical-spring effect as discussed below, yielding $L_{OM} = 1.58$ μm ($g_{OM}/2\pi = 123$ GHz nm^{-1}) and an inferred root-mean-squared amplitude of motion of $x_{rms} \approx 5.8$ pm. This is in good correspondence with both the FEM-simulated optomechanical coupling constant for this device ($L_{OM} = 2.1$ μm for $s = 120$ nm in Fig. 2b) and the expected thermal amplitude for the h_{1d} mode ($\langle x_{th}^2 \rangle^{1/2} = (k_B T/k_{h1})^{1/2} = 6.2$ pm, where T is temperature and k_B is Boltzmann's constant).

The radio-frequency spectrum of the transmitted optical intensity out to 150 MHz is shown in Fig. 3e. Comparison with FEM mechanical simulations[19] allows us to identify many of the resonances in the radio-frequency spectrum. The strength of the corresponding spectral peaks oscillates for odd and even orders of in-plane motion, consistent with the odd-order mechanical modes having an antinode of displacement at the centre of the zipper cavity and the even-order modes having a node. For the odd-order in-plane modes, on the basis of the amplitude of transduced thermal motion, the optomechanical coupling is measured to be roughly constant out to the ninth-order modes with mechanical resonance frequency of $\Omega_M/2\pi \approx 140$ MHz, in good correspondence with theoretical predictions[19]. The mechanical Q-factor of the resonances are measured to vary between $Q_M \approx 50$ and 150, limited by gas damping[23] in the 1-atm nitrogen test environment used in this work. The right-hand axis of Fig. 3e, which is calibrated in units of displacement sensitivity for the h_{1d} mode, shows a photoreceiver-noise-limited (grey curve) sensitivity of 5×10^{-17} m Hz$^{-1/2}$ for the optical input power of $P_i = 12$ μW (~600 stored cavity photons). This is within a factor of four of the resonant displacement sensitivity of the h_{1d} mode set by the standard quantum limit[24] (Supplementary Information). Even at this low probe power, a magnified view of the spectrum around the h_{1c} and h_{1d} resonances (Fig. 3e, inset) shows a slight shift of the resonance peaks to lower resonance frequencies for red laser–cavity detuning and to higher resonance frequencies for blue laser–cavity detuning. As described below, this is due to the optical-spring effect.

The optically driven zipper cavity not only allows for sensitive mechanical-displacement measurement, but can also strongly modify the mechanical motion in two distinct ways. Optical stiffening of the mechanical resonant structure[4,17] (the 'optical spring') results from the component of optical cavity energy (and gradient force) oscillating in phase with the mechanical motion. On the other hand, the finite cavity-photon lifetime introduces a non-adiabatic, time-delayed, component of optical force acting in quadrature with the mechanical motion. This velocity-dependent force results in detuning-dependent amplification or damping of the mechanical motion. A perturbative analysis shows that in the unresolved sideband limit ($\Omega_M \ll \Gamma$, where Γ is the energy decay rate of the waveguide-loaded optical cavity), the effective mechanical frequency (Ω'_M) and damping rate (γ'_M) are given by the following relations (ref. 17 and Supplementary Information):

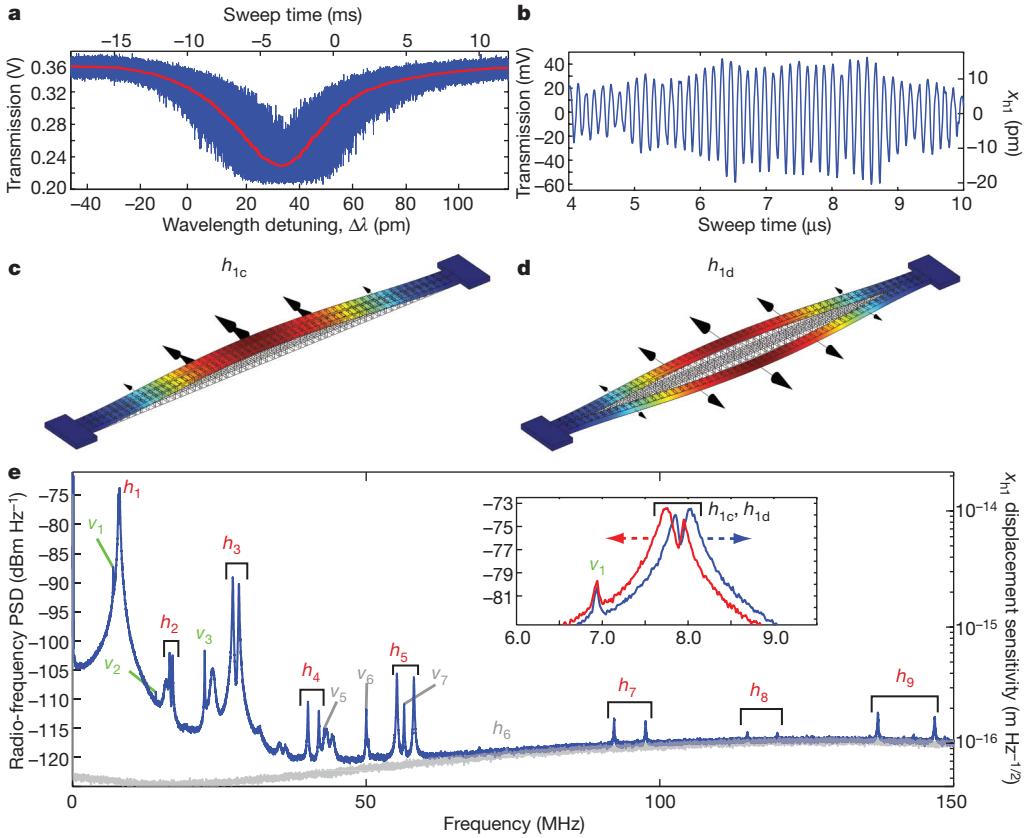


Figure 3 | Mechanical motion transduction. **a**, Optical transmission through the zipper cavity of Fig. 2c. Red curve corresponds to the low-pass-filtered (10 kHz) transmission signal (expressed as raw voltages from the PD_T photodetector) showing the underlying Lorentzian-like cavity resonance. **b**, Temporal oscillations in optical transmission for fixed detuning, showing oscillations of frequency ~8 MHz. The right axis indicates the corresponding displacement amplitude of the h_1 mechanical modes (x_{h1}). **c**, **d**, FEM-modelled lowest-order common (**c**) and differential (**d**) mechanical resonances. Mechanical deformation and colour indicate displacement amplitude; arrows indicate direction. **e**, Detected radio-frequency spectrum with horizontal (h) and vertical (v) cantilever modes of motion indicated. Grey labels indicate modes of questionable description. The grey curve is the detector electronic-noise floor. Inset, magnified radio-frequency spectrum of the fundamental h_1 mechanical modes (red, red-detuned spectrum; blue, blue-detuned spectrum). PSD, power spectral density.

$$\Omega'^2_{\mathrm{M}} = \Omega_{\mathrm{M}}^2 + \left(\frac{2|a_0|^2 g_{\mathrm{OM}}^2}{\Delta^2 \omega_{\mathrm{c}} m_x}\right)\Delta'_{\mathrm{o}}$$

$$\gamma'_{\mathrm{M}} = \gamma_{\mathrm{M}} - \left(\frac{2|a_0|^2 g_{\mathrm{OM}}^2 \Gamma}{\Delta^4 \omega_{\mathrm{c}} m_x}\right)\Delta'_{\mathrm{o}} \qquad (1)$$

Here Ω_{M} and γ_{M} are the bare mechanical properties of the zipper cavity, $|a_0|^2$ is the time-averaged energy stored in the optical cavity, $\Delta'_{\mathrm{o}} \equiv \omega_{\mathrm{l}} - \omega_{\mathrm{c}}$ is the laser–cavity detuning and $\Delta^2 \equiv \Delta'^2_{\mathrm{o}} + (\Gamma/2)^2$.

As is shown in Fig. 4, for higher optical input powers ($P_{\mathrm{i}} = 5$ mW) the internal optical cavity field provides significant stiffness to mechanical motion of modes in the zipper cavity. The large optomechanical coupling ($L_{\mathrm{OM}} = 1.58$ μm) and comparatively small motional mass ($m_x \approx 43$ pg) of the zipper cavity $h_{1\mathrm{d}}$ mechanical mode results in a giant optical-spring effect[4], shifting the mechanical frequency from 8 to 19 MHz (Fig. 4g). This corresponds to an optical stiffness more than five times greater than the intrinsic mechanical stiffness of the silicon nitride nanobeams. An additional feature (Fig. 4c) is the mechanical mode mixing that occurs as the optical spring tunes the $h_{1\mathrm{d}}$ mode through other mechanical resonances. The mode mixing is most prevalent as the $h_{1\mathrm{d}}$ mode sweeps away from the $h_{1\mathrm{c}}$ mode and through the h_2 in-plane modes near $\Omega/2\pi = 16$ MHz (Fig. 4c). This mixing and renormalization of mechanical modes is due to the highly anisotropic and motion-dependent optical stiffness, and will be discussed in more detail elsewhere.

Figure 4d–g compares the measured integrated radio-frequency power in the $h_{1\mathrm{d}}$ mechanical resonance line, and its mechanical frequency, with a nonlinear optical model of the zipper cavity system including the optical gradient force and thermo-optic tuning of the cavity. At low optical power (Fig. 4d, e), a single estimate for g_{OM} based on optical FEM simulations fits both the total measured radio-frequency power (or $\langle x^2 \rangle$) and the frequency of the $h_{1\mathrm{d}}$ mode over a large detuning range. At higher powers (Fig. 4f, g), the same estimated g_{OM} fits the optical-frequency tuning of the $h_{1\mathrm{d}}$ mode but severely overestimates the total radio-frequency power (Fig. 4f, green curve) where the optomechanical interaction is strongest. Damping of the mechanical motion is unexpected for blue-detuned laser excitation[17]. FEM numerical simulations of the zipper cavity indicate that thermo-mechanical effects[25,26] produce a response several orders of magnitude too small to explain the observed damping; however, a theoretical analysis of the cavity dynamics including the thermo-optic effect (Supplementary Information) shows that the severely phase-lagged and damped thermo-optic tuning of the cavity introduces a significant correction to equation (1). Owing to the small heat capacity of the zipper cavity, thermo-optic tuning reverses the sign of the damping coefficient of the bare optomechanical response for blue-detuned pumping (the correction to the optical spring is found to be small, at the 10^{-4} level). The numerical model including the thermo-optic correction to the spring and damping terms is shown as a blue curve in Fig. 4d, f, with the fit to the high-power data now in much better agreement. The model indicates that at a detuning of $\Delta'_{\mathrm{o}} \approx \Gamma/4$ (Fig. 4f), the thermal motion of the $h_{1\mathrm{d}}$ resonance is being damped from $x_{\mathrm{rms}} \approx 7$ pm to $x_{\mathrm{rms}} \approx 1$ pm, at a bath temperature of 360 K.

Beyond the giant optical-spring effect afforded by the large optomechanical coupling and picogram-scale mass of the zipper cavity,

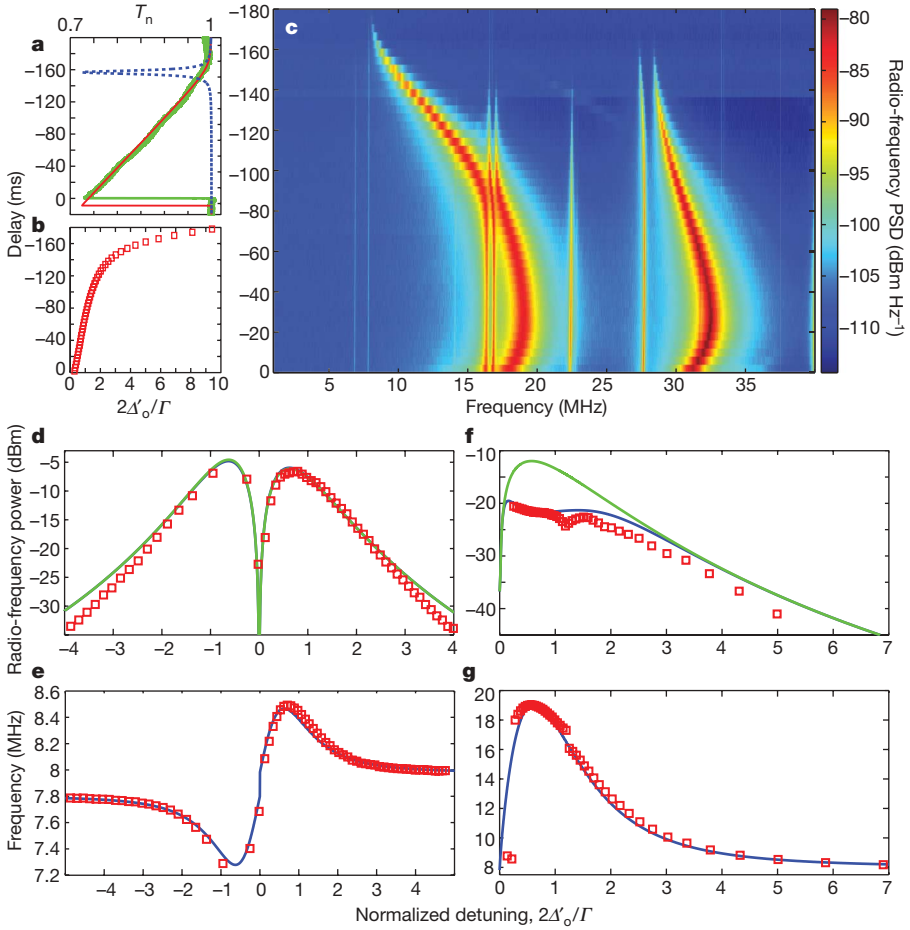


Figure 4 | Optical spring and damping. a, Measured (green curve) and fitted model (red curve) normalized optical transmission (T_n) versus wavelength sweep expressed in terms of sweep time delay. The dashed blue curve is a low-power model curve. **b**, Relation between sweep time delay and normalized cavity detuning, fitted using the red model curve in **a**, with zero time delay corresponding to zero laser–cavity detuning. **c**, Intensity image of the measured radio-frequency power spectrum versus cavity detuning (time delay) of the optical transmission signal from the zipper cavity of Fig. 2c at an input optical power of 5.1 mW. **d**–**g**, Measured and modelled total radio-frequency power (**d** and **f**, respectively) and resonance frequency (**e** and **g**, respectively) of the h_{1d} mechanical mode, versus detuning. **d** and **e** correspond to low optical input power ($P_i = 127\ \mu W$) and **f** and **g** correspond to high optical input power ($P_i = 5.1$ mW). The blue and green curves correspond to models with and, respectively, without optomechanical damping. Red squares, measured data points.

thin-film photonic crystals offer a chip-scale architecture for coupling optical and mechanical degrees of freedom. In the area of quantum cavity optomechanics, significant improvements in optical Q-factor to values approaching 5×10^6 ($\mathcal{F} \approx 10^6$) can be expected with new design and processing techniques[19,27]. Along with the higher frequencies of the higher-order in-plane mechanical modes (>25 MHz), this will push the system into the important resolved-sideband limit[28,29]. Having high mechanical Q-factor is also important for a variety of applications, and preliminary *in vacuo* measurements of the zipper cavity (Supplementary Information) show a stress-enhanced[22] value of $Q_M = 1.6 \times 10^4$ for the fundamental h_{1d} mode. Finally, by combining phononic-crystal[30] with photonic-crystal concepts, simultaneous routing and localization of acoustic and optical waves may be possible. Such a platform would not only provide a means of reducing mechanical clamping losses in higher-frequency photonic-crystal nanobeam cavities, but it would also greatly expand both quantum and classical applications by enabling integration not possible in current cavity optomechanical microsystems.

METHODS SUMMARY

A fibre-taper optical coupling technique[18] is used to in-couple and out-couple light from the zipper cavity with low insertion loss (<3 dB). The optical power reaching the zipper cavity is calibrated by measuring the thermo-optic wavelength shift for power sent in both directions down the fibre and from the total transmission efficiency of the fibre-taper section. With the fibre taper placed in the near-field, and evanescently coupled to the zipper cavity, the mechanical motion and radio-frequency spectra of the cavity are measured by monitoring the fibre-taper transmitted optical power using direct detection on a 125-MHz-bandwidth photoreceiver connected to a high-speed oscilloscope. Two 'duelling' calibrated optical attenuators (Fig. 1g) are used to vary the input power to the cavity while keeping the detected optical power level constant. A fibre-based Mach–Zehnder interferometer accurately converts sweep time to wavelength for the swept-laser-frequency measurements in this work. In conjunction with a nonlinear model incorporating the optomechanical force and thermo-optic cavity tuning, the Mach–Zehnder signal provides an accurate determination of laser–cavity detuning. Radio-frequency spectra of the transmitted light intensity for different detunings are obtained by Fourier transforming a 100-μs time interval around the desired detuning as the laser is swept across the cavity resonance in ~100 ms. The radio-frequency spectra are then averaged over ~100 sweeps to obtain the displayed plots.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 December 2008; accepted 8 April 2009.
Published online 13 May 2009.

1. Arcizet, O., Cohadon, P.-F., Briant, T., Pinard, M. & Heidmann, A. Radiation-pressure cooling and optomechanical instability of a micromirror. *Nature* **444**, 71–73 (2006).
2. Gigan, S. *et al.* Self-cooling of a micromirror by radiation pressure. *Nature* **444**, 67–70 (2006).
3. Schliesser, A., Del'Haye, P., Nooshi, N., Vahala, K. J. & Kippenberg, T. J. Radiation pressure cooling of a micromechanical oscillator using dynamical backaction. *Phys. Rev. Lett.* **97**, 243905 (2006).

4. Corbitt, T. *et al.* Optical dilution and feedback cooling of a gram-scale oscillator to 6.9 mK. *Phys. Rev. Lett.* **99,** 160801 (2007).
5. Thompson, J. D. *et al.* Strong dispersive coupling of a high-finesse cavity to micromechanical membrane. *Nature* **452,** 72–75 (2008).
6. Kippenberg, T. J., Rokhsari, H., Carmon, T., Scherer, A. & Vahala, K. J. Analysis of radiation-pressure induced mechanical oscillation of an optical microcavity. *Phys. Rev. Lett.* **95,** 033901 (2005).
7. Regal, C. A., Tuefel, J. D. & Lehnert, K. W. Measuring nanomechanical motion with a microwave cavity interferometer. *Nature Phys.* **4,** 555–560 (2008).
8. Kippenberg, T. J. & Vahala, K. J. Cavity optomechanics: back-action at the mesoscale. *Science* **321,** 1172–1176 (2008).
9. Stowe, T. D. *et al.* Attonewton force detection using ultrathin silicon cantilevers. *Appl. Phys. Lett.* **71,** 288–290 (1997).
10. Rakich, P. T., Popovic, M. A., Soljacic, M. & Ippen, E. P. Trapping, coralling and spectral bonding of optical resonances through optically induced potentials. *Nature Photon.* **1,** 658–665 (2007).
11. Hossein-Zadeh, M. & Vahala, K. J. Photonic RF down-converter based on optomechanical oscillation. *IEEE Photon. Technol. Lett.* **20,** 234–236 (2008).
12. Notomi, M., Taniyama, H., Mitsugi, S. & Kuramochi, E. Optomechanical wavelength and energy conversion in high-*Q* double-layer cavities of photonic crystal slabs. *Phys. Rev. Lett.* **97,** 023903 (2006).
13. Povinelli, M. L. *et al.* Evanescent-wave bonding between optical waveguides. *Opt. Lett.* **30,** 3042–3044 (2005).
14. Eichenfield, M., Michael, C. P., Perahia, R. & Painter, O. Actuation of micro-optomechanical systems via cavity-enhanced optical dipole forces. *Nature Photon.* **1,** 416–422 (2007).
15. Li, M. *et al.* Harnessing optical forces in integrated photonic circuits. *Nature* **456,** 480–484 (2008).
16. Ashkin, A. History of optical trapping and manipulation of small-neutral particle, atoms, and molecules. *IEEE J. Quantum Electron.* **6,** 841–856 (2000).
17. Kippenberg, T. J. & Vahala, K. Cavity optomechanics. *Opt. Express* **15,** 17172–17205 (2007).
18. Srinivasan, K., Barclay, P. E., Borselli, M. & Painter, O. An optical fiber-based probe for photonic crystal microcavities. *IEEE J. Sel. Areas Comm.* **23,** 1321–1329 (2005).
19. Chan, J., Eichenfield, M., Camacho, R. & Painter, O. Optical and mechanical design of a "zipper" photonic crystal optomechanical cavity. *Opt. Express* **17,** 3802–3817 (2009).
20. Song, B.-S., Noda, S., Asano, T. & Akahane, Y. Ultra-high-*Q* photonic double-heterostructure nanocavity. *Nature Mater.* **4,** 207–210 (2005).
21. McCutcheon, M. W. & Loncar, M. Design of a silicon nitride photonic crystal nanocavity with a quality factor of one million for coupling to a diamond nanocrystal. *Opt. Express* **16,** 19136–19145 (2008).
22. Verbridge, S. S., Parpia, J. M., Reichenbach, R. B., Bellan, L. M. & Craighead, H. G. High quality factor resonance at room temperature with nanostrings under high tensile stress. *J. Appl. Phys.* **99,** 124304 (2006).
23. Verbridge, S. S., Illic, R., Craighead, H. G. & Parpia, J. M. Size and frequency dependent gas damping of nanomechanical resonators. *Appl. Phys. Lett.* **93,** 013101 (2008).
24. Tittonen, I. *et al.* Interferometric measurements of the position of a macroscopic body: towards observation of quantum limits. *Phys. Rev. A* **59,** 1038–1044 (1999).
25. Höhberger, C. & Karrai, K. Cavity cooling of a microlever. *Nature* **432,** 1002–1005 (2004).
26. Ilic, B., Krylov, S., Aubin, K., Reichenbach, R. & Craighead, H. G. Optical excitation of nanoelectromechanical oscillators. *Appl. Phys. Lett.* **86,** 193114 (2005).
27. Takahashi, Y. *et al.* High-*Q* nanocavity with a 2-ns photon lifetime. *Opt. Express* **15,** 17206–17213 (2007).
28. Schliesser, A., Riviere, R., Anetsberger, G., Arcizet, O. & Kippenberg, T. J. Resolved-sideband cooling of a micromechanical oscillator. *Nature Phys.* **4,** 415–419 (2008).
29. Marquardt, F., Harris, J. G. E. & Girvin, S. M. Dynamical multistability induced by radiation pressure in high-finesse micromechanical optical cavities. *Phys. Rev. Lett.* **96,** 103901 (2006).
30. Olson, R. H. III & El-Kady, I. Microfabricated phononic crystal devices and applications. *Meas. Sci. Technol.* **20,** 012002 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank Q. Lin for extensive discussions regarding this work, and for pointing out the origin of the mechanical resonance interference. Funding for this work was provided by a US Defense Advanced Research Projects Agency seedling effort managed by H. Temkin, and through an Emerging Models and Technologies grant from the US National Science Foundation.

Author Contributions M.E. and R.C. performed the majority of the fabrication and testing of devices and J.C. performed the optical and mechanical simulations. O.P., along with M.E. and K.J.V., developed the device concept. O.P., K.J.V., M.E. and R.C. all contributed to planning the measurements. All authors worked together to write the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.P. (opainter@caltech.edu).

METHODS

Optical transmission, measured radio-frequency spectra and motional sensitivity. Radio-frequency spectra are measured by direct detection of the optical power transmitted through the zipper cavity using a 125-MHz-bandwidth photoreceiver (noise equivalent power, NEP = 2.5 pW Hz$^{-1/2}$ from 0 to 10 MHz and 22.5 pW Hz$^{-1/2}$ from 10 to 200 MHz; responsivity, $R = 1$ A W^{-1}; transimpedance gain, $G = 4 \times 10^4$ V A^{-1}) and a high-speed oscilloscope (sampling rate, 2 Gs s^{-1}; bandwidth, 1 GHz). As shown in Fig. 1g, a pair of duelling calibrated optical attenuators is used to vary the input power to the cavity while keeping the detected optical power level constant. The measured electrical-noise floor is set by the circuit noise of the photodetector for the optical power levels considered in this work, which corresponds to -125 dBm Hz^{-1} near 10 MHz. The motional sensitivity of the h_{1d} mechanical mode is measured at 5×10^{-17} m Hz$^{-1/2}$ for an optical input power of 12 μW (corresponding to a dropped power of 3.5 μW and an estimated 660 stored cavity photons). At the power levels considered in this work, optical force noise contribution to the motional sensitivity is negligible (Supplementary Information).

Calibration of input power and intra-cavity photon number. A fibre-taper optical coupling technique is used to in-couple and out-couple light from the zipper cavity. The fibre taper, although extremely low loss on its own (88% transmission efficiency in this work), is put in contact with the substrate near the zipper cavity to mechanically anchor it during all measurements (thus avoiding power-dependent movement of the taper due to thermal and/or optical forces). The total fibre-taper transmission after mechanical anchoring of the taper to the substrate is 53%. To accurately determine the optical power reaching the cavity (determined by the optical loss in the taper section before the cavity) we measure the cavity response at high optical power (resulting in thermo-optic tuning of the cavity and optical bistability in the transmission response) with the input sent first in one direction along the taper and then in the other. From the asymmetry in the thermo-optic tuning in the cavity for both directions, it is possible to determine the asymmetry in the optical loss, and thus determine the optical loss before and after the zipper cavity. Finally, this method, along with calibrated measurements of the optical power at the input and output of the taper, can accurately determine the optical power reaching the zipper cavity (the input power) and that dropped by the cavity. From calibration of the wavelength sweep using the fibre Mach–Zehnder interferometer, it is also possible to measure the cavity linewidth and the corresponding loaded cavity Q-factor accurately. The average stored photon number can then be determined from the dropped power and the loaded cavity Q-factor. We choose to study the $TE_{+,1}$ mode in detail, instead of the $TE_{+,0}$ fundamental mode, because of its spectral alignment with the EDFA gain bandwidth, allowing for the higher power measurements presented in Fig. 4.

Calibration of laser–cavity detuning. The transduction from mechanical motion to modulated intra-cavity power, and consequently the measured RF photodetector spectrum, depends sensitively on the detuning point of the laser from the cavity resonance. Accurate measurement of the laser–cavity detuning, even for large detunings (>5 half-cavity linewidths), is required to compare the theoretical model with measured data for the optical spring and damping shown in Fig. 4. Several methods exist to determine the laser–cavity detuning, including calibration of the transduced modulated photodetector signal for a known mechanical or optical modulation. For the swept measurements presented in this work, we opt to accurately calibrate the laser wavelength versus sweep time using a fibre-based Mach–Zehnder interferometer (free spectral range of 1.57 pm at $\lambda \approx 1{,}480$ nm), and to use this to fit and convert sweep time to laser–cavity detuning by comparison with a nonlinear model of the cavity system that incorporates thermo-optic and gradient-force tuning (Supplementary Information). The thermo-optic cavity tuning with temperature is measured to be 14.9 pm K^{-1} by direct measurement of the resonance wavelength shift over a temperature range of 20 K. The optomechanical coupling constant (g_{OM}) is estimated from both simulation, based on a FEM model of scanning electron microscope images of the cavity geometry, and a fit to the peak measured mechanical-frequency shift. The nonlinear cavity model, incorporating the measured thermo-optic effect and the fitted g_{OM}, then provides an accurate conversion between wavelength and detuning from the cavity. The above method for calibrating laser–cavity detuning is simple to implement with the swept-wavelength method used in this work, and is found to be much more accurate than relying on the low-pass-filtered optical-transmission contrast to infer detuning (especially for large detunings where the transmission contrast is below the 1% level).

Zipper cavity optical loss. As mentioned above, although the optical force dominates the cavity tuning at megahertz frequencies, the static tuning of the cavity is still largely (~80%) provided by the thermo-optic effect through optical absorption and subsequent heat generation within the zipper cavity. Calculation of the thermal resistance of the silicon nitride zipper cavity indicates that optical absorption accounts for approximately 6% of the total optical cavity loss shown in Fig. 4a (an absorption-limited Q-factor of $\sim 4.8 \times 10^5$). We attribute the optical absorption loss in the zipper cavity to surface states of the 'holey' silicon nitride beams, rather than absorption in the bulk of the silicon nitride film, due to the much larger Q values we have measured in less surface-sensitive microdisks formed from the same silicon nitride material. Properly chosen chemical surface treatments should make possible Q-factors approaching the bulk-absorption-limited value of $Q_b \approx 5 \times 10^6$ at $\lambda = 1.5$ μm, and perhaps even higher at shorter wavelengths where optical absorption from overtones of the vibrational modes of the N–H bond is reduced.

Vol 459|28 May 2009|doi:10.1038/nature08031

LETTERS

The effect of permafrost thaw on old carbon release and net carbon exchange from tundra

Edward A. G. Schuur[1]*, Jason G. Vogel[1]*, Kathryn G. Crummer[1], Hanna Lee[1], James O. Sickman[2] & T. E. Osterkamp[3]

Permafrost soils in boreal and Arctic ecosystems store almost twice as much carbon[1,2] as is currently present in the atmosphere[3]. Permafrost thaw and the microbial decomposition of previously frozen organic carbon is considered one of the most likely positive climate feedbacks from terrestrial ecosystems to the atmosphere in a warmer world[1,2,4–7]. The rate of carbon release from permafrost soils is highly uncertain, but it is crucial for predicting the strength and timing of this carbon-cycle feedback effect, and thus how important permafrost thaw will be for climate change this century and beyond[1,2,4–7]. Sustained transfers of carbon to the atmosphere that could cause a significant positive feedback to climate change must come from old carbon, which forms the bulk of the permafrost carbon pool that accumulated over thousands of years[8–11]. Here we measure net ecosystem carbon exchange and the radiocarbon age of ecosystem respiration in a tundra landscape undergoing permafrost thaw[12] to determine the influence of old carbon loss on ecosystem carbon balance. We find that areas that thawed over the past 15 years had 40 per cent more annual losses of old carbon than minimally thawed areas, but had overall net ecosystem carbon uptake as increased plant growth offset these losses. In contrast, areas that thawed decades earlier lost even more old carbon, a 78 per cent increase over minimally thawed areas; this old carbon loss contributed to overall net ecosystem carbon release despite increased plant growth. Our data document significant losses of soil carbon with permafrost thaw that, over decadal timescales, overwhelms increased plant carbon uptake[13–15] at rates that could make permafrost a large biospheric carbon source in a warmer world.

We measured ecosystem carbon (C) dynamics at a tundra site in Alaska where permafrost thaw has been documented since 1990, and was occurring in the area before that time, probably owing to regional climate change[12]. The Eight Mile Lake watershed is located in the northern foothills of the Alaska Range (Supplementary Information and Supplementary Fig. 1). Permafrost temperature has been monitored annually on a gentle north-facing slope in a 30-m-deep borehole that was installed in 1985 before the permafrost started to thaw[12]. Although permafrost thaw can sometimes result in water ponding, depending on local topography[16,17], this landscape consists largely of relatively well-drained uplands (Supplementary Fig. 2). Permafrost thaw at the local level occurs as a series of positive feedbacks between temperature change, ground subsidence, and hydrologic redistribution that causes further thawing in a spatially heterogeneous pattern generated by water flow paths. In this watershed, we monitored three sites that represented minimal, moderate and extensive amounts of change as a result of the duration of permafrost thaw, based on observations of ground subsidence, depth of thaw, historical ground temperature measurements, and historical and current photographs (Supplementary Table 1)[12,18–20]. Although an observational study cannot unequivocally rule out unknown pre-existing differences across sites that might give rise to differences in carbon fluxes, all available historical evidence points to the soil and plant community originally being relatively similar across this hillslope. Here we assume that the site differences in onset of permafrost thaw are due to spatially random processes that occur at a local level as a series of positive feedbacks between temperature change, ground subsidence, hydrologic redistribution, and further thawing via thermal erosion from moving water, and therefore that the differences in permafrost thaw are the main driver of the differences in soil and plant community that we observe today. Across this gradient of thaw, we measured net carbon dioxide (CO_2) exchange between the tundra and the atmosphere over 3 years (Supplementary Figs 3 and 4), coupled with radiocarbon measurements (expressed as $\Delta^{14}C$) of respired C as a fingerprint for identifying the decomposition of old organic C that has been stored in these permafrost soils[21]. This is, to our knowledge, the first study to quantitatively demonstrate the link between old C emissions and ecosystem C losses.

The tundra ecosystem at Eight Mile Lake showed net C uptake during the summer months (June–August), with differences among sites and years (Mixed Model; site, $P = 0.053$; year, $P = 0.002$; site × year, $P = 0.473$). The moderate and extensive thaw sites trended towards greater C uptake, averaged across years, than the minimal thaw site (Tukey's pairwise comparisons; $P = 0.018$ and $P = 0.100$, respectively) (Fig. 1a). The pattern of significantly increased gross primary productivity at the moderate and extensive thaw sites, averaged across years for the full growing season (May–September), indicated that plant C uptake was stimulated by permafrost thaw (Supplementary Tables 2 and 3; Tukey's pairwise comparisons versus minimal thaw site; $P = 0.009$ and $P = 0.005$, respectively). Carbon gain in the summer was offset by net C loss in spring and autumn (May, September) (Fig. 1a) and during the winter (October–April) (Fig. 1b, Supplementary Table 2), when plant C uptake is low or absent while microbial decomposition proceeds even at subzero temperatures (Supplementary Fig. 4). Overall, C loss during winter accounted for 15–18% of ecosystem respiration (R_{eco}) on average across all sites. Winter C loss was large enough to switch the minimal and extensive sites, which were net C sinks in the growing season, to annual net C sources on average across years (Supplementary Table 2).

On an annual basis across years, the extensive thaw site had significantly higher R_{eco} ($P < 0.001$) with a trend towards greater total C loss than the minimal thaw site, while the moderate site had intermediate R_{eco} between the two (Supplementary Tables 2 and 3). The extensive thaw site was a net C source to the atmosphere, based on calculated annual net ecosystem C exchange, losing an average of $32 \pm 22\ g\,C\,m^{-2}\,yr^{-1}$ (mean $\neq 0$, $P = 0.019$). However, there was significant interannual variability; this site lost $136\ g\,C\,m^{-2}$ over the first two years of the study (Fig. 1c), but then gained

[1]Department of Biology, University of Florida, Gainesville, Florida 32611, USA. [2]Department of Environmental Science, University of California, Riverside, California 92521, USA. [3]Geophysical Institute, University of Alaska, Fairbanks, Alaska 99775, USA.
*These authors contributed equally to this work.

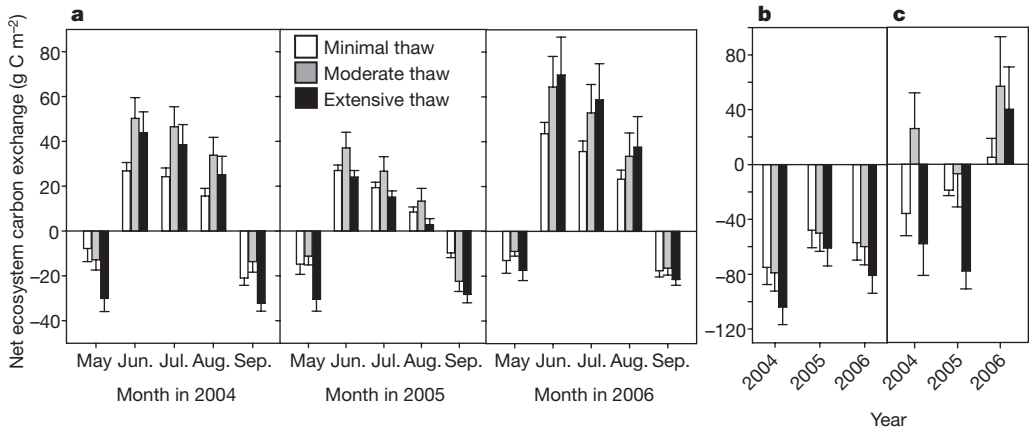


Figure 1 | Net exchange of CO_2 between tundra and the atmosphere for three sites that differ in the extent of permafrost thaw. **a**, For the growing season (May–September) over a 3-year period from 2004–06; **b**, for the winter (October–April); and **c**, on an annual basis, which is the net of the growing season and the winter. Values represent total C uptake (positive) or release (negative) per month (**a**), per winter (**b**) and per year (**c**). Data show mean ± s.e.

$40 \pm 31\ g\,C\,m^{-2}$ in 2006. The moderate and minimal thaw sites had similar interannual variability, with greatest uptake in 2006. Overall, the minimal thaw site lost $17 \pm 11\ g\,C\,m^{-2}\,yr^{-1}$ but this value was not significantly different from neutral C balance (mean $\neq 0$, $P = 0.215$), while the moderate thaw site contrasted with the other two sites as a net sink of atmospheric C, gaining $25 \pm 29\ g\,C\,m^{-2}\,yr^{-1}$ (mean $\neq 0$, $P = 0.060$).

Flux measurements alone cannot determine the influence of permafrost C on ecosystem fluxes, but $\Delta^{14}C$ provides a fingerprint for determining the source of ecosystem C loss. Total R_{eco} is a mixture of CO_2 derived from three sources: plant metabolism, decomposition of recently dead plant tissue (recent detritus), and decomposition of older soil organic matter (old C). Each of these sources has a characteristic $\Delta^{14}C$ value as a result of radioactive decay combined with the recent change in atmospheric ^{14}C abundance caused by aboveground thermonuclear weapons testing (Supplementary Fig. 5a). Radiocarbon values of R_{eco} across sites were generally elevated relative to current atmospheric values, and on average they declined over the 3-year period (Fig. 2a). This R_{eco} $\Delta^{14}C$ pattern demonstrates the influence of (1) recent photosynthate, whose $\Delta^{14}C$ value is closely tied to the contemporary atmosphere that is currently declining by ~5‰ annually[22], and (2) decomposition of recently dead plant tissue that grew over the past several decades and thus contained elevated 'bomb' $\Delta^{14}C$ values. Also contributing to R_{eco} is the decomposition

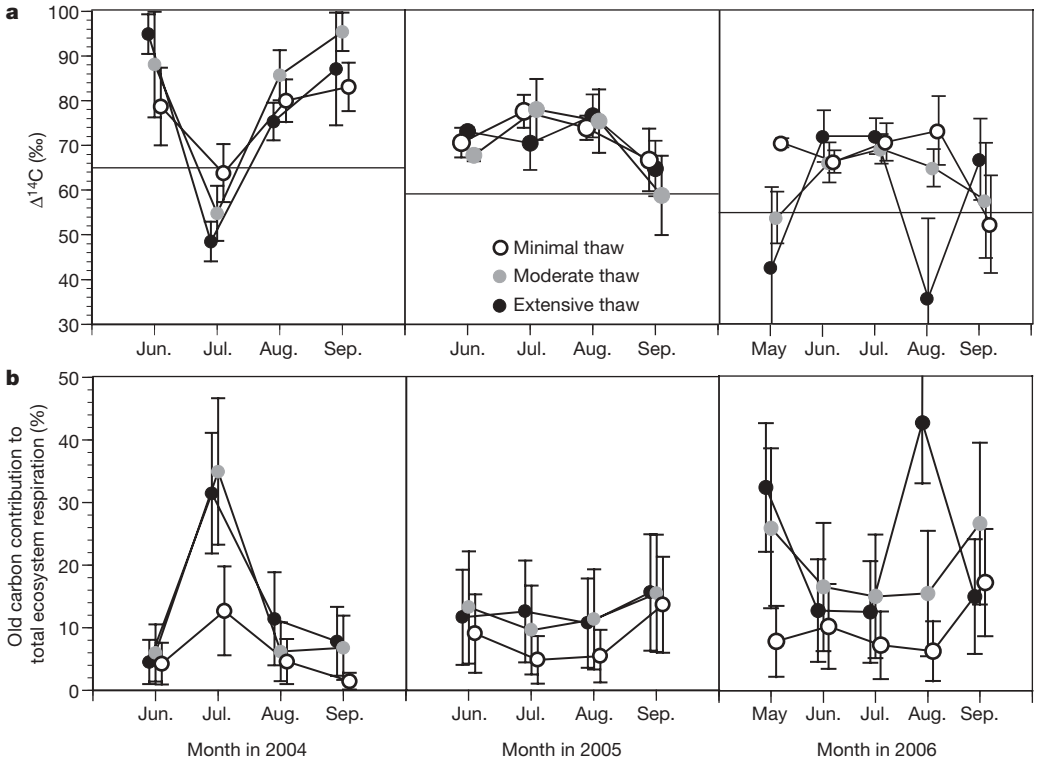


Figure 2 | Radiocarbon values of ecosystem respiration and the proportional contribution from old carbon for three sites that differ in the extent of permafrost thaw. **a**, Radiocarbon values (mean ± s.e.) of R_{eco} over a 3-year period from 2004–06. Horizontal lines across the graphs represent the average atmospheric value for each year. Radiocarbon is reported in $\Delta^{14}C$ units referenced to a 1890 wood standard defined as 0‰. **b**, Statistical partitioning estimate of the contribution of old C to R_{eco} (mean ± s.d.), based on $\Delta^{14}C$ measurements from incubations of surface soil, deep soil, and plants.

of old C with negative $\Delta^{14}C$ values as a result of radioactive decay. The imprint of old C is unequivocally observed in R_{eco} in months, or in sites, where R_{eco} $\Delta^{14}C$ values drop below the current atmospheric value (Fig. 2a), as it is the only CO_2 source pool that can lower R_{eco} $\Delta^{14}C$ values below that of the current atmosphere (Supplementary Fig. 5b).

Significant temporal variation in R_{eco} $\Delta^{14}C$ values ($P < 0.001$) is thought to reflect changes in the relative contribution of plant metabolism, recent detritus, and old C. To constrain the contribution of the three C sources to R_{eco}, we employed a statistical isotope mass balance approach using the laboratory incubations of plants and soil to estimate the proportional contribution of each source[23] (Supplementary Information, and Supplementary Table 4). The mean proportional contribution of deep C to growing season R_{eco} generally ranged from 7 ± 4% to 23 ± 9% (mean ± s.d.) across all sites and years, with individual site means reaching as high as 41 ± 10% at the extensive thaw site in particular months (Fig. 2b; Supplementary Table 5). If we combine the mean proportional contributions from old C loss with R_{eco} flux values (Supplementary Table 2), then the extensive and moderate thaw sites are projected to have lost an average of 63 and 47 g C m^{-2} yr^{-1} respectively from the deep soil layers during the growing season; this was approximately 2 or 3 times greater than the deep C loss calculated for the minimal thaw site (22 g C m^{-2} yr^{-1}) (Fig. 3a). Applying a similar methodology to winter flux measurements increases the old C contribution to total annual R_{eco} at all sites, and reinforces the general pattern across sites of increasing old C loss with increased permafrost thaw (Supplementary Information, Supplementary Fig. 6).

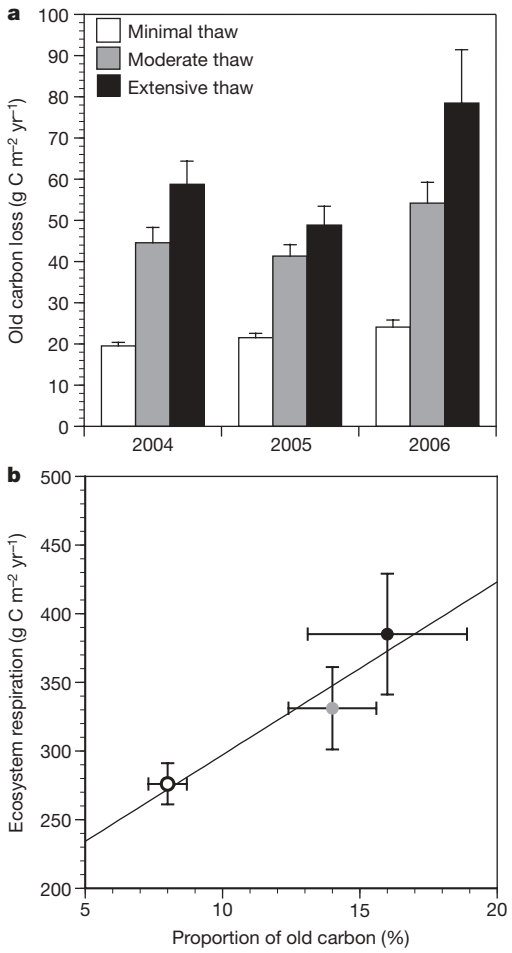


Figure 3 | Old carbon loss and its relationship to total ecosystem respiration for three sites that differ in the extent of permafrost thaw. **a**, Growing-season loss of old C from deeper in the soil profile, based on statistical partitioning estimates of mean proportional old C loss multiplied by R_{eco} flux measurements. Error bars represent the spatial variability of R_{eco} fluxes. **b**, The relationship between total R_{eco} and proportional old C loss for the growing season across sites. Error bars represent the interannual variability in C loss estimates; the regression line is shown for $n = 3$ sites.

Although our estimates of the proportional contribution of old C loss from statistical modelling can only be expressed as a possible range of values, the positive relationship observed between growing season R_{eco} and proportional old C loss across sites and years ($R^2 = 0.92$; Fig. 3b) supports the idea that the mean predicted contributions provide a reasonable picture of old C losses across sites. The relationship between these two independently estimated measurements across our gradient of sites shows that permafrost thaw and ground subsidence stimulates the release of old C, and that both the relative and absolute old C release increases with thawing. This observation confirms initial hypotheses regarding the response of permafrost C to warming[1,2,4–7,24]. More surprising is the finding that increased plant C uptake can offset the release of old C, at least in the initial decades of permafrost thaw.

How could these observed changes in flux have affected C stocks over time since thawing began at Eight Mile Lake? Our gradient of permafrost thaw provides some detail about decadal-scale trends. If C fluxes over time at a single site followed the trajectory observed across our gradient, the extensive thaw site would have accumulated an additional net 0.20 kg C m^{-2} over the first 15 years following thaw, even as old C deep in the soil was increasingly destabilized (Supplementary Information). Increased plant C uptake cannot fully offset continued increases in old C respiration, thus by the time the extensive thaw site reached the present-day state, it would have already lost an amount almost equivalent to the initial C gain (net gain since thawing initiated = 0.03 kg C m^{-2}) and perhaps more than double that amount (net loss 0.32 kg C m^{-2}), depending on the exact year that permafrost thaw initiated. Although these gains and losses are substantial, they are still small relative to the soil C stocks at this site, thus these changes would be impossible to detect with soil C stock measurements against background variability at this time.

Extending observed C exchange rates into the future must be done with caution, but provides insight into possible permafrost C losses and the feedback to climate change. Assuming that surface C gains and losses approach a dynamic equilibrium with the increased plant C uptake levels, the rate of old C loss at the extensive thaw site suggests that a net loss of 4.4–6.0 kg C m^{-2} is possible by the end of this century, or about 9.4–12.9% of the approximately 47 kg C m^{-2} contained in the 80-cm active layer of soil in the Eight Mile Lake watershed. Although our observations cannot account for future thawing that will expose a larger pool of permafrost C to decomposition, we can frame the observed C loss rate from our single gradient at Eight Mile watershed in a global context by applying this rate to the global surface permafrost C pool (818 Pg) to demonstrate that 0.8–1.1 Pg C yr^{-1} could be lost if surface permafrost thaws—as some models[25], albeit controversial[26], have predicted for this century (Supplementary Information). The actual emission rate will, of course, depend on future thaw rates, the forms of C gases released, and other positive and negative feedbacks to decomposition that may be expressed differently in the future or in different ecosystem types, such as changes in nutrient availability[27] and/or litter quality[28]. But the calculation is consistent with laboratory incubations of permafrost soil[29], and serves to illustrate that this biospheric feedback from permafrost C has the potential to be large as warming continues, and, at some point, could possibly be similar in magnitude to the current biospheric flux from land use change (1.5 ± 0.5 Pg C yr^{-1})[30].

METHODS SUMMARY

We measured net CO_2 exchange between tundra and the atmosphere over 3 years using a combination of static and automatic clear chambers. An infrared gas analyser measured chamber CO_2 concentrations to quantify ecosystem C exchange. Growing season measurements began within a week of snowmelt

(early May), and ended at snowfall (end of September). During the winter, flux measurements were made either at the snow surface or in shallow snow pits on a campaign basis that covered all winter months over the study period. Growing season, winter and annual C exchange were estimated using gap filling with response functions to measured environmental variables, and by interpolating mean estimates between time points using continuous environmental measurements to understand whether these tundra sites were gaining or losing C. Gross primary productivity was estimated as the difference between the integrated net ecosystem exchange and R_{eco} values.

At monthly intervals during the growing season, we also collected ecosystem respiration CO_2 from dark chambers, as well as soil CO_2 from soil profile gas wells, for $\Delta^{14}C$ analysis. In the laboratory, CO_2 was then purified and analysed for $\Delta^{14}C$ using an accelerator mass spectrometer, while a subsample was analysed for $\delta^{13}C$ using an isotope ratio mass spectrometer. Radiocarbon provides an indication of the age of respired C, and thus can be used as a fingerprint for identifying the decomposition of old organic C that has been stored in these permafrost soils. We used these field $\Delta^{14}C$ measurements in combination with $\Delta^{14}C$ measured from soil and plant incubations to estimate the contribution from soil and plants to ecosystem respiration. To do this, we used two and three pool isotope mixing models to constrain the proportional contribution of the component sources to total ecosystem respiration, in particular to determine the contribution of old C to ecosystem C exchange.

Received 24 August 2008; accepted 25 March 2009.

1. Schuur, E. A. G. *et al.* Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle. *Bioscience* **58,** 701–714 (2008).
2. Zimov, S. A., Schuur, E. A. G. & Chapin, F. S. Permafrost and the global carbon budget. *Science* **312,** 1612–1613 (2006).
3. Field, C. B., Sarmiento, J. & Hales, B. in *The First State of the Carbon Cycle Report (SOCCR) — Synthesis and Assessment Product 2.2* (eds King, A.W. *et al.)* 21–28 (National Oceanic and Atmospheric Administration, National Climatic Data Center, 2007).
4. Field, C. B., Lobell, D. B., Peters, H. A. & Chiariello, N. R. Feedbacks of terrestrial ecosystems to climate change. *Annu. Rev. Environ. Resour.* **32,** 1–29 (2007).
5. Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440,** 165–173 (2006).
6. Heimann, M. & Reichstein, M. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451,** 289–292 (2008).
7. Oechel, W. C. *et al.* Recent change of Arctic tundra ecosystems from a net carbon dioxide sink to a source. *Nature* **361,** 520–523 (1993).
8. Harden, J. W., Sundquist, E. T., Stallard, R. F. & Mark, R. K. Dynamics of soil carbon during deglaciation of the Laurentide ice sheet. *Science* **258,** 1921–1924 (1992).
9. Schirrmeister, L. *et al.* Paleoenvironmental and paleoclimatic records from permafrost deposits in the Arctic region of Northern Siberia. *Quat. Int.* **89,** 97–118 (2002).
10. Zimov, S. A. *et al.* Permafrost carbon: Stock and decomposability of a globally significant carbon pool. *Geophys. Res. Lett.* **33,** doi:10.1029/2006GL027484 (2006).
11. Smith, L. C. *et al.* Siberian peatlands a net carbon sink and global methane source since the early Holocene. *Science* **303,** 353–356 (2004).
12. Osterkamp, T. E. Characteristics of the recent warming of permafrost in Alaska. *J. Geophys. Res.* **112** (F2), doi:10.1029/2006JF000578 (2007).
13. Myneni, R. B., Tucker, C. J., Asrar, G. & Keeling, C. D. Interannual variations in satellite-sensed vegetation index data from 1981 to 1991. *J. Geophys. Res.* **103** (D6), 6145–6160 (1998).
14. Sturm, M., Racine, C. & Tape, K. Climate change – Increasing shrub abundance in the Arctic. *Nature* **411,** 546–547 (2001).
15. Chapin, F. S. *et al.* Role of land-surface changes in Arctic summer warming. *Science* **310,** 657–660 (2005).
16. Oechel, W. C. *et al.* Acclimation of ecosystem CO_2 exchange in the Alaskan Arctic in response to decadal climate warming. *Nature* **406,** 978–981 (2000).
17. Vitt, D. H., Halsey, L. A. & Zoltai, S. C. The changing landscape of Canada's western boreal forest: The current dynamics of permafrost. *Can. J. For. Res.* **30,** 283–287 (2000).
18. Osterkamp, T. E. The recent warming of permafrost in Alaska. *Glob. Planet. Change* **49,** 187–202 (2005).
19. Osterkamp, T. E. & Romanovsky, V. E. Evidence for warming and thawing of discontinuous permafrost in Alaska. *Permafrost Periglac. Process.* **10,** 17–37 (1999).
20. Schuur, E. A. G., Crummer, K. G., Vogel, J. G. & Mack, M. C. Plant species composition and productivity following permafrost thaw and thermokarst in Alaskan tundra. *Ecosystems* **10,** 280–292 (2007).
21. Trumbore, S. Age of soil organic matter and soil respiration: Radiocarbon constraints on belowground C dynamics. *Ecol. Appl.* **10,** 399–411 (2000).
22. Levin, I. & Hesshaimer, V. Radiocarbon – A unique tracer of global carbon cycle dynamics. *Radiocarbon* **42,** 69–80 (2000).
23. Phillips, D. L. & Gregg, J. W. Source partitioning using stable isotopes: Coping with too many sources. *Oecologia* **136,** 261–269 (2003).
24. Shaver, G. R. *et al.* Global change and the carbon balance of Arctic ecosystems. *Bioscience* **42,** 433–441 (1992).
25. Lawrence, D. M. & Slater, A. G. A projection of severe near-surface permafrost degradation during the 21st century. *Geophys. Res. Lett.* **32,** L24401, doi:10.1029/2005GL025080 (2005).
26. Delisle, G. Near-surface permafrost degradation: How severe during the 21st century? *Geophys. Res. Lett.* **34** (9) doi:10.1029/2007GL029323 (2007).
27. Mack, M. C. *et al.* Ecosystem carbon storage in arctic tundra reduced by long-term nutrient fertilization. *Nature* **431,** 440–443 (2004).
28. Hobbie, S. E. Temperature and plant species control over litter decomposition in Alaskan tundra. *Ecol. Monogr.* **66,** 503–522 (1996).
29. Dutta, K., Schuur, E. A. G., Neff, J. C. & Zimov, S. A. Potential carbon release from permafrost soils of Northeastern Siberia. *Glob. Change Biol.* **12,** 2336–2351 (2006).
30. Canadell, J. G. *et al.* Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl Acad. Sci. USA* **104,** 18866–18870 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was made possible by assistance from G. Adema, T. Chapin, S. DeBiasio, L. Gutierrez, M. Mack, M. Schieber, E. Tissier, C. Trucco, W. Vicars, E. Wilson, C. Wuthrich, L. Yocum, and the researchers and technicians of the Bonanza Creek LTER. This work relied on funds from the following sources: NASA New Investigator Program, NSF Bonanza Creek LTER Program, NSF DEB Ecosystems Program, and a cooperative agreement with the National Park Service.

Author Contributions E.A.G.S. conceived the experiment. E.A.G.S. and J.G.V. designed the experiment and wrote the paper. E.A.G.S., J.G.V., K.G.C. and H.L. performed research. All authors commented on the analysis and presentation of the data and were involved in the writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.A.G.S. (tschuur@ufl.edu).

Vol 459|28 May 2009|doi:10.1038/nature08018

LETTERS

Preserving noble gases in a convecting mantle

Helge M. Gonnermann[1] & Sujoy Mukhopadhyay[2]

High ^{3}He/^{4}He ratios sampled at many ocean islands are usually attributed to an essentially undegassed lower-mantle reservoir with high ^{3}He concentrations[1–4]. A large and mostly undegassed mantle reservoir is also required to balance the Earth's ^{40}Ar budget, because only half of the ^{40}Ar produced from the radioactive decay of ^{40}K is accounted for by the atmosphere and upper mantle[5]. However, geophysical[6,7] and geochemical observations[8] suggest slab subduction into the lower mantle, implying that most or all of Earth's mantle should have been processed by partial melting beneath mid-ocean ridges and hotspot volcanoes. This should have left noble gases in both the upper and the lower mantle extensively outgassed, contrary to expectations from ^{3}He/^{4}He ratios and the Earth's ^{40}Ar budget. Here we suggest a simple solution: recycling and mixing of noble-gas-depleted slabs dilutes the concentrations of noble gases in the mantle, thereby decreasing the rate of mantle degassing and leaving significant amounts of noble gases in the processed mantle. As a result, even when the mass flux across the 660-km seismic discontinuity is equivalent to approximately one lower-mantle mass over the Earth's history, high ^{3}He contents, high ^{3}He/^{4}He ratios and ^{40}Ar concentrations high enough to satisfy the ^{40}Ar mass balance of the Earth can be preserved in the lower mantle. The differences in ^{3}He/^{4}He ratios between mid-ocean-ridge basalts and ocean island basalts, as well as high concentrations of ^{3}He and ^{40}Ar in the mantle source of ocean island basalts[4], can be explained within the framework of different processing rates for the upper and the lower mantle. Hence, to preserve primitive noble gas signatures, we find no need for hidden reservoirs or convective isolation of the lower mantle for any length of time.

The requirement of a ^{3}He- and ^{40}Ar-rich reservoir provides the strongest arguments for a relatively undegassed lower mantle[1–3,5,9]. However, seismological observations of plate subduction into the lower mantle[6,7] appear to be inconsistent with a relatively undegassed lower mantle. To reconcile the high ^{3}He/^{4}He ratios observed in ocean island basalts (OIBs) with the expectation of a processed and outgassed lower-mantle source, residues of mantle melting[10,11], depleted in uranium and thorium relative to ^{3}He, have been suggested to carry the high ^{3}He/^{4}He signatures. However, the preservation of high ^{3}He/^{4}He ratios in residues of mantle melting is apparently inconsistent with recently determined noble gas partition coefficients[12]. Because helium is nevertheless incompatible, the low concentrations in residual mantle require convective isolation of the plume source over most of the Earth's lifetime. Convective isolation over geologically long periods (>1 Gyr) has also been suggested in other models for the evolution of helium in the OIB mantle source[13].

By contrast, we find that the solution to the apparent discrepancy between an outgassed lower mantle and high concentrations of primordial ^{3}He in the lower mantle lies in the ubiquitous process of recycling and mixing of degassed slabs back into the mantle. During partial melting, noble gases behave incompatibly and strongly partition into the basaltic melt, which forms the ocean crust and leaves the residual mantle severely depleted of noble gases. However, ocean crust is extensively degassed and when slabs, comprised of the degassed oceanic crust and residual mantle, are recycled back into the mantle, they contain negligible mantle-derived helium[14]. Mixing of these ^{3}He-depleted slabs with ambient mantle will dilute average helium concentrations in the mixed mantle assemblage. Because the rate at which noble gases are lost from the mantle is proportional to the gas concentration within the mantle being processed by melting, the monotonically decreasing noble gas concentrations in the evolving mixed mantle assemblage will result in a decreasing outgassing rate at any given processing rate (Fig. 1). In the simplest case of slabs mixing instantaneously with mantle, the average concentration of a primordial noble gas, such as ^{3}He, in a convectively mixed mantle assemblage evolves as

$$\bar{C}(t) = \mathrm{e}^{(f-1)qt} \approx \mathrm{e}^{-qt} \qquad (1)$$

Here $\bar{C} = C/C_0$, with C the concentration in the mantle reservoir at time t and C_0 the initial concentration, $f \approx 0$ is the fraction of a given noble gas that remains in the mantle residue after melting and q is the rate at which the mantle reservoir is processed through melting over 4.5 Gyr. Present-day concentrations are given by e^{-N}, where N is the number of times the reservoir has been processed through partial melting over 4.5 Gyr. This simple analytic expression demonstrates that even after one reservoir mass has undergone processing by partial melting ($N = 1$), ~37% of primordial ^{3}He is retained in the reservoir (Fig. 2, dashed line).

Mixing of subducted slabs with ambient mantle is not instantaneous, but is thought to have an exponential dependence on time[15,16]. Convective motions stretch and juxtapose recycled slab and ambient mantle at increasingly smaller length scales over a characteristic mixing time, τ, which may range between a few hundred million years to a few billion years[15–17]. Partial melting of this mixed mantle assemblage will extract noble gases in proportion to their average concentrations within the melting region. If mixing of recycled slab to length scales that are smaller than the melting region occurs at a rate proportional to $q\mathrm{e}^{-s/\tau}$, where s is the time after slab formation, then (Methods)

$$\bar{C} = \exp\left[\frac{1}{q\tau - 1}\ln\left(1 - q\tau + q\tau \mathrm{e}^{-t/\tau}\right) + \frac{qt}{q\tau - 1}\right] \qquad (2)$$

Figure 2 shows that for any reasonable estimate of τ, at least 30% of primordial ^{3}He will be retained in a mantle reservoir even after one reservoir mass has been processed. Therefore, preserving high concentrations of ^{3}He in the Earth's lower mantle does not require convective isolation or inefficient mixing. Because incompatible trace elements have been preferentially sequestered in the continental crust, a processed ^{3}He-rich lower mantle should have depleted strontium and neodymium isotopic characteristics relative to compositions of bulk silicate Earth (BSE), consistent with the mantle component that carries the signature of high ^{3}He/^{4}He ratios in OIBs (the focus zone, or FOZO)[13,18]. A mantle reservoir that has been processed repeatedly

[1]Department of Earth Science, Rice University, Houston, Texas 77005, USA. [2]Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

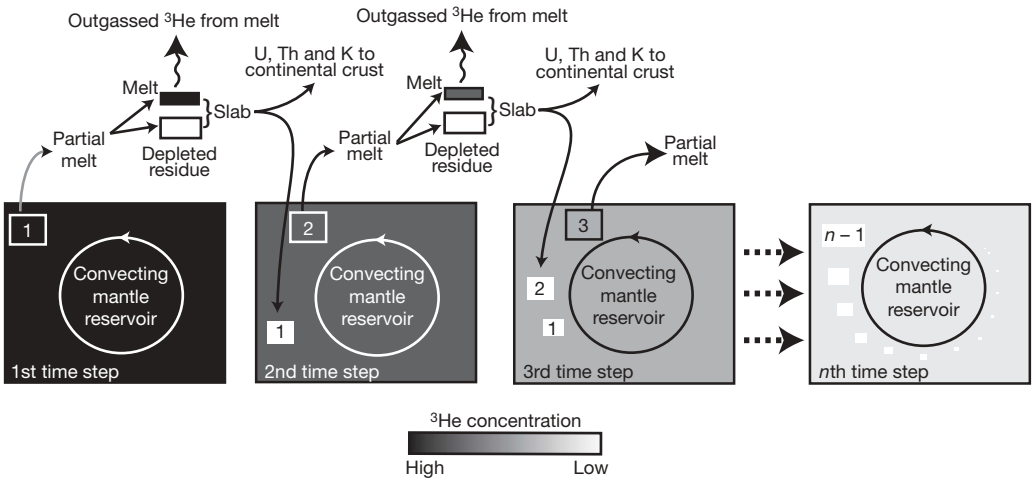


Figure 1 | Conceptual model of mantle degassing. Idealized processing of a mantle reservoir through partial melting, shown schematically in discrete time steps. Uranium, thorium, potassium and the noble gases (^{3}He shown here) are incompatible and partition preferentially into the basaltic melt phase, leaving a depleted mantle residue. The basaltic melt degasses before it solidifies to form ocean crust, but remains enriched in incompatible elements. Ocean crust plus mantle residue comprise slab, which is subducted back into the mantle. During subduction, any residual helium is lost from the slab[14] and some of the uranium, thorium and potassium are extracted from the slab and incorporated into the continental crust. The noble-gas-depleted slab eventually becomes mixed in with the remaining mantle reservoir over a characteristic timescale, τ. Because of slab recycling, ^{3}He concentrations in the mixed mantle assemblage decrease over time and the rate of ^{3}He degassing at a given processing rate decreases monotonically. Consequently, a significant fraction of ^{3}He is retained, despite extensive processing.

over the Earth's history ($N = 2$–4) is considerably more degassed, but, depending on τ, can still contain a few per cent primordial ^{3}He. This is consistent with the ^{3}He/ ^{4}He of the mid-ocean-ridge basalt (MORB) mantle source.

To test under what conditions mixing of noble-gas-depleted slabs back into the mantle is quantitatively consistent with observed ^{3}He/^{4}He ratios, ^{3}He concentrations and the Earth's ^{40}Ar budget, we use geochemical reservoir modelling[19–21] (see Methods and Supplementary Information for a detailed description). We model two mantle reservoirs corresponding in mass to the Earth's upper mantle (MORB source) and lower mantle (OIB source), respectively. We start with both the upper and the lower mantle being homogeneous and having BSE compositions (Fig. 1). The reservoirs degas by processing through partial melting near the Earth's surface to produce oceanic crust and an underlying residual mantle, which together comprise the slab. We assume that the ocean crust loses all helium and argon before being subducted back into the mantle, which introduces chemical heterogeneity into the mantle. The recycled slabs mix with the ambient mantle (Fig. 1) on a characteristic timescale, τ, which can differ between the upper and the lower mantle and is allowed to vary from 0–4 Gyr. Thus, the lower- and upper-mantle reservoirs evolve from a homogeneous reservoir to one that contains a well-mixed assemblage and unmixed slabs (Fig. 1), the amount of which depends on τ.

In our model, we assume that only the mixed mantle assemblage undergoes processing by partial melting and we allow the fraction of slab that subducts into the lower mantle to vary from zero to one. The mass flux of slab into the lower mantle is balanced by a subsequent

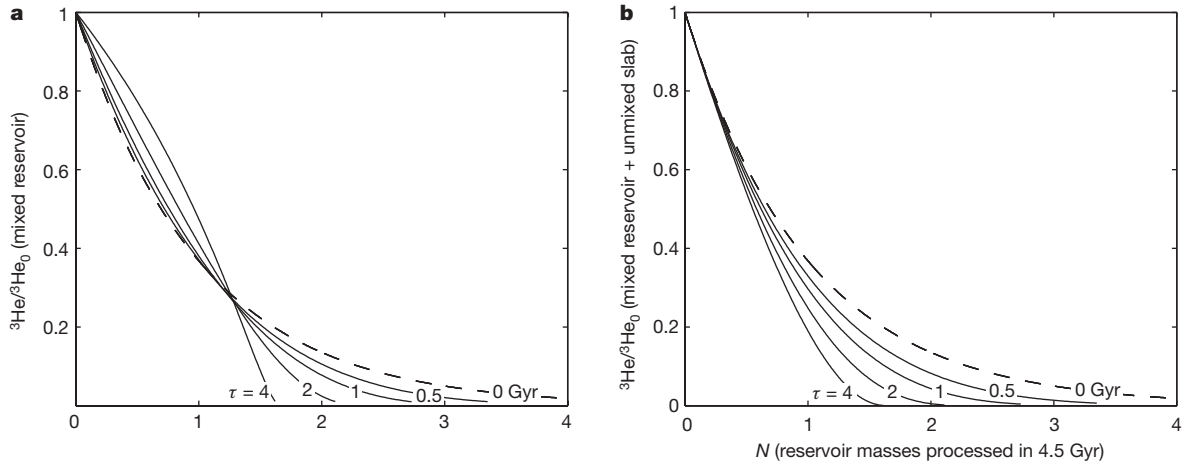


Figure 2 | Primordial noble gas concentration in the processed mantle reservoir. a, Concentration of ^{3}He in the mixed mantle reservoir after 4.5 Gyr (normalized to the initial concentration, ^{3}He$_0$), plotted as a function of the number of times the reservoir has been processed, N. Each curve corresponds to a given value of τ. The case of instantaneous mixing of slabs with the ambient mantle is represented by $\tau = 0$ Gyr, shown as the dashed curve. Even for values of τ greater than 1 Gyr, processing of a mantle reservoir at a rate of one reservoir mass over the Earth's history ($N = 1$) leads to the concentration of primordial nobles gases in the mixed mantle reservoir being at least 30% of its initial value. By contrast, repeated processing ($N \geq 2$) will result in significant degassing and primordial helium concentrations of about 1% of initial values. With the exception of the case in which $\tau = 0$ Gyr, the intersection of the curves with the horizontal axis (^{3}He/^{4}He = 0) represents the point at which the entire mantle reservoir consists only of unmixed slabs. We note that for $N < 1$ and large values of τ, little of the recycled slab has mixed back into the mantle. **b**, Same as in **a**, but for the entire mantle reservoir (mixed reservoir plus unmixed slab). Over a wide range of conditions, the total ^{3}He retained within a convecting mantle reservoir is a substantial fraction of the initial ^{3}He. Concentrations calculated using equations (1) and (2).

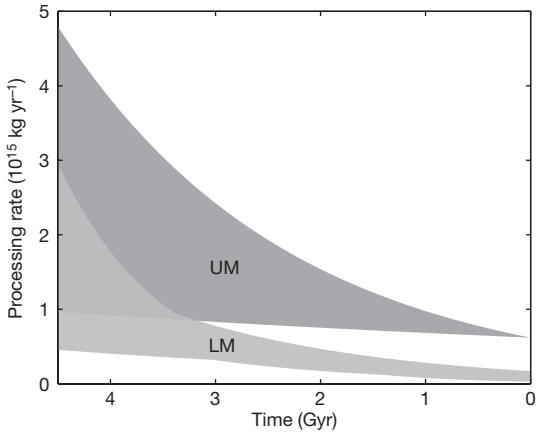


Figure 3 | Mantle processing rate. Range of instantaneous mantle processing rate as a function of time for the upper mantle (UM) and the lower mantle (LM), for models that meet geochemical constraints. The processing rate decreases exponentially in time to present-day processing rates beneath mid-ocean ridges (7×10^{14} kg y^{-1}) for UM[26], and to plume flux estimates ($(2-20) \times 10^{13}$ kg y^{-1}) for LM[27]. The shaded areas represent the regions of overlap for individual curves; any kinks in the shaded areas are a consequence of intersecting curves.

return flow out of the lower mantle. This is either processed directly, by partial melting, presumably at hotspots and large igneous provinces, or a fraction of it is mixed into the upper mantle with only the remainder being directly processed. The residual mantle generated from partial melting of upwelling lower mantle can mix directly with the upper mantle or become part of the slab. We model a wide range of mantle processing rates that decrease exponentially in time[15] to estimated present-day mid-ocean-ridge processing rates for the upper mantle and plume flux rates for the lower mantle (Fig. 3).

Within the framework of our model, we find that the high ^{3}He/^{4}He ratios measured in OIBs are entirely consistent with a lower-mantle source that has been depleted over the Earth's lifetime. Recycling of slabs into the lower mantle results in processing of 0.5–1 lower-mantle masses over the Earth's history, and the present day ^{3}He concentrations in the mixed regions of the lower mantle are consistent with estimates from OIB degassing models[4] (Fig. 4). The well-mixed lower-mantle assemblage has a non-primordial neodymium isotopic composition (Supplementary Information) and contains the high ^{40}Ar concentrations that are required to balance the Earth's ^{40}Ar budget. Hence, we find no argon concentration paradox[5]. The amount of upper mantle processed over the Earth's history falls within the range of 2.5–7.5 upper-mantle masses, consistent with a more depleted upper mantle. Characteristic mixing times for the lower mantle are up to ~2 Gyr and less than 1 Gyr for the upper mantle. Model results predict that at present ~30% of slabs subduct into the lower mantle with an average of ~50% over the Earth's history. Direct mixing of lower mantle with upper mantle comprises less than 10% of the lower-mantle flux across the 660-km seismic discontinuity.

These results can be interpreted to indicate that over much of the Earth's history, slab subduction into the lower mantle[6,7] is balanced by a lower-mantle return flow that rises with little mixing through the upper mantle to shallow depths where partial melting results in outgassing and produces large igneous provinces, hotspot volcanoes or plume-influenced MORBs. Consequently, there is limited direct input of noble gases from the lower to the upper mantle. Over the Earth's history, the lower-mantle return flow results in a mass exchange of approximately one lower-mantle mass across the 660-km discontinuity. Hence, the noble gas view suggests that up to 50% of slabs produced over the Earth's history have subducted into the lower mantle across a 660-km discontinuity that opposes unfettered mixing between the upper and the lower mantle. This is consistent with the seismic evidence that slab stagnation is common in the transition zone[22] and with the significantly different spatial patterns of shear-wave velocity anomalies on either side of the 660-km discontinuity, which indicate limited direct coupling of flow between the upper and the lower mantle[23,24].

From our analytic treatment (Fig. 2) and geochemical reservoir modelling (Fig. 4), we find that the recycling and mixing of outgassed slabs allows the preservation of high noble gas concentrations in a processed mantle reservoir. We find no noble gas requirements for convective isolation or hidden reservoirs in the lower mantle. The differences in measured ^{3}He/^{4}He ratios, as well as helium and argon concentrations, between MORBs and OIBs can be explained by an upper mantle that has been processed several times with respect to its mass, whereas the lower mantle has been processed approximately

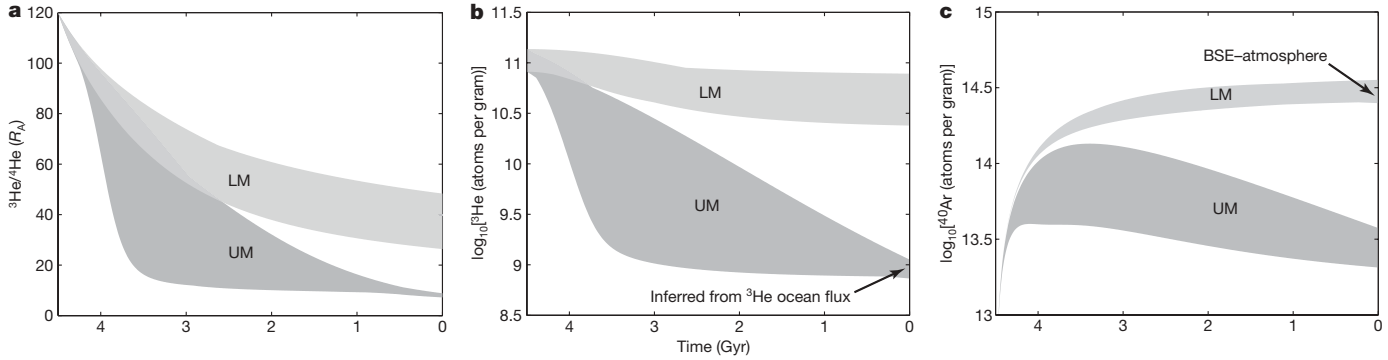


Figure 4 | Results from geochemical reservoir modelling. Initial ^{3}He/^{4}He ratios are $120R_A$, where R_A denotes the atmospheric ^{3}He/^{4}He ratio[28]. The initial ^{3}He concentration is $(7.5–15) \times 10^{10}$ atoms per gram, within the range of previous estimates[3,26], and there is no primordial ^{40}Ar. Uranium, thorium and potassium concentrations at 4.5 Gyr are BSE values for both UM and LM (^{238}U, 40 p.p.b.; ^{235}U, 12 p.p.b.; ^{232}Th, 95 p.p.b.; ^{40}K, 327 p.p.b.). The shaded areas represent model simulations where present-day upper-mantle ^{3}He/^{4}He ratios are $(8.5 \pm 0.5)R_A$; uranium and potassium concentrations are within 3.2–5.4 p.p.b. and 50–68.4 p.p.m., respectively[29,30]; lower-mantle ^{3}He/^{4}He ratios are larger than $35R_A$; the predicted ^{3}He outgassing rate at mid-ocean ridges is within 75–100% of the estimated ^{3}He flux to the oceans[31] (1,060 mol yr^{-1}); and approximately 40–60% of radiogenic ^{40}Ar produced over 4.5 Gyr still resides within the Earth's mantle. **a**, Modelled evolution in ^{3}He/^{4}He ratios in the mixed UM and LM reservoirs, as a function of time. **b**, The resultant range and evolution of ^{3}He in the mantle reservoirs. **c**, Predicted range and evolution of radiogenic ^{40}Ar for zero primordial ^{40}Ar. Despite processing of up to one LM mass, present-day ^{40}Ar in the LM ranges between approximately 2.5×10^{14} and 3.6×10^{14} atoms per gram, which accounts for 40–60% of the ^{40}Ar produced over the Earth's history, with most of the remaining ^{40}Ar outgassed to the atmosphere. This is consistent with the Earth's atmospheric ^{40}Ar budget. Because of repeated processing, the ^{40}Ar concentration of the UM is substantially lower, ranging between approximately 2.1×10^{13} and 3.8×10^{13} atoms per gram. This is within the range of values predicted from the measured ^{3}He flux to the Earth's oceans, assuming a ^{3}He/^{4}He ratio of about $8.5R_A$ and a production ratio of three[28] for ^{4}He/^{40}Ar.

once. A simple explanation for this is that the upper mantle is efficiently processed on timescales of 0.5–1 Gyr as a consequence of mid-ocean-ridge migration[25] and relatively low viscosity, whereas the lower mantle is inefficiently processed as a result of its higher viscosity and because mantle flow across the 660-km discontinuity is impeded[22–24].

High concentrations of ^{40}Ar and ^{3}He and ^{3}He/^{4}He ratios of about $50R_A$, noble gas characteristics that are normally attributed to a primitive mantle or hidden reservoirs, can be preserved in a convecting and processed lower mantle. Moreover, these noble gas characteristics are expected to be associated with depleted strontium and neodymium isotopic compositions, which are characteristic of the mantle component FOZO[13,18] (Supplementary Information). Although FOZO is often thought of as an ancient mantle reservoir[18], our modelling suggests that its noble gas signature can be attributed to a depleted mantle reservoir that has continuously evolved through slab recycling. A continuum of compositions between relatively well-mixed lower mantle assemblage (FOZO) and incompletely mixed recycled slabs is likely to produce the observed variability of ^{3}He/^{4}He ratios in OIBs.

METHODS SUMMARY

The free parameters in our geochemical reservoir model are the τ values for the upper and the lower mantle, mantle processing rates through time, rates of sequestration of uranium, thorium and potassium in the continental crust, the fraction of slab subducted into the lower mantle, the amount of direct mixing of the lower mantle with the upper mantle and the fraction of residual plume mantle that mixes directly with the upper mantle. We use Monte Carlo simulations in which free parameters are allowed to vary independently to determine their permissible range with respect to observational constraints. For each simulation, the mass balances for ^{3}He, ^{4}He, ^{40}Ar, uranium, thorium and potassium in the continental crust and in the well-mixed and unmixed slabs of the upper- and lower-mantle reservoirs are calculated as functions of time by numerical integration (see Methods and Supplementary Information for details). The model is constrained by a set of initial conditions (the initial ^{3}He/^{4}He ratio and the initial ^{3}He, ^{40}Ar, uranium, thorium and potassium concentrations) and a set of present-day observations (the upper-mantle ^{3}He concentration, the ^{3}He/^{4}He ratios and ^{40}Ar concentrations in the upper and the lower mantle, the upper-mantle processing rate as given by global ocean-crust production rate, lower-mantle processing rates as given by the plume mass flux and present-day uranium and potassium concentrations in the depleted MORB mantle. The vast majority of model runs do not meet the present-day observational constraints, and only successful runs that simultaneous satisfy all such constraints are shown in Fig. 4.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 December 2008; accepted 24 March 2009.

1. Kurz, M. D., Jenkins, W. J. & Hart, S. R. Helium isotopic systematics of oceanic islands and mantle heterogeneity. *Nature* **297,** 43–47 (1982).
2. Allègre, C. J., Staudacher, T. & Sarda, P. Rare-gas systematics: formation of the atmosphere, evolution and structure of the Earth's mantle. *Earth Planet. Sci. Lett.* **81,** 127–150 (1987).
3. Harper, C. L. & Jacobsen, S. B. Noble gases and Earth's accretion. *Science* **273,** 1814–1818 (1996).
4. Gonnermann, H. M. & Mukhopadhyay, S. Non-equilibrium degassing and a primordial source for helium in ocean-island volcanism. *Nature* **449,** 1037–1040 (2007).
5. Allègre, C. J., Hofmann, A. & O'Nions, K. The argon constraints on mantle structure. *Geophys. Res. Lett.* **23,** 3555–3557 (1996).
6. Grand, S. P., van der Hilst, R. D. & Widiyantoro, S. Global seismic tomography: a snapshot of convection in the Earth. *GSA Today* **7,** 1–7 (1997).
7. van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386,** 578–584 (1997).
8. Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. Nb and Pb in oceanic basalts - new constraints on mantle evolution. *Earth Planet. Sci. Lett.* **79,** 33–45 (1986).
9. O'Nions, R. K. & Tolstikhin, I. N. Limits on the mass flux between lower and upper mantle and stability of layering. *Earth Planet. Sci. Lett.* **139,** 213–222 (1996).
10. Albarède, F. Time-dependent models of U-Th-He and K-Ar evolution and the layering of mantle convection. *Chem. Geol.* **145,** 413–429 (1998).
11. Parman, S. W. Helium isotopic evidence for episodic mantle melting and crustal growth. *Nature* **446,** 900–903 (2007).
12. Heber, V. S., Brooker, R. A., Kelley, S. P. & wood, B. J. Crystal-melt partitioning of noble gases (helium, neon, argon, krypton, and xenon) for olivine and clinopyroxene. *Geochim. Cosmochim. Acta* **71,** 1041–1061 (2007).
13. Class, C. & Goldstein, S. L. Evolution of helium isotopes in the Earth's mantle. *Nature* **436,** 1107–1112 (2005).
14. Staudacher, T. & Allègre, C. J. Recycling of oceanic-crust and sediments—the noble-gas subduction barrier. *Earth Planet. Sci. Lett.* **89,** 173–183 (1988).
15. Allègre, C. J. & Turcotte, D. L. Implications of a two-component marble-cake mantle. *Nature* **323,** 123–127 (1986).
16. Coltice, N. & Schmalzl, J. Mixing times in the mantle of the early Earth derived from 2-D and 3-D numerical simulations of convection. *Geophys. Res. Lett.* **33,** doi:10.1029/2006GL027707 (2006).
17. Albarede, F. Rogue mantle helium and neon. *Science* **319,** 943–945 (2008).
18. Hart, S. R., Hauri, E. H., Oschmann, L. A. & Whitehead, J. A. Mantle plumes and entrainment - isotopic evidence. *Science* **256,** 517–520 (1992).
19. Jacobsen, S. B. & Wasserburg, G. J. Mean age of mantle and crustal reservoirs. *J. Geophys. Res.* **84,** 7411–7427 (1979).
20. Albarède, F. Radiogenic ingrowth in systems with multiple reservoirs: applications to the differentiation of the mantle-crust system. *Earth Planet. Sci. Lett.* **189,** 59–73 (2001).
21. Kellogg, J. B., Jacobsen, S. B. & O'Connell, R. J. Modeling the distribution of isotopic ratios in geochemical reservoirs. *Earth Planet. Sci. Lett.* **204,** 183–202 (2002).
22. Fukao, Y., Widiyantoro, S. & Obayashi, M. Stagnant slabs in the upper and lower mantle transition region. *Rev. Geophys.* **39,** 291–323 (2001).
23. Gu, Y. J., Dziewoński, A. M., Su, W. & Ekström, G. Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities. *J. Geophys. Res.* **106,** 11169–11199 (2001).
24. Kustowski, B., Ekström, G. & Dziewoński, A. M. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model. *J. Geophys. Res.* **113,** doi:10.1029/2007JB005169 (2008).
25. O'Connell, R. J. & Hager, B. H. Ridge migration and mantle differentiation. *Eos* **61,** 373 (1980).
26. Porcelli, D. & Elliott, T. The evolution of He isotopes in the convecting mantle and the preservation of high ^{3}He/^{4}He ratios. *Earth Planet. Sci. Lett.* **269,** 175–185 (2008).
27. Kellogg, L. H. & Wasserburg, G. J. The role of plumes in mantle helium fluxes. *Earth Planet. Sci. Lett.* **99,** 276–289 (1990).
28. Graham, D. W. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. & Wieler, R.) 247–317 (Rev. Mineral. Geochem. Vol. 47, Mineralogical Society of America, 2002).
29. Workman, R. K. & Hart, S. R. Major and trace element composition of the depleted MORB mantle (DMM). *Earth Planet. Sci. Lett.* **231,** 53–72 (2005).
30. Boyet, M. & Carlson, R. W. A new geochemical model for the Earth's mantle inferred from ^{146}Sm–^{142}Nd systematics. *Earth Planet. Sci. Lett.* **250,** 254–268 (2006).
31. Farley, K. A., Maierreimer, E., Schlosser, P. & Broecker, W. S. Constraints on mantle ^{3}He fluxes and deep-sea circulation from an oceanic general-circulation model. *J. Geophys. Res.* **100,** 3829–3839 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. O'Connell, A. Dziewonski, C. Langmuir, E. Hellebrand for discussions and D. Graham for a thorough review of the manuscript. H.M.G. was in part supported by the University of Hawaii, SOEST Young Investigator programme. S.M. was in part supported by US National Science Foundation grant EAR 0509721.

Author Contributions H.M.G. and S.M. together developed the ideas presented here. H.M.G. developed the geochemical reservoir model and performed the numerical modelling. H.M.G. and S.M. analysed the results and co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.M.G. (helge@rice.edu) or S.M. (sujoy@eps.harvard.edu).

METHODS

Instantaneous mixing. The principal parameter that controls outgassing is the amount of processing that a given mantle reservoir undergoes over the Earth's history. If the depletion of a non-radiogenic noble gas is calculated for a mantle reservoir, under the assumption of 'instantaneous' recycling (instant mixing of the processed and degassed slab with the ambient reservoir), then the rate of change in concentration within this mixed reservoir is given by

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -C(1-f)q$$

If q is a constant, integrating the equation gives us

$$C(t) = C_0 \mathrm{e}^{-(1-f)qt}$$

For any feasible helium partition coefficient and melting model, and because helium is degassed from the oceanic crust, the $f = 0$ approximation is reasonable, resulting in

$$C(t) \approx C_0 \mathrm{e}^{-qt}$$

or

$$C(t = 4.5\ \mathrm{Gyr}) \approx C_0 \mathrm{e}^{-N}$$

Inefficient mixing. Mixing in the mantle is not likely to be instantaneous. An analytic formulation for a finite mixing time can be derived on the assumption that $f = 0$ and that complete degassing of the melt fraction occurs. We assume that the fraction of slab produced at time s and remaining unmixed at time t is given by $\mathrm{e}^{(s-t)/\tau}$, where τ is the characteristic mixing time. The cumulative mass of slab (normalized by the initial reservoir mass, M_0) produced at time t and remaining unmixed is therefore given by

$$\frac{M_\mathrm{s}}{M_0} = \int_0^t q\mathrm{e}^{(s-t)/\tau}\mathrm{d}s = q\tau - q\tau\mathrm{e}^{-t/\tau}$$

The complementary mass of the mixed mantle reservoir, M, is

$$\frac{M}{M_0} = 1 - \int_0^t q\mathrm{e}^{(s-t)/\tau}\mathrm{d}s = 1 - q\tau + q\tau\mathrm{e}^{-t/\tau}$$

The rate of change in mass of the mixed reservoir at time t is given by

$$\frac{1}{M_0}\frac{\mathrm{d}M}{\mathrm{d}t} = -q\mathrm{e}^{-t/\tau}$$

Using this expression, we find the rate of change in concentration of the mixed reservoir to be

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \frac{C}{M}(q\mathrm{e}^{-t/\tau} - q)$$

which upon integration results in

$$C = C_0 \exp\left[-\int \frac{q - q\mathrm{e}^{-t/\tau}}{1 - \tau q + \tau q\mathrm{e}^{-t/\tau}}\mathrm{d}t\right]$$

or

$$C = C_0 \exp\left[\frac{qt + \ln(1 - \tau q + \tau q\mathrm{e}^{-t/\tau})}{\tau q - 1}\right]$$

Mass-balance formulation for reservoir model. Please see Supplementary Fig. S4 for an illustration of the conceptual model. The mass lost from the mixed lower mantle at time step i is given by

$$\Delta M_{\mathrm{L},i} = Q_{\mathrm{L},i}\Delta t$$

and that lost from the upper mantle by

$$\Delta M_{\mathrm{U},i} = Q_{\mathrm{U},i}\Delta t$$

We define Δt as the duration of the time step and Q as the mass processing rate, given by

$$Q = Q_\mathrm{present}\mathrm{e}^{\alpha t}$$

Here t is time before present normalized to 4.5 Gyr and α is a parameter that varies independently for the upper and the lower mantle and for each simulation.

The mass of lower mantle that mixes with upper mantle at time step i is

$$\Delta M_{\mathrm{LU},i} = Q_{\mathrm{P},i}\Delta t$$

where $Q_{\mathrm{P},i}$ is the plume flux. The mass of residue generated from partial melting of plumes that mixes directly with upper mantle at time step i is

$$\Delta M_{\mathrm{PU},i} = x_{\mathrm{P},i}(1-F)Q_{\mathrm{P},i}\Delta t$$

where x_P is the fraction of the residue generated from plume melting that mixes directly with the upper mantle and F is the melt fraction. The mass of ocean crust produced at time step i is

$$\Delta M_{\mathrm{O},i} = F(Q_{\mathrm{P},i} + Q_{\mathrm{U},i})\Delta t$$

and the mass of residual mantle that becomes part of the slab at time step i is

$$\Delta M_{\mathrm{R},i} = (1-F)\left((1-x_{\mathrm{P},i})Q_{\mathrm{P},i} + Q_{\mathrm{U},i}\right)\Delta t$$

Ocean crust and residual mantle comprise slab and the mass of slab produced at time step i equals $M_{\mathrm{O},i} + M_{\mathrm{R},i}$. The mass of slab subducted into upper mantle at time step i is

$$\Delta M_{\mathrm{SU},i} = (1-x_\mathrm{S})Q_{\mathrm{S},i}\Delta t$$

whereas that subducted into lower mantle is

$$\Delta M_{\mathrm{SL},i} = x_\mathrm{S}Q_{\mathrm{S},i}\Delta t$$

Here x_S is the fraction of slabs subducted into the lower mantle and Q_S is the mass flux of subducting slab. The mass of slab subducted at time steps $j = 1, 2, \ldots, i$ that becomes mixed with the mixed subreservoir of the upper mantle at time step i is

$$\delta M_{\mathrm{SU},i} = \sum_{j=1}^{i} \Delta M_{\mathrm{SU},j}(\mathrm{e}^{-(i-j-1)\Delta t/\tau} - \mathrm{e}^{-(i-j)\Delta t/\tau})$$

and the mass of slab subducted at time steps $j = 1, 2, \ldots, i$ that becomes mixed with the mixed subreservoir of the lower mantle at time step i is

$$\delta M_{\mathrm{SL},i} = \sum_{j=1}^{i} \Delta M_{\mathrm{SL},j}(\mathrm{e}^{-(i-j-1)\Delta t/\tau} - \mathrm{e}^{-(i-j)\Delta t/\tau})$$

The concentrations of uranium, thorium, potassium, helium and argon as functions of time for each of the reservoirs follow directly from these mass balances. Additional considerations that are incorporated with these mass balances are radioactive decay of uranium, thorium and potassium and concurrent production of radiogenic ^{4}He and ^{40}Ar. These are based on the standard decay equations for these elements. The fractionation of radioactive elements and noble gases between melt (ocean crust) and residual mantle are based on the melting model discussed in Supplementary Information, with the added caveat that all noble gases are lost from the melt to the atmosphere and a fraction, x_C, of radioactive elements is sequestered from ocean crust into continental crust upon subduction.

LETTERS

De novo establishment of wild-type song culture in the zebra finch

Olga Fehér[1], Haibin Wang[2], Sigal Saar[1], Partha P. Mitra[2] & Ofer Tchernichovski[1]

Culture is typically viewed as consisting of traits inherited epigenetically, through social learning. However, cultural diversity has species-typical constraints[1], presumably of genetic origin. A celebrated, if contentious, example is whether a universal grammar constrains syntactic diversity in human languages[2]. Oscine songbirds exhibit song learning and provide biologically tractable models of culture: members of a species show individual variation in song[3] and geographically separated groups have local song dialects[4,5]. Different species exhibit distinct song cultures[6,7], suggestive of genetic constraints[8,9]. Without such constraints, innovations and copying errors should cause unbounded variation over multiple generations or geographical distance, contrary to observations[9]. Here we report an experiment designed to determine whether wild-type song culture might emerge over multiple generations in an isolated colony founded by isolates, and, if so, how this might happen and what type of social environment is required[10]. Zebra finch isolates, unexposed to singing males during development, produce song with characteristics that differ from the wild-type song found in laboratory[11] or natural colonies. In tutoring lineages starting from isolate founders, we quantified alterations in song across tutoring generations in two social environments: tutor–pupil pairs in sound-isolated chambers and an isolated semi-natural colony. In both settings, juveniles imitated the isolate tutors but changed certain characteristics of the songs. These alterations accumulated over learning generations. Consequently, songs evolved towards the wild-type in three to four generations. Thus, species-typical song culture can appear *de novo*. Our study has parallels with language change and evolution[12–14]. In analogy to models in quantitative genetics[15,16], we model song culture as a multigenerational phenotype partly encoded genetically in an isolate founding population, influenced by environmental variables and taking multiple generations to emerge.

Young male zebra finches develop individually distinct song by imitating adult males[17]. The adult wild-type (WT) song includes stereotyped syllables repeated in fixed order (song motifs; Fig. 1a) in both wild and domesticated zebra finch colonies. Birds deprived of song during vocal development develop a less structured isolate (ISO) song with more noisy, broadband notes and high-pitch upsweeps[11] (Fig. 1b). ISO syllables are often prolonged, monotonic or stuttered, and the songs appear to have an irregular rhythm. Despite these anomalies, young zebra finches readily imitate songs of adult isolates[18] even in the presence of WT adults[11].

We quantified the differences between WT and ISO songs on three timescales. At the 10-ms timescale, we used spectral-frame features (for example frequency modulation; Supplementary Information, section 4a). At the 10–100-ms timescale, we used the correlation time of the spectral shape, termed the duration of acoustic state (DAS; Supplementary Information, section 4b). At longer (200–1,000-ms) timescales, we used measures of song rhythm (Supplementary Information, section 4d)[19]. Feature probability distributions across birds differed between ISO and WT songs (Fig. 1c–e). ISO songs had lower frequency modulation, longer DAS and less structured rhythms.

These distributions provide a high-dimensional song phenotype for each bird. We reduced the dimensionality by applying principal component analysis (PCA) to the collection of feature distributions of all birds (WT plus ISO), and retained the first two principal components (PC1 and PC2) to obtain two-dimensional song phenotype values (Supplementary Information, section 4e). Principal components at all three timescales show separable clusters for ISO and WT songs along a continuum (Fig. 2a–c). The mean values of PC1 were significantly different between ISO and WT at all timescales of song structure ($P < 0.001$, t tests, $n_{WT} = 52$ birds, $n_{ISO} = 17$ birds, adjusted for false discovery rate; Supplementary Information, section 5). We found that these differences are largely an outcome of tutoring deprivation and not of social isolation (Supplementary Information, section 3f).

To examine the imitation of isolate songs, we trained 13 juvenile birds (pupils) one to one with isolate tutors in a sound-isolated chamber. This allowed us to control genetic relatedness, and to minimize

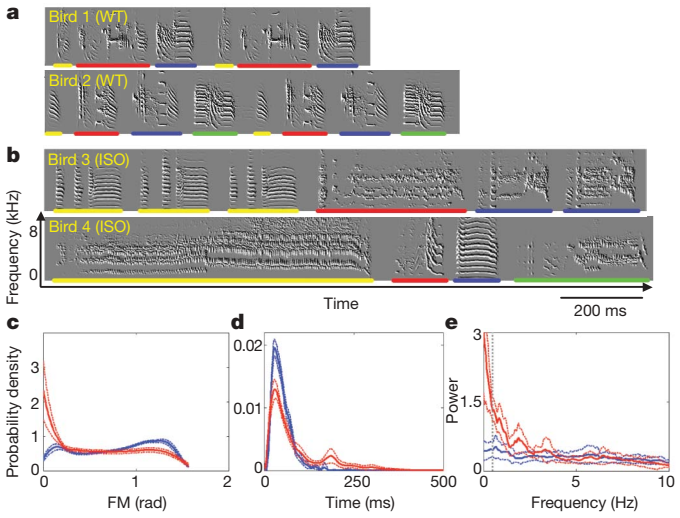


Figure 1 | Wild-type songs versus isolate songs. **a**, Spectral derivatives[30] of two WT song bouts. Different syllable types are underlined in different colours. Syllables show stereotypical organization into song motifs and rapid acoustic transitions within syllables. **b**, ISO song bouts. Some syllables are extremely long (bird 4, yellow) and others are stuttered (bird 3, yellow and blue). **c**, Mean distribution histogram of frequency modulation (FM) in WT birds (blue, $n = 52$) and ISO birds (red, $n = 17$). Dotted lines delineate the 95% confidence intervals. **d**, Histogram of DAS, showing longer durations in ISO. **e**, Spectra of rhythm frequencies, showing less structured rhythm in ISO. The dotted grey line marks the minimum frequency that we used for further analysis (0.5 Hz).

[1]Department of Biology, City College, City University of New York, New York 10031, USA. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

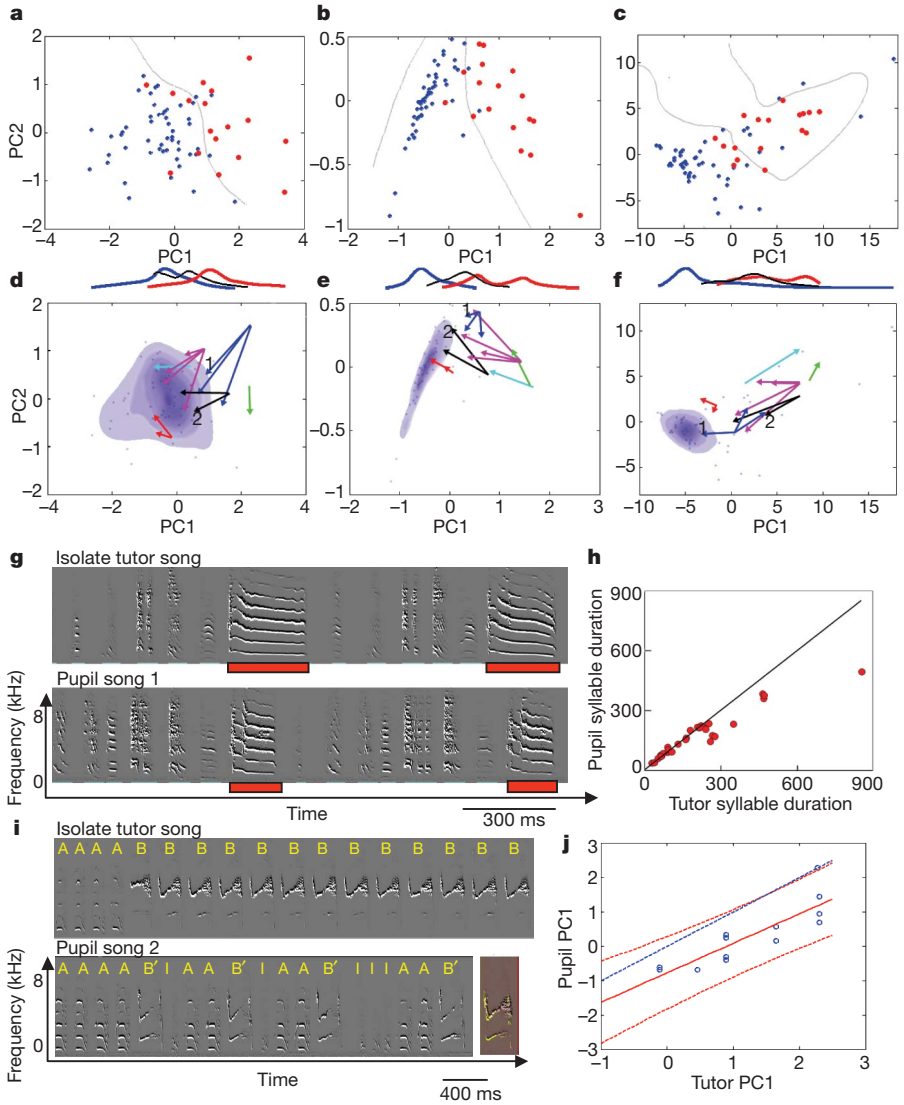


Figure 2 | Progression towards WT song in pupils of isolates. First two principal components constructed from spectral features (**a**), DAS (**b**) and rhythm frequencies (**c**). Dots represent individual WT (blue, $n = 52$) and ISO (red, $n = 17$) birds. Bayes classification lines are shown in grey. The histogram (bottom) of PC1 in first-generation pupils (black, $n = 13$) falls between those of WT and ISO. **d–f**, Same data as in **a–c**. Arrows originate at the tutors and point towards pupils. Different colours represent different tutors. Purple shading indicates the centre of the WT cluster. Numerals indicate the arrows corresponding to the songs in **g** and **i**. **g**, **h**, Biased copying of syllable durations (milliseconds). **i**, Biased copying of syllable abundance and emergence of song motif. Shaded rectangle, overlay of syllable B and its imitation, B′. **j**, Correlation between PC1s of pupil and tutor, indicating biased imitation. Dashed red lines delineate the 95% confidence band, and the dashed blue line is the identity line.

social effects, for example to eliminate feedback from female listeners. Four isolate tutors, with songs stable over the course of tutoring, were used two to four times to train unrelated pupils. We projected the feature distributions of the pupils onto the principal components derived earlier from the WT–ISO data (Fig. 2a–c), and displayed vectors connecting each ISO tutor to his pupils (Fig. 2d–f). As shown, most of these vectors point in the direction of the WT cluster, indicating a shift towards WT features in pupils of ISO tutors. The mean values of PC1 for the first-generation pupils differed significantly from both ISO and WT means for the spectral-frame features and for DAS ($P = 0.018$–0.001, $n = 13$), but not for rhythm. Feature distributions of most individual pupil songs were closer to WT songs than were their tutor's songs (12 of 13 at at least one timescale, 10 of 13 at all timescales, false discovery rate significance of 0.01, binomial test, $n = 52$; Supplementary Information, section 5d).

Although pupils typically imitated all of the tutor syllables[20] and did not invent new syllables (Supplementary Information, section 2), pupil songs deviated consistently from tutor songs. In Fig. 2g, we show an example in which a long ISO syllable (red bar; mean duration of 367 ms, $\sigma = 29$ ms) was copied by a pupil, but was shortened by about 30% (mean duration of 243 ms, $\sigma = 7.6$ ms). Across all the syllables and all pupils, the durations of pupil syllables accurately matched those of the corresponding ISO tutor syllables for syllables shorter than 230 ms (correlation coefficient $r^2 = 0.98$, slope of 0.97, $n = 20$ syllables; Fig. 2h). Copies of longer ISO syllables, however, were shorter than the originals ($r^2 = 0.84$, slope of 0.56, $n = 11$ syllables). Across birds, the ratio between the longest and shortest syllables within a bout was significantly smaller in pupils than in their ISO tutors ($P < 0.01$, $n = 13$, Wilcoxon sign test; Supplementary Information, section 4c). Overall, the range over which durations of ISO syllables were accurately copied is similar to the range of WT syllable durations (25–75 percentile range of 67–180 ms, $n = 52$ WT birds). In addition, pupils only copied the abundance (relative frequency) of syllables when it was within the WT range (up to about 30%). In cases in which one syllable dominated the ISO song (Fig. 2i), in pupils its abundance decreased to 20–30% (Supplementary Fig. 5), thereby creating more structured song motifs.

Imitation of spectral features, as judged by PC1 of the feature distribution, was also biased: linear regression analysis of pupil versus tutor yielded a non-zero intercept and a slope slightly less than one

(Fig. 2j). The equality line, corresponding to faithful copying (identical PC1s; dashed blue line), was rejected in favour of the alternative hypothesis represented by the linear fit shown in red ($P < 0.001$, likelihood ratio test, $n = 13$). We note that imitation that was inaccurate but unbiased would have only increased the spread around the equality line.

Because the songs of ISO-tutored birds differed significantly from both their respective ISO tutors and WT, we examined whether recursive tutoring would cause further progression towards WT over multiple generations. We used four of the first-generation pupils as tutors of a second generation of unrelated pupils, and continued recursively over two to five generations (Fig. 3a). Similarity to WT songs increased over three to four generations, as can be appreciated from Supplementary Audio 1 and the three examples of multiple generations of recursive tutoring in Fig. 3b. In the first example, both ISO syllables become shorter in the songs of the first- and second-generation pupils (blue and red rectangles), but the second syllable is also differentiated into three distinct notes. The middle panel shows spectral and temporal differentiation of syllables, and omission by the third-generation pupil. In the rightmost lineage, the duration of the final syllable (red rectangle) decreased over two generations and then stabilized. The spectral structure, however, continued to change in the third and fourth generations.

To judge whether or not the imitation of ISO song progressed towards WT song over multiple generations, we displayed vectors in the principal-component space (as in Fig. 2d–f) with each tutoring lineage labelled using a different colour (Fig. 3c–e). As shown, the multigenerational trajectories penetrate more deeply into the WT cluster (purple shading). Direct comparisons across first- and later-generation pupils reach significance only for DAS ($P = 0.02$), but multigenerational comparisons suggest further progression towards WT for all song traits. For spectral-frame features, we found that PC1 of song features changes monotonically towards WT over generations. Its mean values for ISO, first generation, later generations and WT songs were 1.3, 0.3, 0.03 and −0.4, respectively. PC1 values for later-generation songs were significantly different from ISO song ($P < 0.005$, t test, $n = 8$ for later generations) but not from WT songs ($P = 0.17$). For DAS, PC1 values also decreased monotonically with increasing generation: 1.1, 0.3, 0.02, −0.3. Later-generation songs were significantly different ($P < 0.01$) from both WT and ISO song, suggesting that WT approximation was not complete. For rhythm, PC1 values also decreased monotonically with increasing generation (4.1, 2.2, 1.4, −2), and differences from WT and ISO were marginally significant ($P = 0.02$ and 0.056, respectively).

Although the one-to-one training provided a well-defined learning environment, the multigenerational changes that would occur in a complex social setting may be more representative of natural evolutionary processes. Therefore, we established a semi-natural island colony (Supplementary Information, section 3d) starting with one of our isolate tutors and three unrelated females in a large sound chamber (Supplementary Fig. 1).

In this social situation, too, the isolate colony approached the WT cluster over a few generations (Fig. 4). To judge the transition towards WT clusters, we examined principal-component projections, with the isolate tutor song marked as a red dot. Comparing the trajectory shown in Fig. 4e with that in the rightmost panel of Fig. 3b (originating from the same tutor), we see that the outcome in the colony is similar to that observed in one-to-one tutoring. Although the outcome of the colony experiment can only be judged qualitatively, we find it noteworthy that despite intense social interactions, female presence and mating competition, there were only mild differences between birds in the two conditions. In the colony, juveniles also imitated sibling syllables and female long calls, leading to more complex songs (Supplementary Information, section 1c). In

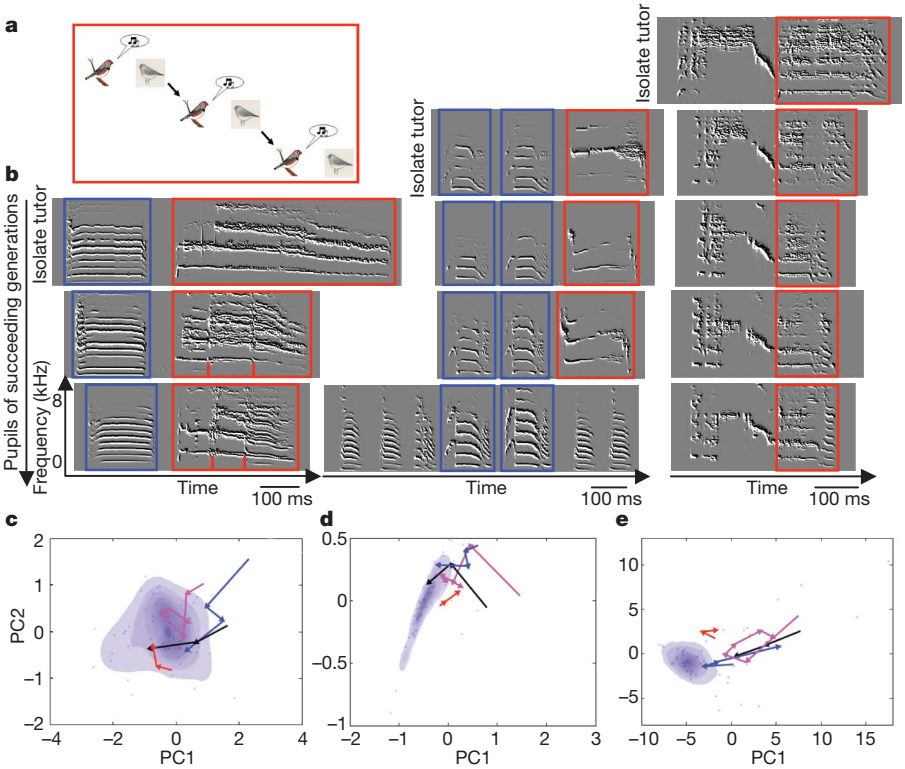


Figure 3 | Multigenerational progression towards WT song. **a**, Schematic diagram of the experimental set-up. Pupils become tutors when they reach adulthood (day 120–140). **b**, Three examples of the songs of isolate tutors and the succeeding generations of learners. Blue and red boxes show individual syllable types that are altered by pupils. Long, monotonic syllables become shorter and more differentiated (left- and right-hand panels). Rarely, syllables were omitted (middle panel) in later generations of learners. **c**–**e**, PCA of spectral features (**c**), state duration (**d**) and rhythm spectra (**e**). As in Fig. 2d–f, arrows originate at the tutors and point towards pupils. The progression towards the WT cluster (purple ovals) continues over generations.

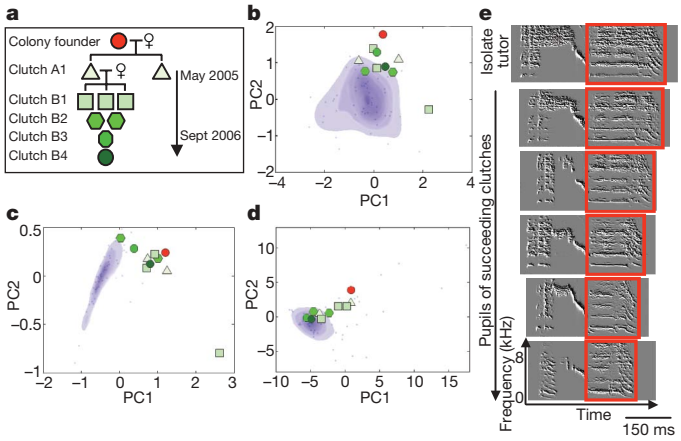


Figure 4 | Progression towards WT song in an isolated colony. **a**, Family relationships in the first five clutches, based on behavioural observations. **b–d**, PCA of spectral features (**b**), DAS (**c**) and rhythm (**d**) (as in Fig. 2d–f). The colony founder is marked with a red dot. Colours and symbols identify individuals in **a**. Successive clutches approach the WT cluster (purple shading) in spectral features, DAS, and most strongly in rhythm frequencies. **e**, A long syllable that dominates the founder isolate song motif, and its imitations in successive clutches.

contrast to one-to-one tutoring, the farthest progress towards WT song occurred in rhythm, perhaps because birds incorporated additional syllable types into their song motifs.

Our findings resemble the well-known case of deaf children in Managua, Nicaragua, spontaneously developing sign language[21], as well as linguistic phenomena such as creolization. Models of language change and evolution[12,13,15], which contain a developmental account of the language acquisition process, are germane to our study (Supplementary Material, model details, section 3).

We further discuss our findings using a simple recursive model that motivated this study. Principal components of feature distributions (Fig. 2) give us phenotypic measures of song. Consider the distribution of a quantitative phenotype P in the ISO population. Because some of the variation in ISO song is heritable, we partition P into genotypic and environmental values, as $P = G + E$, assuming an additive model for genetic variance[22], $V_P = V_G + V_E$.

We consider an isolated-lineages model, in which the environmental component of the pupil phenotype in the $(n+1)$th generation, $E(n+1)$, is further divided into a portion, $E_0(n+1)$, independent of the tutor, and a portion, $c_0P(n)$, proportional to the tutor song phenotype. We therefore have the recursion

$$P(n+1) = G(n+1) + c_0P(n) + E_0(n+1) \qquad (1)$$

The partitioning of the phenotypic variance is analogous to the parental-effects model in quantitative genetics[1,23]. In the one-to-one study, tutor and pupil genotypic values are approximately uncorrelated, and c_0 may be estimated by regressing the pupil against the tutor (see Fig. 2j, where $c_0 = 0.86$ and $\sigma = 0.15$). The literature on cultural transmission[24,25] also contains models analogous to equation (1) and has similar implications. Half-sib or cross-fostering experimental designs[26] should be useful for separating the genetic[27] and learning-related components of song transmission in future studies[28].

Our one-to-one experimental design may be modelled using equation (1) by initializing $P(1) = G(1) + E(1)$ for the ISO generation. The recursion then causes the distribution of phenotypic values to exponentially relax to an asymptotic 'WT' distribution, the relaxation being rapid if c_0 is near 0. The largest changes occur in the first generation (consistent with our results). The case $c_0 = 1$ corresponds to a simple random walk ($V_{P(n)} \approx \sqrt{n}$), in which the song phenotype would drift indefinitely (unbiased song copying with errors). The 'copying bias' $(1 - c_0)$ plays the role of a spring constant, confining the walker to a parabolic potential well. Notably, the WT variance in the model is a combination of the ISO variance and the learning parameter, emphasizing how ISO song and learning ability combine to produce WT song. Extensions of the model predict that both genetic relatedness between tutor and pupil and horizontal transmission alter the asymptotic 'WT' distributions (Supplementary Material, model details, sections 1 and 2). Therefore, we would expect our two designs to yield slightly different song cultures.

In a sense, the results of our study show that song culture is the result of an extended developmental process, a 'multigenerational' phenotype partly genetically encoded in a founding population and partly in environmental variables, but taking multiple generations to emerge. The functional significance of our findings—that is, their bearing on whether WT females prefer the songs of multigeneration pupils to those of ISO tutors—remains open. Because our findings suggest that song culture is the result of an extended developmental process, it would be interesting to examine whether changes in gene expression, neuronal reorganization or neurogenesis associated with song development show orderly multigenerational progression during the evolution of song culture.

METHODS SUMMARY

Animal care. All experiments were performed in accordance with guidelines of the US National Institutes of Health and have been reviewed and approved by the Institutional Animal Care and Use Committee of City College, City University of New York.

Experimental design. We used zebra finches (*Taeniopygia guttata*) from the City College breeding colony. Colony management and isolation procedures have been described previously[29]. Except for the colony experiment, all birds were kept either singly (isolates) or in pairs (one-to-one tutored) in sound attenuation chambers (Supplementary Information, section 3e) betweens days 30 and 120 post hatch. WT songs ($n = 52$) were obtained from birds raised in two well-established colonies. Isolates ($n = 17$) were raised by their mothers from day 7 to day 29 post hatch and were kept in complete isolation from day 30 until day 120 or later. One-to-one-tutored birds ($n = 13$ and 8 for first and later generations, respectively) were randomly selected from 40 breeding pairs and paired with one of six isolate tutors on day 30. For the colony setting, we made a sound isolation chamber from an inoperational 20-ft^3 refrigerator (Supplementary Fig. 1). All birds in the colony (except for the three female founders) were the descendants of the founder male.

Data analysis. All the analysis was performed using MATLAB 7, except for spectral-feature calculations, which were done using SOUND ANALYSIS PRO 2. Isolate song syllables are often prolonged and monotonic. To quantify this notion, we estimated the time interval over which acoustic features remain highly correlated and named this feature the DAS (Supplementary Information, section 4b). Rhythm spectrum[19] was used to detect periodicity in song features at the syllabic and song-motif levels (Supplementary Information, section 4d). We constructed song-feature principal components by first computing cumulative frequency distributions for each feature time series (Supplementary Fig. 8). These cumulative frequency distributions were the input vectors for the PCA (Fig. 2a–c). Statistical tests are described Supplementary Information, section 5.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 October 2008; accepted 16 March 2009.
Published online 3 May 2009.

1. Marler, P. & Sherman, V. Innate differences in singing behavior of sparrows reared in isolation from adult conspecific song. *Anim. Behav.* **33**, 57–71 (1985).
2. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, 1965).
3. Catchpole, C. K. & Slater, P. B. J. *Bird Song: Biological Themes and Variations* (Cambridge Univ. Press, 2008).
4. Marler, P. & Tamura, M. Culturally transmitted patterns of vocal behavior in sparrows. *Science* **146**, 1483–1486 (1964).
5. Olofsson, H. & Servedio, M. R. Sympatry affects the evolution of genetic versus cultural determination of song. *Behav. Ecol.* **19**, 596–604 (2008).
6. Soha, J. A. & Marler, P. A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Anim. Behav.* **60**, 297–306 (2000).
7. West, M. J. & King, A. P. in *Issues in the Ecological Study of Learning* (eds Johnston, T. D. & Pietrewicz, A. T.) 245–274 (Erlbaum, 1985).
8. Gardner, T. J., Naef, F. & Nottebohm, F. Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science* **308**, 1046–1049 (2005).
9. Adret, P. In search of the song template. *Ann. NY Acad. Sci.* **1016**, 303–324 (2004).

10. Volman, S. F. & Khanna, H. Convergence of untutored song in group-reared zebra finches (*Taeniopygia guttata*). *J. Comp. Psychol.* **109,** 211–221 (1995).
11. Williams, H., Kilander, K. & Sotanski, M. L. Untutored song, reproductive success and song learning. *Anim. Behav.* **45,** 695–705 (1993).
12. Niyogi, P. *The Computational Nature of Language Learning and Evolution* (MIT Press, 2006).
13. Nowak, M. A., Komarova, N. L. & Niyogi, P. Computational and evolutionary aspects of language. *Nature* **417,** 611–617 (2002).
14. Kirby, S., Cornish, H. & Smith, K. Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl Acad. Sci. USA* **105,** 10681–10686 (2008).
15. Cheverud, J. M. Evolution in a genetically heritable social environment. *Proc. Natl Acad. Sci. USA* **100,** 4357–4359 (2003).
16. Dickerson, G. Comparison of hog carcasses as influenced by heritable differences in rate and economy of gain. *Iowa Agric. Exp. Sta. Res. Bull.* **354,** 489–524 (1947).
17. Jones, A. E., ten Cate, C. & Slater, P. J. B. Early experience and plasticity of song in adult male zebra finches (*Taeniopygia guttata*). *J. Comp. Psychol.* **110,** 354–369 (1996).
18. Böhner, J. Song learning in the zebra finch (*Taeniopygia guttata*): selectivity in the choice of a tutor and accuracy of song copies. *Anim. Behav.* **31,** 231–237 (1983).
19. Saar, S. & Mitra, P. P. A technique for characterizing the development of rhythms in bird song. *PLoS ONE* **3,** e1461 (2008).
20. Nelson, D. A. & Marler, P. Selection-based learning in bird song development. *Proc. Natl Acad. Sci. USA* **91,** 10498–10501 (1994).
21. Kegl, J. in *Language Acquisition in Signed Languages* (eds Morgan, G. & Woll, B.) 207–254 (Cambridge Univ. Press, 2002).
22. Falconer, D. S. & Mackay, T. C. F. *Introduction to Quantitative Genetics* 4th edn (Prentice Hall, 1996).
23. Willham, R. L. The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19,** 18–27 (1963).
24. Cavalli-Sforza, L. L. & Feldman, M. W. *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton Univ. Press, 1981).
25. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (University of Chicago Press, 1985).
26. Boake, C. R. B. *Quantitative Genetic Studies of Behavioral Evolution* 94–98 (University of Chicago Press, 1994).
27. Mundinger, P. C. Behaviour-genetic analysis of canary song: inter-strain differences in sensory learning, and epigenetic rules. *Anim. Behav.* **50,** 1491–1511 (1995).
28. Soha, J. A. & Marler, P. Cues for early discrimination of conspecific song in the white-crowned sparrow (*Zonotrichia leucophrys*). *Ethology* **107,** 813–826 (2001).
29. Tchernichovski, O., Lints, T., Mitra, P. P. & Nottebohm, F. Vocal imitation in zebra finches is inversely related to model abundance. *Proc. Natl Acad. Sci. USA* **96,** 12901–12904 (1999).
30. Tchernichovski, O., Nottebohm, F., Ho, C. E., Bijan, P. & Mitra, P. P. A procedure for an automated measurement of song similarity. *Anim. Behav.* **59,** 1167–1176 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Wallman and H. Williams for critical reading of the manuscript and consultation and F. Nottebohm, C. Harding and N. Leader for recordings of WT songs. The study was supported by US National Institutes of Health (NIH) grants to O.T. and P.P.M., by a NIH Research Centers in Minority Institutions grant to City College, City University of New York, and by the Crick-Clay Professorship to P.P.M.

Author Contributions The idea for the study originated with P.P.M., with important modifications by O.T. and O.F. The experiments were carried out by O.F. and O.T. The model was developed by P.P.M. with help from H.W. All authors participated in the data analysis, with major efforts by H.W. and O.F.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.F. (olcifeher@gmail.com).

METHODS

For a more thorough description of our methods and supplementary data, see Supplementary Information.

Animal care. All experiments were performed in accordance with guidelines of the US National Institutes of Health and have been reviewed and approved by the Institutional Animal Care and Use Committee of City College, City University of New York.

Experimental design. We used zebra finches (*Taeniopygia guttata*) from the City College breeding colony.

Isolation from song culture: All birds (except for the isolated colony birds) were raised by their parents in a dedicated cage until day 7 post hatch. The father was then removed and the cage was taken to a nursery area housing mothers (who do not sing) and chicks only. Zebra finches do not imitate songs heard before day 20 (ref. 31), and measurements obtained on day 10 post hatch show elevated auditory thresholds[32]. Therefore, possible exposure to the father WT songs before day 7 and later exposure to siblings is unlikely to have an effect (Supplementary Fig. 2). Except for the colony experiment, all birds were kept either singly (isolates) or in pairs (one-to-one tutored) in sound-attenuating chambers from day 30 to day 120 post hatch. Our custom-made sound chambers provide a sufficient level of box-to-box isolation for zebra finch songs (Supplementary Information, section 3e).

Experimental groups. WT songs ($n = 52$): To obtain a baseline for WT songs produced by domesticated zebra finches, we used recordings from birds raised in two well-established colonies, The Rockefeller University colony and Hunter College colony. Both colonies have existed for over 20 years.

Isolates ($n = 17$): Birds kept visually and acoustically separated from other birds during the sensitive period for song learning are called isolates. Birds were isolated from songs from day 7 to day 29 post hatch. They were kept in complete isolation from day 30 until day 120 post hatch or later. Six of these birds were used as tutors. Tutors' ages ranged between 140 and 1,571 days (median age at beginning of first tutoring, 316 days).

One-to-one tutoring ($n = 13$ and 8 for first and later generations, respectively): We randomly selected 13 hatchlings from 40 breeding pairs and paired them with one of six isolate tutors on day 30. The isolate tutor and his pupil were kept together for 90 days in a sound-isolated chamber. The isolate tutors were removed when the pupils were 120 days old. At this time, we recorded the song of the pupil and obtained an additional recording from the isolate tutor (in a separate box) to test whether his song remained unchanged. After confirming that the isolate tutor's song remained stable over the tutoring period, a new pupil was placed with the tutor and the new training began.

Tutoring lineages: For four of the six isolate tutors, we established a line of learners, where the first-generation pupils tutored another generation that, upon reaching adulthood (between days 120 and 140), tutored another generation, and so on. This allowed us to track the same song as it was passed down over a few generations.

Isolate colony setting: We constructed a large isolation chamber from an inoperational 20-ft^3 refrigerator (Supplementary Fig. 1). We used one of our isolate tutors to establish the 'island colony'. Three females and this isolate male were placed into the chamber and kept completely isolated, acoustically as well as socially, from other birds over a period of two years. Once a pair bond had been established, we followed (by audio and video recording) the evolution of this colony. All birds in the colony (except for the three female founders) were the descendants of the founder male. However, he fathered only one clutch, after which one of his sons from this first clutch paired up with another of the original female founders and produced all the successive clutches. On the basis of partial video observations, we suspect that the rest of the colony birds were all descendants of this pair. We allowed the colony to grow, with the occasional removal of female offspring, over five generations of learners. The colony founder was removed just before the hatching of the fifth-generation learner.

Data analysis. All the analysis presented below was performed using MATLAB 7, except for feature calculations, which were done using SOUND ANALYSIS PRO 2.

Spectral-frame (10-ms) features: We used SOUND ANALYSIS PRO 2 to calculate song features in 10-ms windows. We analysed 20 s of singing bouts for each bird. To construct the PCA shown in Fig. 2a–c, we used the three features that showed the best separation between ISO and WT in the cumulative frequency distributions (Supplementary Fig. 5): frequency modulation, amplitude modulation and goodness of pitch.

DAS: Isolate song syllables and notes are often prolonged and monotonic. To quantify this notion, we estimated correlation time, namely, the interval over which acoustic features remain highly correlated. This procedure was implemented in SOUND ANALYSIS PRO 2. See Supplementary Information, section 4b, for details.

Rhythm spectrum: Rhythm spectrum[19] was used to detect periodicity (rhythm) in song features on the timescale of the song bout. Rhythm frequencies can capture patterns of repetitions at the syllabic level and at the song-motif level. See Supplementary Information, section 4d, for details.

Constructing the song features PCA: The song of each individual bird is described by a set of feature vectors spanning multiple timescales. For example, when analysing 20 s of singing, we obtained several spectral-frame features (pitch, frequency modulation and so on), with 20,000 time-series values for each feature. We first computed cumulative frequency distributions for each feature time series (Supplementary Fig. 8). The cumulative frequency distributions, which summarize the distribution of each feature in a song, are the input vectors of the PCA shown in Fig. 2a–c. Note that each red dot is a two-dimensional projection of cumulative frequency distributions (red lines) presented in Supplementary Fig. 8 for each bird (combining frequency modulation, amplitude modulation and goodness of pitch).

Statistical tests: See Supplementary Information, section 5.

31. Roper, A. & Zann, R. The onset of song learning and song tutor selection in fledgling zebra finches. *Ethology* **112**, 458–470 (2006).
32. Amin, N., Doupe, A. & Theunissen, F. E. Development of selectivity for natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* **97**, 3517–3531 (2007).

LETTERS

Autism genome-wide copy number variation reveals ubiquitin and neuronal genes

Joseph T. Glessner[1], Kai Wang[1], Guiqing Cai[2], Olena Korvatska[3], Cecilia E. Kim[1], Shawn Wood[4], Haitao Zhang[1], Annette Estes[3], Camille W. Brune[5], Jonathan P. Bradfield[1], Marcin Imielinski[1], Edward C. Frackelton[1], Jennifer Reichert[2], Emily L. Crawford[6], Jeffrey Munson[3], Patrick M. A. Sleiman[1], Rosetta Chiavacci[1], Kiran Annaiah[1], Kelly Thomas[1], Cuiping Hou[1], Wendy Glaberson[1], James Flory[1], Frederick Otieno[1], Maria Garris[1], Latha Soorya[2], Lambertus Klei[4], Joseph Piven[7], Kacie J. Meyer[8], Evdokia Anagnostou[2], Takeshi Sakurai[2], Rachel M. Game[6], Danielle S. Rudd[8], Danielle Zurawiecki[2], Christopher J. McDougle[10], Lea K. Davis[8], Judith Miller[9], David J. Posey[10], Shana Michaels[4], Alexander Kolevzon[2], Jeremy M. Silverman[2], Raphael Bernier[3], Susan E. Levy[11], Robert T. Schultz[11], Geraldine Dawson[3], Thomas Owley[5], William M. McMahon[9], Thomas H. Wassink[8], John A. Sweeney[5], John I. Nurnberger Jr[10], Hilary Coon[9], James S. Sutcliffe[6], Nancy J. Minshew[12], Struan F. A. Grant[1,11], Maja Bucan[13], Edwin H. Cook Jr[5], Joseph D. Buxbaum[2,14], Bernie Devlin[4], Gerard D. Schellenberg[15] & Hakon Hakonarson[1,11]

Autism spectrum disorders (ASDs) are childhood neurodevelopmental disorders with complex genetic origins[1–4]. Previous studies focusing on candidate genes or genomic regions have identified several copy number variations (CNVs) that are associated with an increased risk of ASDs[5–9]. Here we present the results from a whole-genome CNV study on a cohort of 859 ASD cases and 1,409 healthy children of European ancestry who were genotyped with ~550,000 single nucleotide polymorphism markers, in an attempt to comprehensively identify CNVs conferring susceptibility to ASDs. Positive findings were evaluated in an independent cohort of 1,336 ASD cases and 1,110 controls of European ancestry. Besides previously reported ASD candidate genes, such as *NRXN1* (ref. 10) and *CNTN4* (refs 11, 12), several new susceptibility genes encoding neuronal cell-adhesion molecules, including *NLGN1* and *ASTN2*, were enriched with CNVs in ASD cases compared to controls ($P = 9.5 \times 10^{-3}$). Furthermore, CNVs within or surrounding genes involved in the ubiquitin pathways, including *UBE3A*, *PARK2*, *RFWD2* and *FBXO40*, were affected by CNVs not observed in controls ($P = 3.3 \times 10^{-3}$). We also identified duplications 55 kilobases upstream of complementary DNA *AK123120* ($P = 3.6 \times 10^{-6}$). Although these variants may be individually rare, they target genes involved in neuronal cell-adhesion or ubiquitin degradation, indicating that these two important gene networks expressed within the central nervous system may contribute to the genetic susceptibility of ASD.

ASDs, including autism, are neurodevelopmental disorders characterized by impairments in social and communication skills, as well as stereotyped and repetitive behaviours and/or a restricted range of interests. Current prevalence estimates in the United States are 0.1–0.2% for autism and 0.6% for ASDs[1,2].

Linkage and candidate gene association studies have implicated several chromosomal regions in autism[3,4]. However, positive findings in one study often fail to replicate in other studies, and a consistent picture of susceptibility loci in autism is still lacking. Some telling clues about ASD genetics arose from recent studies on CNVs[5], including the association of *de novo* CNVs with ASDs[6]. Although *de novo* CNVs that disrupt specific genes may contribute to the pathogenesis of ASDs, heritable CNVs are much more common but have been less studied as risk factors of ASDs. A family-based genome-wide linkage and CNV analysis by the Autism Genome Project Consortium using Affymetrix 10K single nucleotide polymorphism (SNP) arrays implicated chromosome 11p12-13 and neurexin 1 (*NRXN1*) as candidate loci[7]. A study using the Affymetrix 500K SNP array in a Canadian population reported 277 rare CNVs that were only observed in ASD patients but not in 1,652 healthy controls or in the Database of Genomic Variants[8]. Furthermore, 16p11.2 deletions and duplications have been reported in independent cohorts of autism patients[9]. These studies concordantly implicate a role for CNVs in the genetic susceptibility to ASDs.

To search systematically for CNVs that confer a risk to ASDs, we used a genome-wide approach with the Illumina HumanHap550 BeadChip. We assembled an Autism Case-Control (ACC) cohort by collecting 859 ASD cases (from a total of 1,246 ACC cases, parents, and siblings) of European ancestry, and 1,409 healthy controls. Among these case subjects, all met the diagnostic criteria for autism on the basis of Autism Diagnostic Interview (ADI), and 124 met the criteria for other ASDs on the basis of Autism Diagnostic Observation Schedule (ADOS)[13]. Fifty-four per cent were from simplex families, the rest were from multiplex families. In addition, we also analysed 1,336 ASD cases (from a total of 3,398 cases, parents, and siblings) in the Autism Genetic Resource Exchange (AGRE)[14] collection, as well as 1,110 control subjects as a replication cohort. Among the AGRE cases, 5% were from simplex families and 95%

[1]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [2]Seaver Autism Center for Research and Treatment, Department of Psychiatry, Mount Sinai School of Medicine, New York, New York 10029, USA. [3]University of Washington, Seattle, Washington 98105, USA. [4]Departments of Psychiatry and Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [5]Institute for Juvenile Research and Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60608, USA. [6]Center for Molecular Neuroscience and Vanderbilt Kennedy Center, Vanderbilt University, Nashville, Tennessee 37235, USA. [7]Neurodevelopmental Disorders Research Center and Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27412, USA. [8]University of Iowa, Iowa City, Iowa 52242, USA. [9]University of Utah, Salt Lake City, Utah 84112, USA. [10]Indiana University, Indianapolis, Indiana 46202, USA. [11]Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [12]Departments of Psychiatry and Neurology, University of Pittsburg School of Medicine, Pittsburg, Pennsylvania 15260, USA. [13]Dept of Genetics, Biology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [14]Departments of Neuroscience, and Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA. [15]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

Table 1 | CNVs in gene regions previously implicated in ASDs

Gene/region	Region of significance	Type	Validation	ACC case	ACC control	AGRE case	AGRE control	Inh	Combined *P* value	Permuted *P* value
UBE3A 15q11-13	chr15: 21200234–26208861	Dup	MLPA, qPCR	8	0	7 (5)	0	6:0:2 (75%)	1×10^{-5}	<0.001
NRXN1	chr2: 51120644–51147600	Del	qPCR	4	0	6 (4)	0	4:3:0 (100%)	4.7×10^{-4}	0.002
CNTN4	chr3: 1915190–1915922	Del	qPCR	3	0	7 (4)	0	5:3:0 (100%)	4.7×10^{-4}	0.004
22q11.21	chr22: 19351264–19358946	Dup	MLPA	5	0	4	0	2:4:0 (100%)	0.0010	0.008
CNTN4	chr3: 2548148–2548531	Dup	qPCR	1	1	8 (6)	0	2:3:1 (83%)	0.0078	0.013
16p11.2	chr16: 29554844–30085308	Dup	Affymetrix	2	3	7 (3)	1	3:4:1 (88%)	0.162	0.408
16p11.2	chr16: 29554844–30085308	Del	Affymetrix	3	2	5 (4)	2	0:0:5 (0%)	0.246	0.300
AUTS2	chr7: 68576554–68967283	Dup	Affymetrix	0	0	1	0	1:0:0 (100%)	0.466	0.425
NLGN3	chrX: 70288980–70460080	Dup	Affymetrix	0	0	1	1	0:0:0 (NA)	0.601	0.601
SHANK3	chr22: 49456858–49524956	Del	Affymetrix	2	1	0	1	4:3:0 (100%)	1	1

The number in parenthesis refers to count of unrelated siblings or distinct unrelated families. The sample included 859 ASD cases from the ACC cohort, 1,336 ASD cases from the AGRE cohort, and 2,519 unaffected controls (1,409 ACC discovery controls and 1,110 AGRE replication controls). All CNVs, except 16p11.2, *AUTS2*, *NLGN3* and *SHANK3*, were experimentally validated in the ACC cohort. Regions listed represent the optimal overlap of cases and significance with respect to controls, as described in the Methods and Supplementary Fig. 5, upper panel. 'Inh' column lists the inheritance pattern of each CNV from parents to cases in the format <inherited from mother>:<inherited from father>:<denovo>. Pedigrees provided in Supplementary Fig. 9. The percentage of inheritance is listed in parenthesis next to these three values. Note that parents were not available for all cases.

were from multiplex families: 1,202 met the criteria for autism and 134 for other ASDs[13] (Supplementary Tables 1 and 2).

We generated 78,490 CNV calls (22,581 in the ACC series and 55,909 in the AGRE series) from all the ASD subjects and their family members that met strictly established data quality thresholds (Methods). On average, 15.5 CNV calls were made for each individual using the PennCNV software[15], with similar frequency observed in cases and controls (Supplementary Fig. 2).

We first examined eight genomic regions that have been previously implicated in ASDs. Among those, CNVs involving the 15q11-13, 22q11.21, and *NRXN1* regions have well-established associations with autism[10]. CNVs affecting *CNTN4* in ASD cases have also been reported in independent studies[11,12]. We statistically adjusted for relatedness of cases with permutation and our results demonstrate that duplications of 15q11-q13 and the 22q11.21 region, deletions of *NRXN1*, as well as deletions and duplications of *CNTN4* replicate in our cohorts (Table 1). Conversely, we did not obtain statistical support for several other genomic regions previously shown to associate with ASD, including *AUTS2* (ref. 16), *NLGN3* (ref. 17), *SHANK3* (ref. 18) and 16p11.2 (ref. 9) (Table 1). We observed a similar frequency of deletions and duplications of the 16p11.2 locus in the ASD cases (~0.3%) as previously reported[9]; however, the CNV frequency in the control subjects at this locus was also comparable to that of the cases (Supplementary Fig. 3). It is noteworthy that CNVs at the 16p11.2 locus do not segregate to all cases in three of the affected families and they are also transmitted to unaffected siblings (Supplementary Fig. 4). These results indicate that CNVs at the 16p11.2 locus may not be sufficient to be causal variants in ASD.

To identify other new genomic loci contributing to ASDs, we applied a segment-based scoring approach that scans the genome for consecutive SNPs with more frequent copy number changes in cases compared to controls. This approach defines copy number variation regions, or CNVRs (Supplementary Fig. 5, upper panel). In the ACC cohort, we identified four CNVRs that were observed in cases but not in controls, as well as five CNVRs that had significantly higher frequency in cases versus controls (Table 2).

To replicate the CNVRs exclusively observed in ACC cases, we examined the AGRE case-control data set; of the four case-specific CNVRs, two were also exclusive to AGRE cases (*PARK2* and *RFWD2*), whereas the other two (*AK057321* and *FBXO40*) were not observed in either the cases or controls (combined *P* values ranging from 3.57×10^{-6} to 0.1, unadjusted for multiple testing) (Table 2). Interestingly, four genes (*UBE3A*, *PARK2*, *RFWD2* and *FBXO40*) that were significantly enriched for CNVs and observed in the ASD cases only, belong to the ubiquitin gene family (UniProt category 'Ubl conjugation pathway', $P = 3.3 \times 10^{-3}$). The other five CNVRs, as well as being enriched in the ACC cases compared with controls, were over-represented in the AGRE cases compared with the independent controls (Table 2). Figure 1 shows the most statistically significant locus, a duplication 55 kb upstream of *AK123120*, using UCSC Genome Browser[19] with Build 36 (March 2006) of the human genome. To ensure reliability of our CNV detection method, we experimentally validated all the significant CNVRs using other methods, including quantitative PCR (qPCR) and multiplex ligation-dependent probe amplification (MLPA) (Fig. 2). Affymetrix 5.0 array data were also available for a subset of the AGRE subjects for validation.

Table 2 | New common CNVRs over-represented in ASD patients

CNVR	Gene	Description	Type	ACC case	ACC control	AGRE case	AGRE control	Inh	Combined statistics: OR	Combined statistics: *P* value	Combined statistics: Permuted *P* value
chr6: 162584576–162587001	*PARK2*	E3 ubiquitin-protein ligase	Del	3	0	4 (3)	0	0:5:0 (100%)	NA	***0.0047***	0.005
chr1: 174500555–174543675	*RFWD2, PAPPA2*	E3 ubiquitin-protein ligase	Dup	3	0	3	0	1:1:1 (67%)	NA	**0.0102**	0.011
chr7: 32667087–32770713	*AK057321*	Testis expression	Dup	4	0	0	0	0:0:1 (0%)	NA	0.0469	0.034
chr3: 122826190–122870474	*FBXO40, GOLGB1*	Protein-ubiquitin ligase	Dup	3	0	0	0	0:0:0 (NA)	NA	0.1009	0.094
chr2: 13119667–13165898	*AK123120*	Cerebellum expression	Dup	9	3	24 (20)	4	16:8:1 (96%)	5.547	***3.57 × 10⁻⁶***	<0.001
chr3: 4199731–4236304	*UNQ3037*	Sulphatase	Del	3	1	12 (8*)	2	6:7:0 (100%)	5.804	***0.0017***	0.003
chr10: 87941666–87949029	*GRID1*	Glutamate receptor	Del	4	1	10 (8)	2	4:7:0 (100%)	5.412	***0.0031***	0.011
chr3: 174754378–174771975	*NLGN1*	Neuroligin	Dup	40	52	74 (57†)	40	43:29:2 (97%)	1.471	**0.0101**	0.039
chr4: 144847402–144854579	*GYPELOC441046*	Glycophorin E precursor	Dup	3	0	7 (5)	2	2:2:2 (67%)	5.782	**0.0166**	0.037

The number in parenthesis refers to count of unrelated siblings or distinct unrelated families. The sample included 859 ASD cases from the ACC cohort, 1,336 ASD cases from the AGRE cohort and 2,519 unaffected controls (1,409 ACC discovery controls and 1,110 AGRE replication controls). All loci were validated with qPCR. 'Inh' column lists the inheritance pattern of each CNV from parents to cases in the format <inherited from mother>:<inherited from father>:<denovo>. The percentage of inheritance is listed in parenthesis. Italicized *P* values denote CNVs that survive multiple testing with Bonferroni adjustment in the discovery phase (that is, $P < 0.05$ after correction for five deletion and nine duplication CNVRs), and bold *P* values denote CNVRs that survived both the replication and experimental validation. OR, odds ratio.

* AGRE family 574 has three affected siblings.

† Three families had three affected siblings (AGRE families 656, 955 and 1559).

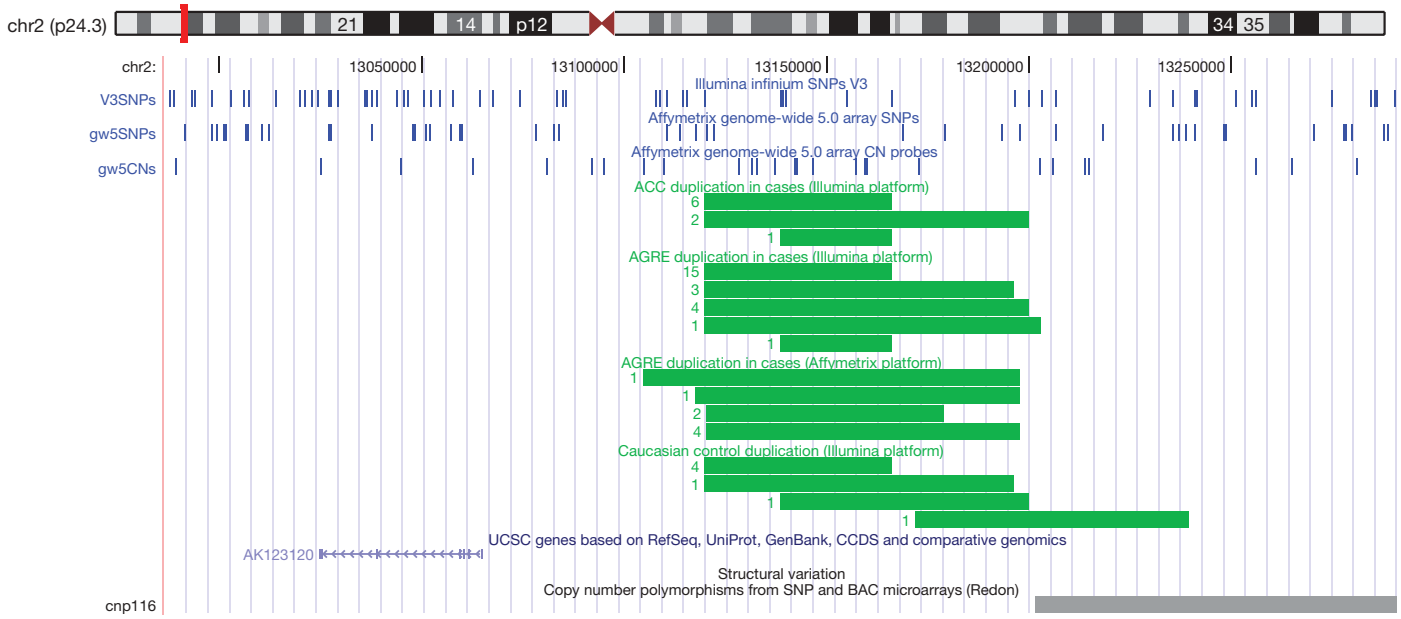


Figure 1 | *AK123120*: example of overrepresented CNVs. *AK123120* chromosome 2 (chr2): 12,986,750–13,291,000 divided into subsections with headers for ACC CNVs, AGRE CNVs, AGRE Affymetrix validation CNVs, and control CNVs. The AGRE Affymetrix Replication track is on the basis of genome-wide 5.0 SNP genotyping data from the Broad Institute (see Supplementary Methods and Acknowledgements), and were generated using the PennCNV-Affy algorithm (see Supplementary Methods), to serve as a further means to validate the Illumina-based CNV calls. SNP and copy number (CN) probe coverage are shown as blue lines across the top. Produced with custom tracks listing CNV calls uploaded to http://genome.ucsc.edu. Figures for all loci are included in Supplementary Information.

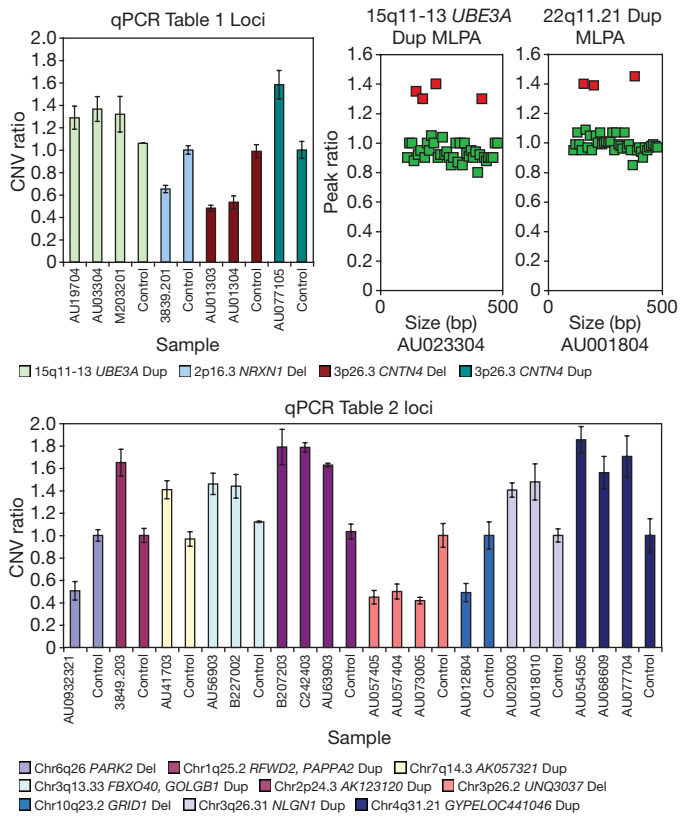


Figure 2 | Independent validation using qPCR and MLPA. Fluorescent probe-based qPCR assays using Roche Universal probe library and/or MLPA were designed to validate every candidate CNV with a completely independent test (representative series shown for each locus). Error bars denote the s.d. of quadruplicate runs. bp, base pairs; Del, deletion; Dup, duplication.

Besides a segment-based scoring approach for CNV association, an alternative method is the gene-based scoring approach that examines CNV calls affecting any region of the gene (Supplementary Fig. 5, lower panel). Using this approach, we further identified seven genes with an increased frequency of CNVs in ASD cases versus controls (Supplementary Table 3). For each gene, most CNVs target different parts of the gene and would have been missed by the segment-based approach. Of note, four of the genes identified by the segment- and gene-based approaches are involved in neuron development (*NRXN1*, *CNTN4*, *ASTN2* and *NLGN1*) (Gene Ontology term 'neuron development', $P = 9.5 \times 10^{-3}$). Therefore, by combining evidence from two complementary CNV association approaches, the large sample size has enabled us to implicate two specific gene networks or biological pathways in ASDs: the ubiquitination system and neuronal cell-adhesion molecules.

The genes from the ubiquitin pathway (*UBE3A*, *PARK2*, *RFWD2* and *FBXO40*) represent a new CNV finding in ASD susceptibility. Ubiquitination is a post-translational modification which can rapidly alter protein function and target proteins for proteasome-mediated degradation. The ubiquitin–proteasome system operates pre- and post-synaptic compartments, regulating synaptic attributes, including neurotransmitter release, synaptic vesicle recycling in pre-synaptic terminals, and dynamic changes in dendritic spines and the post-synaptic density (PSD)[20]. Of the four ubiquitin-related genes highlighted in our study, *UBE3A*, a ubiquitin protein ligase, has been the most extensively studied in the context of autism. *PARK2* is a ubiquitin-protein ligase, mutations of which cause autosomal recessive juvenile Parkinson's disease[21], and *RFWD2* and *FBXO40* are also ubiquitin-protein ligases, but neither has been previously associated with disease-causing mutations. The role of ubiquitin in the turnover of synaptic components such as the neuronal cell-adhesion molecules in a process involving regulation of activity-dependent synaptic plasticity presents a mechanism that links these two major gene networks. In addition to the genes described above, several ubiquitin-related genes are involved in human neurological diseases. These include *NHLRC1*, *UBR*, *CUL4B*, *BRWD3* and *HUWE1*, genes that encode ubiquitin protein E3 ligases. Mutations in the latter three and in

UBE2A, an E2 ubiquitin-conjugating enzyme, cause syndromes that include intellectual disability[22].

Genes from the second group of genes implicated in our study, neuronal cell-adhesion molecules, are critical in the development of the nervous system, contributing to axonal guidance, synaptic formation and plasticity, and neuronal–glial interactions. Recent genetic evidence has suggested associations between autism susceptibility and neuronal cell-adhesion molecules, including *NRXN1* (ref. 10), *CNTNAP2* (ref. 23), *NLGN3* (ref. 17), *NLGN4X* (ref. 17), and specific cadherins. Our results provide support for some previously reported genes (*NRXN1* and *CNTN4*), and also implicate additional genes with cell-adhesion functions, including *NLGN1* and *ASTN2*. Mutations in neuroligin superfamily members have previously been found in individuals with autism, and have subsequently been shown to be functionally relevant[24]. *ASTN1*, a well-studied homologue of *ASTN2*, is a neuronal protein receptor integral in the process of glial-guided granule cell migration during development[25], and *ASTN2* deletions have recently been associated with schizophrenia[26].

Using a genome-wide approach for high-resolution CNV detection, we have identified candidate genomic loci with enrichment of CNVs in ASD cases as compared to controls, and have replicated many of them using an independent set of cases. Most of these genes fall within two pathways/networks involving neuronal cell-adhesion and ubiquitin degradation. The enrichment of genes within these molecular systems suggests new susceptibility mechanisms for ASDs. Our results call for functional and expression assays to be completed to assess the biological effects of CNVs in these candidate genes.

METHODS SUMMARY

All genome-wide SNP genotyping was performed using the InfiniumII HumanHap550 BeadChip at the Center for Applied Genomics at The Children's Hospital of Philadelphia (CHOP). We called CNVs with the PennCNV algorithm[15], which combines multiple values, including Log R Ratio, B Allele Frequency, SNP spacing and population frequency of the B allele into a hidden Markov model. The term 'CNV' represents individual CNV calls, whereas 'CNVR' refers to population-level variation. Quality control thresholds included a high success rate of attempted SNPs, low standard deviation of normalized intensity, genetically inferred European ancestry, low genomic wave artefacts, count of CNV calls per subject, and genotypic duplicate removal (Supplementary Table 4). CNV frequency between cases and controls was evaluated at each SNP using Fisher's exact test. We report statistical local minimums in reference to a region of nominal significance including SNPs residing within 1 Mb of each other (Supplementary Fig. 6). Resulting significant CNVRs were excluded if they were (1) residing on telomere or centromere proximal cytobands; (2) arising from a 'peninsula' of common CNV (Supplementary Fig. 7); (3) genomic regions with extremes in GC content[27]; or (4) samples contributing to multiple CNVRs. To adjust for siblings in the AGRE data, we calculated a permutation-based *P* value (×1,000), in which disease labels for siblings were permutated together. DAVID (Database for Annotation, Visualization, and Integrated Discovery)[28] assessed the significance of functional annotation clustering. We considered loci significant between cases and controls ($P < 0.05$) where ACC discovery cases had overlapping variation, replicated in AGRE or were not observed in control subjects, and validated with another method (qPCR Roche Universal Probe Library using qBase[29], MRC-Holland MLPA, and Affymetrix 5.0 from Broad). Statistical correction of five deletion and nine duplication CNVRs, on the basis of discovery cohort (ACC) significance and signal review is appropriate for our study ('CNV Filtering Steps' in Supplementary Materials).

Received 13 November 2008; accepted 27 February 2009.
Published online 28 April 2009.

1. Autism and Developmental Disabilities Monitoring Network. ⟨http://www.cdc.gov/mmwr/pdf/ss/ss5601.pdf⟩ (2007).
2. Newschaffer, C. J. *et al.* The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health* **28,** 235–258 (2007).
3. Gupta, A. R. & State, M. W. Recent advances in the genetics of autism. *Biol. Psychiatry* **61,** 429–437 (2007).
4. Klauck, S. M. Genetics of autism spectrum disorder. *Eur. J. Hum. Genet.* **14,** 714–720 (2006).
5. Vorstman, J. A. S. *et al.* Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry* **11,** 18–28 (2006).
6. Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316,** 445–449 (2007).
7. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39,** 319–328 (2007).
8. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–488 (2008).
9. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
10. Kim, H. G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 199–207 (2008).
11. Roohi J. . *et al.* Disruption of contactin 4 in three subjects with autism spectrum disorder. *J. Med. Genet.* **46,** 176–182 (2008).
12. Fernandez, T. *et al.* Disruption of *Contactin 4* (*CNTN4*) results in developmental delay and other features of 3p deletion syndrome. *Am. J. Hum. Genet.* **82,** 1385 (2008).
13. Le Couteur, A. *et al.* Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *J. Autism Dev. Disord.* **38,** 362–372 (2008).
14. Geschwind, D. H. *et al.* The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* **69,** 463–466 (2001).
15. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17,** 1665–1674 (2007).
16. Kalscheuer, V. M. *et al.* Mutations in *autism susceptibility candidate 2* (*AUTS2*) in patients with mental retardation. *Hum. Genet.* **121,** 501–509 (2007).
17. Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genet.* **34,** 27–29 (2003).
18. Moessner, R. *et al.* Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* **81,** 1289–1297 (2007).
19. Kent W. J. . *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
20. Yi, J. J. & Ehlers, M. D. Ubiquitin and protein turnover in synapse function. *Neuron* **47,** 629–632 (2005).
21. Kitada, T. *et al.* Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* **392,** 605–608 (1998).
22. Tai, H.-C. & Schuman, E. Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. *Nature Rev. Neurosci.* **9,** 826–838 (2008).
23. Alarcón, M. *et al.* Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82,** 150–159 (2008).
24. Chubykin, A. A. *et al.* Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. *J. Biol. Chem.* **280,** 22365–22374 (2005).
25. Zheng, C., Heintz, N. & Hatten, M. E. CNS gene encoding astrotactin, which supports neuronal migration along glial fibers. *Science* **272,** 417–419 (1996).
26. Kahler, A. K. *et al.* Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **147B,** 1089–1100 (2008).
27. Diskin, S. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36,** e126 (2008).
28. Dennis, G. Jr *et al.* DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* **4,** R60 (2003).
29. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8,** R19 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully thank all the ASD children and their families at the participating study sites who were enrolled in this study and all the control subjects who donated blood samples to Children's Hospital of Philadelphia for genetic research purposes. We thank the technical staff in the Center for Applied Genomics, Children's Hospital of Philadelphia for generating the genotypes used in this study. We thank S. Diskin for her contribution to the discussion on the effect of wave artifacts on CNV calling. We also thank S. Kristinsson, L. A. Hermannsson and A. Krisbjörnsson for their software design and contribution. This research was financially supported by The Children's Hospital of Philadelphia, Autism Speaks, and NICHD (HD35476). We also gratefully acknowledge the resources provided by the AGRE Consortium (D. H. Geschwind, M. Bucan, W. T. Brown, J. D. Buxbaum, R. M. Cantor, J. N. Constantino, T. C. Gilliam, C. M. Lajonchere, D. H. Ledbetter, C. Lese-Martin, J. Miller, S. F. Nelson, G. D. Schellenberg, C. A. Samango-Sprouse, S. Spence, M. State, R. E. Tanzi) and the participating families. AGRE is a program of Autism Speaks and is at present supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to C. M. Lajonchere (PI), and formerly by grant MH64547 to D. H. Geschwind (PI). The AGRE data set was genotyped by the Center for Applied Genomics at the Children's Hospital of Philadelphia, which funded all genotyping for this study, and the complete sets of genotyping data have been released to the public domain. AGRE-approved academic researchers can acquire the data sets from AGRE (http://www.agre.org). The study is supported in part by Research Award from the Margaret Q. Landenberger Foundation (H.H.), UL1-RR024134-03 (H.H.), a Research Development Award from the Cotswold Foundation (H.H. and S.F.A.G.), the Beatrice and Stanley A. Seaver Foundation (J.D.B.), the Department of Veterans Affairs (G.D.S.), and National Institute of Health grants HD055782-01 (J.Munson, A.E., O.K., G.D. and G.D.S.), MH0666730 (J.D.B.), MH061009 and NS049261 (J.S.S.), HD055751 (E.H.C.), U10MH66766-02S1

(C.J.M.), MH69359, M01-RR00064 and the Utah Autism Foundation (H.C., J.Miller, and W.M.M.). All genotyping for this study was supported by an Institutional Development Award to the Center for Applied Genomics (H.H.) from the Children's Hospital of Philadelphia. We acknowledge the Autism Genome Project Consortium (J.P., C.W.B., T.H.W., W.M.M., H.C., J.I.N., J.S.S., E.H.C., J.Munson, A.E., O.K., J.D.B., B.D. and G.D.S.) funded by Autism Speaks, the Medical Research Council (UK) and the Health Research Board (Ireland). We are grateful for the public access to the Affymetrix 5.0 data available for a subset of AGRE families (http://www.agre.org), which served as important validation of CNV calls made.

Author Contributions H.H. and G.D.S. designed the study and supervised the data analysis and interpretation. J.T.G., K.W. and B.D. conducted the statistical analyses. C.E.K. and E.C.F. directed the genotyping of stage 1. J.D.B. coordinated the validation. G.C. and O.K. performed qPCR and MLPA validation of CNVs and edited the manuscript. J.T.G., K.W., H.H., G.D.S. and B.D. drafted the manuscript. G.D.S., N.J.M., E.H.C., W.M.M., H.C., T.H.W., J.D.B., T.O., J.I.N., E.A., L.S., J.R., T.S., C.B., C.J.M., D.J.P. and D.Z. collected samples and contributed phenotype data for the study and assisted with data collection and manuscript preparation. E.C., S.F.A.G., P.S., M.I., B.D., L.K., S.W. and K.W. reviewed the data, assisted with interpretation of the data, and edited the manuscript. Other authors contributed to sample acquisition and processing or to data analysis and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.H. (hakonarson@chop.edu) or G.D.S (gerardsc@mail.med.upenn.edu).

Vol 459|28 May 2009|doi:10.1038/nature08029

LETTERS

Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors

Stéphane Rivière[1]*, Ludivine Challet[1]*, Daniela Fluegge[2]*†, Marc Spehr[2]† & Ivan Rodriguez[1]

Mammals rely heavily on olfaction to interact adequately with each other and with their environment[1]. They make use of seven-transmembrane G-protein-coupled receptors to identify odorants and pheromones. These receptors are present on dendrites of olfactory sensory neurons located in the main olfactory or vomeronasal sensory epithelia, and pertain to the odorant[2], trace amine-associated receptor[3] and vomeronasal type 1 (ref. 4) or 2 (refs 5–7) receptor superfamilies. Whether these four sensor classes represent the complete olfactory molecular repertoire used by mammals to make sense of the outside world is unknown. Here we report the expression of formyl peptide receptor-related genes by vomeronasal sensory neurons, in multiple mammalian species. Similar to the four known olfactory receptor gene classes, these genes encode seven-transmembrane proteins, and are characterized by monogenic transcription and a punctate expression pattern in the sensory neuroepithelium. *In vitro* expression of mouse formyl peptide receptor-like 1, 3, 4, 6 and 7 provides sensitivity to disease/inflammation-related ligands. Establishing an *in situ* approach that combines whole-mount vomeronasal preparations with dendritic calcium imaging in the intact neuroepithelium, we show neuronal responses to the same molecules, which therefore represent a new class of vomeronasal agonists. Taken together, these results suggest that formyl peptide receptor-like proteins have an olfactory function associated with the identification of pathogens, or of pathogenic states.

In an attempt to understand better the molecular tools used by mammals to interact socially, we searched for currently unidentified mouse vomeronasal chemoreceptors. We performed a screen based on polymerase chain reaction with reverse transcription (RT–PCR) on mouse vomeronasal sensory tissue using primers corresponding to known and putative heptahelical receptor transcripts (Supplementary Table 1); we identified five transcripts, corresponding to members of the formyl peptide receptor (FPR) gene family. FPRs are found in all mammals and are encoded in humans and mice by three and seven genes, respectively[8] (Supplementary Fig. 1a, b). In mice, FPR genes are located on chromosome 17, all of them found in a region encompassing 2.7 megabases (Supplementary Fig. 1a). Intrafamily sequence identities between FPRs range from 67% to 96%, whereas they share 38% identity with their closest relative, the G-protein-coupled receptor Cmklr1. Two mouse FPR genes not transcribed in vomeronasal neurons, *Fpr1* and *Fpr2* (also called *Fpr-rs2*), have been shown to have a role in immunity and to be expressed in immune cells and other cell types[8]. We evaluated by quantitative RT–PCR the expression of all mouse FPR genes in a broad range of tissues and organs. Presence of *Fpr-rs1*, *Fpr-rs3*, *Fpr-rs4*, *Fpr-rs6* and *Fpr-rs7* transcripts was exclusively observed in vomeronasal tissue extracts (Fig. 1i and Supplementary Fig. 2).

In situ analysis of *Fpr-rs1*, *Fpr-rs3*, *Fpr-rs4*, *Fpr-rs6* and *Fpr-rs7* transcription showed strong and punctate expression in the vomeronasal neuroepithelium (Fig. 1b–f), similar to patterns observed for vomeronasal type 1 (V1R) and type 2 (V2R) genes[4–7]. *Fpr-rs1*, *Fpr-rs3*, *Fpr-rs4*, *Fpr-rs6* and *Fpr-rs7*-expressing neurons represented 0.7%, 0.8%, 0.6%, 1.2% and 0.8% of vomeronasal neurons, respectively, which corresponds to a total of 4.1%. To investigate further FPR-rs expression we raised a rabbit antiserum against FPR-rs3 (Supplementary Fig. 3). Immunohistochemical localization of FPR-rs3 on vomeronasal sections showed a punctate expression pattern corresponding to the staining of single sensory neurons, similar to the one obtained by *in situ* hybridization. The FPR-rs3 receptor was present in sensory neuron somata as well as dendritic endings; that is, where contact with the outside world takes place (Fig. 1g). We evaluated the potential expression of *Fpr-rs3* in rat vomeronasal slices and found punctate messenger RNA (Supplementary Fig. 4a, b) and protein expression (Fig. 1h) similar to what was observed in mice. Comparable results were obtained on gerbil vomeronasal sensory tissue (Supplementary Fig. 4c). The

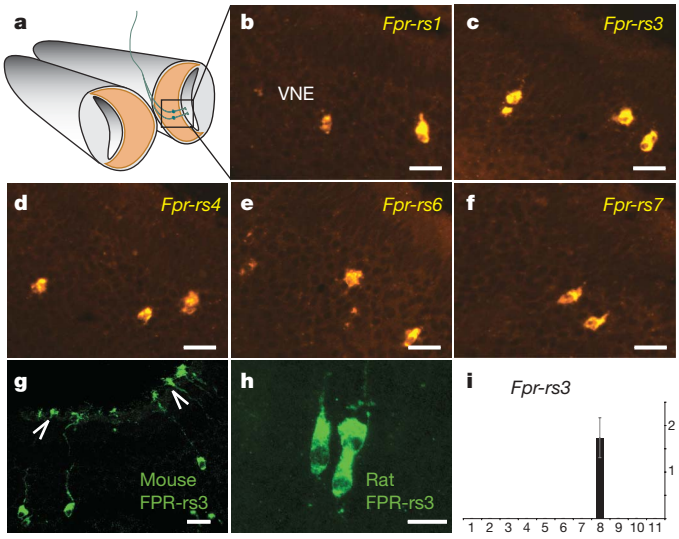


Figure 1 | FPR genes are expressed by vomeronasal sensory neurons. **a**, Schematic of a vomeronasal organ sectioned coronally. **b**–**f**, DIG-labelled antisense probes of *Fpr-rs1*, *Fpr-rs3*, *Fpr-rs4*, *Fpr-rs6* and *Fpr-rs7* hybridized to coronal sections of mouse vomeronasal neuroepithelium show punctate expression. **g**, Immunohistochemical localization of FPR-rs3 in mouse vomeronasal neuron somata and dendritic endings (arrows). **h**, Expression of FPR-rs3 in rat vomeronasal neuroepithelium. **i**, RT–qPCRs with primers specific for mouse *Fpr-rs3* performed on muscle (1), kidney (2), intestine (3), lung (4), spleen (5), thymus (6), cerebellum (7), vomeronasal organ (8), main olfactory epithelium (9), tongue (10) and blood (11) extracts. *Fpr-rs3* transcripts are only found in vomeronasal tissue. VNE, vomeronasal neuroepithelium. Scale bars, 40 μm.

[1]Department of Zoology and Animal Biology, and National Center of Competence 'Frontiers in Genetics', University of Geneva, 1205 Geneva, Switzerland. [2]Department of Cellular Physiology, Ruhr University, 44780 Bochum, Germany. †Present address: Department of Chemosensation, Institute of Biology II, RWTH Aachen University, 52074 Aachen, Germany. *These authors contributed equally to this work.

vomeronasal neuroepithelium is divided into two molecularly and possibly functionally distinct zones, one referred to as apical, expressing V1Rs and $G\alpha_{i2}$, the other basal and expressing V2Rs and $G\alpha_o$. Neurons transcribing *Fpr-rs3*, *Fpr-rs4*, *Fpr-rs6* and *Fpr-rs7* were localized in the apical $G\alpha_{i2}$-expressing layer of the neuroepithelium (Supplementary Fig. 5).

Members of a given olfactory receptor gene family are not coexpressed with members of another family. To evaluate whether this characteristic was shared by FPR genes, we took an approach based on double *in situ* hybridizations, using a large panel of probes recognizing most chemoreceptors expressed by vomeronasal sensory neurons[9]. A total of 2,293 vomeronasal neurons expressing *Fpr-rs3* were examined; none showed co-labelling with any of the probes (Fig. 2a–k), indicating a transcription mechanism that excludes the coexpression of vomeronasal chemoreceptors pertaining to other families. A second peculiar characteristic shared by most olfactory chemoreceptor genes is monogenic expression; that is, the transcription of a single chemoreceptor gene, from a given family, per sensory neuron. We investigated the potential co-transcription of different members of the FPR family. Vomeronasal sections were co-hybridized with an *Fpr-rs3*-specific probe and probes recognizing *Fpr-rs1*, *Fpr-rs4*, *Fpr-rs6* or *Fpr-rs7*. A total of 911 *Fpr-rs3*-expressing vomeronasal neurons were visualized; none showed coexpression (Fig. 2l–o), indicating a monogenic transcription similar to the one observed for odorant or V1R receptor genes[9].

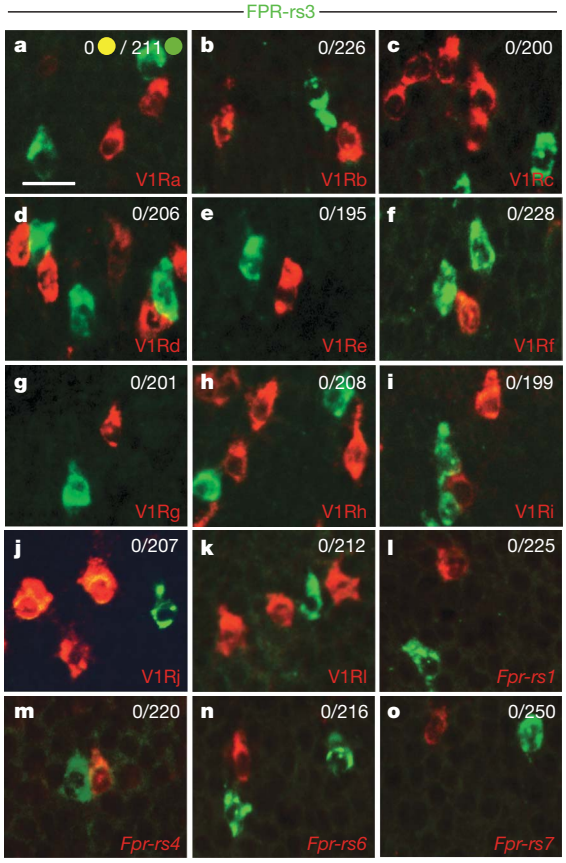


Figure 2 | Exclusive expression of FPR genes in vomeronasal neurons. **a–k**, Double *in situ* hybridizations with probes covering the whole V1R repertoire (red) and *Fpr-rs3* (green) show the absence of coexpression between *Fpr-rs3* and V1R genes. The *x*/*y* ratio indicates the observed number (*x*) of neurons coexpressing *Fpr-rs3* and a member of the tested V1R family, among *y* *Fpr-rs3*-positive neurons. Intrafamily monogenic expression, which is observed for V1Rs, also characterizes *Fpr-rs3* (**l–o**), the transcripts of which do not co-localize with *Fpr-rs1*, *Fpr-rs4*, *Fpr-rs6* and *Fpr-rs7* transcripts. The *x*/*y* ratio indicates the number (*x*) of neurons coexpressing *Fpr-rs3* and a member of the tested FPR family, among *y* *Fpr-rs3*-positive neurons. Scale bar, 20 μm.

FPR1 and FPR2 are expressed by granulocytes, monocytes, or macrophages, and recognize multiple agonists. These include molecules associated with inflammation or disease, such as formyl peptides which are released by Gram-negative bacteria, HIV-derived peptides, the antimicrobial peptide CRAMP, lipoxin A4, SMUPAR-HIS (uPAR) and the acute phase protein SAA[8,10]. No ligands are known for mouse FPR-rs3, FPR-rs4, FPR-rs6 or FPR-rs7 receptors. To explore their agonist characteristics, we transiently expressed FPR-rs1 as well as FPR-rs3, FPR-rs4, FPR-rs6 and FPR-rs7 in HEK293 cells. We hypothesized that FPR-rs1, FPR-rs3, FPR-rs4, FPR-rs6 and/or FPR-rs7 could respond to FPR1 or FPR-rs2 agonists; these potential ligands were evaluated by measuring agonist-induced calcium transients in cells loaded with the ratiometric Ca^{2+}-sensitive reporter dye fura-2/AM. Ca^{2+} transients were recorded after addition of 9 μM *N*-formyl-methionyl-leucyl-phenylalanine (fMLF), 0.9 μM lipoxin A4, 5 μM CRAMP, and 1 μM uPAR (Fig. 3a–d). Every transfected FPR-rs provided responsiveness to at least one ligand. Desensitization was observed after exposure to a single agonist (Supplementary Fig. 6). FPR-rs responded to varying degrees to most stimuli (Fig. 3d), an observation consistent with the very widely tuned agonist characteristics of FPRs expressed in the immune system, such as FPR1 and FPR2 (refs 8, 10, 11).

Because our findings suggested a capacity of the vomeronasal organ to perceive compounds related to bacteria and inflammation, we measured potential calcium transients elicited on vomeronasal neurons by the addition of these compounds. We took two independent approaches: (1) we measured cytosolic Ca^{2+} signals in dissociated vomeronasal sensory neurons; and (2) we established an intact whole-mount preparation allowing confocal Ca^{2+} imaging in individual dendritic knobs without disrupting the integrity of the sensory epithelium or axonal projections.

First, acutely dissociated vomeronasal sensory neurons, loaded with fura-2/AM, were stimulated by application of the formyl peptide fMLF (0.9 μM and 9 μM), CRAMP (0.9 μM), lipoxin A4 (0.9 μM), diluted urine (1:100) and an elevated extracellular K^+ solution (40 mM) (Fig. 4b, c). The integrated fluorescence ratio (f_{340}/f_{380}) in defined regions of interest covering large areas of individual cell somata was monitored as a function of time. From a total of 598 K^+-sensitive vomeronasal neurons investigated in 105 independent experiments (14 animals), 143 neurons (23.9%) responded to diluted urine. At a concentration of 0.9 μM, 5.5%, 4.6% and 11.6% of the recorded vomeronasal neurons responded to fMLF, lipoxin A4 and CRAMP application respectively (Supplementary Fig. 7 and Supplementary Movie 1). A higher concentration of fMLF (9 μM) triggered responses in 9.9% of neurons (Supplementary Fig. 7).

We then tested the response profiles of ligand-sensitive vomeronasal neurons. When fMLF, CRAMP and lipoxin A4 (0.9 μM each) were applied sequentially (in random order), response profiles were not homogenous: 23 out of 28 CRAMP-sensitive vomeronasal sensory neurons (82.1%), 7 out of 11 lipoxin A4-sensitive neurons (63.6%) and 8 out of 11 fMLF-sensitive cells (72.7%) responded exclusively to CRAMP, lipoxin A4 and fMLF, respectively. The remaining neurons were sensitive to two out of the three compounds (two vomeronasal sensory neurons responded to CRAMP and fMLF; three vomeronasal sensory neurons responded to CRAMP and lipoxin A4; one vomeronasal sensory neuron responded to lipoxin A4 and fMLF), whereas no vomeronasal neuron ($n = 50$) responded to all three ligands (Fig. 4c). These data are consistent with the expression of different FPR-rs genes (and therefore non-identical agonist profiles) in randomly chosen vomeronasal neurons.

The second strategy was based on the analysis of single neuron sensitivity to FPR-rs agonists in the intact vomeronasal sensory neuroepithelium. We established a vomeronasal organ whole-mount imaging assay based on Ca^{2+} transients, termed *en face* Ca^{2+} imaging, and recorded responses from individual dendritic knobs of fluo-4/AM-loaded vomeronasal neurons (Fig. 4a, b). This approach presents several advantages: (1) dendritic tips are kept covered by mucus;

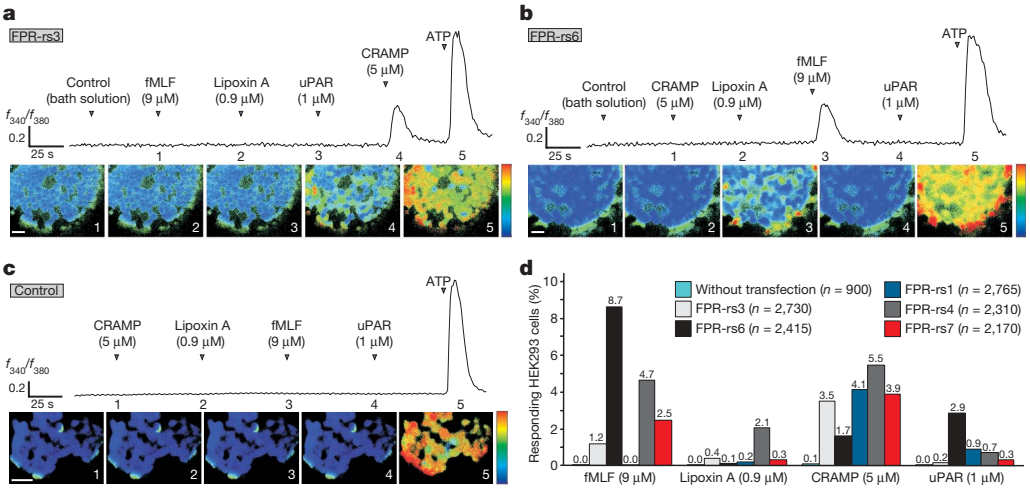


Figure 3 | Recognition by FPRs of pathogen/inflammation-related compounds. **a**–**c**, Representative original recordings of Ca^{2+} signals in either HEK293 cells expressing FPR-rs3 (**a**) or FPR-rs6 (**b**) or untransfected controls (**c**). Relative cytosolic Ca^{2+} concentrations are depicted in pseudocolour (rainbow spectra: low Ca^{2+}, blue; high Ca^{2+}, red). The integrated fluorescence ratio (f_{340}/f_{380}) in a user-defined region of interest is depicted as a function of time. Numbered images correspond to individual frames at indicated time points (1–5). FPR-rs3-transfected cells respond to CRAMP (5 μM), whereas FPR-rs6-expressing cells respond to fMLF (9 μM). Untransfected HEK293 cells did not respond to any of the stimuli. **d**, Semi-quantitative bar chart illustrating the ligand–receptor interactions as measured by Ca^{2+} assays in FPR-expressing HEK293 cells. The response 'probability' is plotted in per cent as a function of both stimulus and receptor identity. Recordings ($n \geq 50$) from at least four independent transfections were repeated multiple times for each possible receptor–ligand combination. Scale bars, 50 μm.

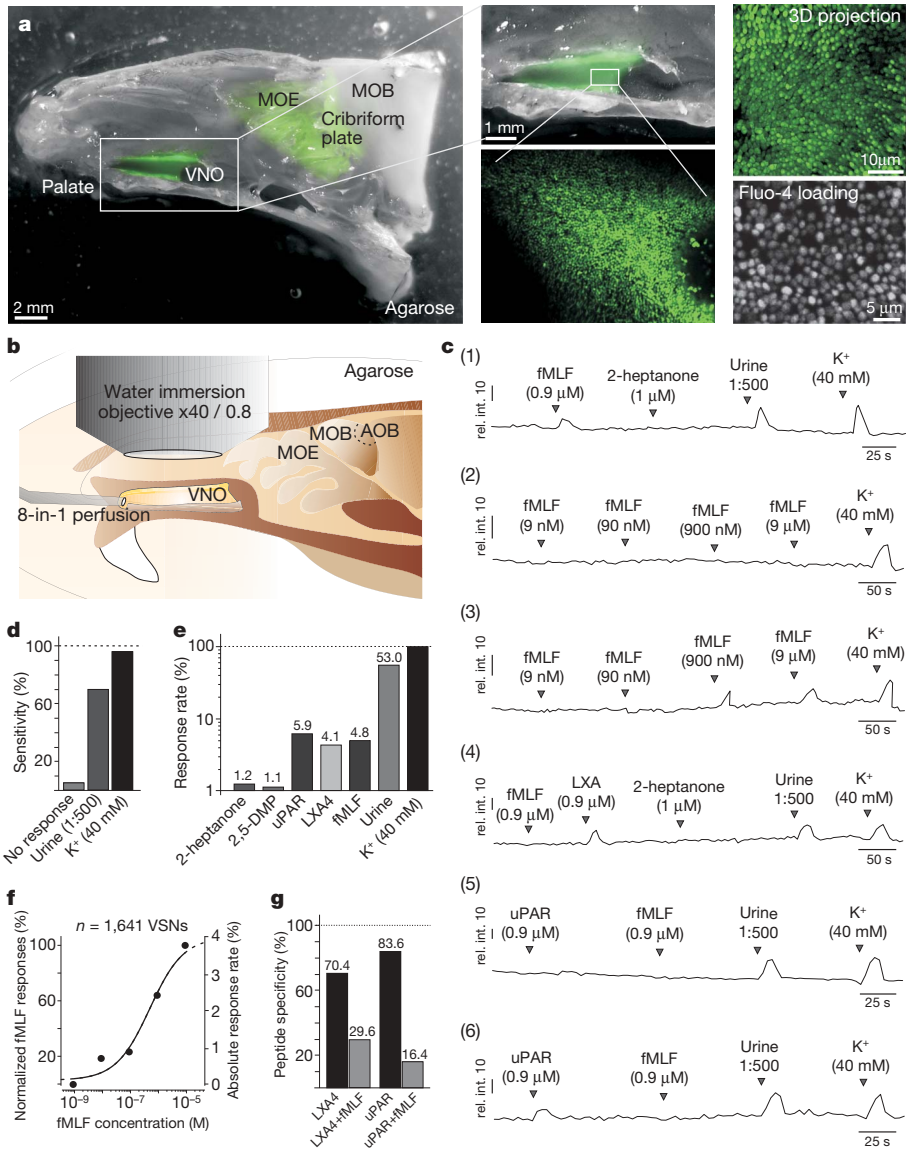


Figure 4 | Formyl-peptide-evoked Ca^{2+} signals in the intact vomeronasal epithelium. **a**, The *en face* confocal Ca^{2+} imaging approach established in this study. Left: merged macroscopic bright-field and fluorescence images of the hemisected rostral head of an OMP–GFP mouse. The vomeronasal organ was opened laterally and the cavernous tissue was removed (see inset at higher magnification; middle panel, top). Middle panel, bottom: low-magnification *en face* confocal scan of the dendritic surface of the vomeronasal organ sensory epithelium area shown above. Right panel, top: three-dimensional transparency projection of a high-magnification confocal *z*-stack through the apical part of the sensory epithelium. Right panel, bottom: high-magnification confocal section of fluo-4/AM-loaded vomeronasal neurons. **b**, Schematic diagram showing the experimental set-up. **c**, Representative original recordings (panels 1 to 6) of cytosolic Ca^{2+} signals in single dendritic vomeronasal neuron knobs in response to diverse stimuli. The integrated fluorescence intensities in user-defined regions of interest are displayed in arbitrary units and viewed as a function of time. **d**, Bar graph showing general vomeronasal neuron viability as 95% of all neurons investigated being sensitive to membrane depolarization. **e**, Semi-logarithmic bar chart illustrating the stimulus-dependent response rate in K^+-sensitive vomeronasal neurons. **f**, Quasi-dose–response curve depicting the concentration dependence of fMLF sensitivity. The percentage of responding vomeronasal neurons was normalized to the number of neurons activated by 9 μM fMLF. Data points are fit to a Hill equation. **g**, Tuning profile of formyl-peptide-sensitive vomeronasal neurons sequentially exposed to uPAR, lipoxin A4 (LXA4) or fMLF (0.9 μM each). AOB, accessory olfactory bulb; MOB, main olfactory bulb; MOE, main olfactory epithelium.

(2) neurons stay surrounded by sustentacular cells; and (3) axonal projections to the accessory olfactory bulb are kept intact. Fifteen to twenty-five individual knobs were typically visualized per experiment (Supplementary Fig. 8b), among which 95% responded to K^+-dependent membrane depolarization with a transient increase in cytosolic Ca^{2+} (Fig. 4d and Supplementary Movie 2). The remaining K^+-insensitive neurons (5%) were discarded from further analysis. Application of known vomeronasal stimuli including urine, soiled bedding, 2-heptanone, 2,5-dimethylpyrazine (2,5-DMP) and the SYFPEITHI peptide induced repeated Ca^{2+} transients in vomeronasal dendrites (Supplementary Fig. 8a–c and Supplementary Movie 2). In 25 independent control experiments, repeated focal application of standard bath solution never elicited a response (data not shown).

In a total of 221 different experiments (21 animals), the vomeronasal surface was sequentially exposed to brief pulses (10 s) of the formyl peptide fMLF, uPAR, or lipoxin A4 (Fig. 4c). Similar to observations in isolated vomeronasal neurons (Supplementary Fig. 7), we recorded responses in 4.8%, 4.1% and 5.9% of vomeronasal neurons after perfusion with fMLF, lipoxin A4 and uPAR, respectively (Fig. 4e). As expected from previously reported data[12], Ca^{2+} transients in response to described vomeronasal stimuli were observed less frequently: 1.2% for 2-heptanone and 1.1% for 2,5-DMP. When increasing concentrations of fMLF were sequentially applied over a five orders of magnitude concentration range (0.9 nM to 9 μM; $n = 1{,}641$ vomeronasal sensory neurons over 82 experiments, Fig. 4c, panels 2 and 3), initial Ca^{2+} signals could be detected at 90 nM fMLF (Fig. 4f). Neurons responding to a given concentration of fMLF invariably responded to higher concentrations of fMLF (Supplementary Fig. 9). Similar to data obtained from dissociated neurons, vomeronasal neurons in the intact epithelium displayed some variability in their individual ligand profile when tested for their responses to fMLF, lipoxin A4 and uPAR (Fig. 4g).

Resulting from our initial identification of FPR-rs receptors in vomeronasal olfactory sensory neurons, our data shed light on novel vomeronasal agonists. These compounds are not part of a group defined by chemical composition or structure, but rather by a link to pathogens, or to inflammation, which may result from an immune response to germs. FPR-rs agonists are mostly produced in tissues and serum after inflammation processes. What could thus be their physiological link to the olfactory system? These molecules are also present in bodily fluids, such as urine, in which CRAMP, uPAR and SAA can be found in situations associated with various diseases[13–15] (a situation reminiscent of the vomeronasal response to sulphated glucocorticoids, molecules whose concentrations in urine depend on the stress level of the animal[16]). Although the functional role of FPR-rs1, FPR-rs3, FPR-rs4, FPR-rs6 and FPR-rs7 receptors in olfactory sensory neurons remains to be determined, their agonist characteristics suggest that they may allow the detection of various contaminated compounds such as spoiled food, or alternatively, enable the identification of unhealthy conspecifics via the evaluation of bodily secretions, a job at which mice excel[17–19].

METHODS SUMMARY

A detailed description of the methods is provided in Supplementary Information. Animals were housed and handled in accordance with the guidelines and regulations of the institution and of the state of Geneva. For immunohistochemistry, a polyclonal anti-FPR-rs3 antiserum was raised by immunization of rabbits with the KSSNFSSCPADSEL peptide. HEK293 cells were transiently transfected with plasmids driving the expression of FPR-rs. Cells were loaded with dye 48 h after transfection to record agonist-induced calcium transients. For preparation of intact vomeronasal sensory epithelia (*en face* confocal Ca^{2+} imaging), we adopted and modified a previously published technique for electro-vomeronasogram recordings[12]. Vomeronasal organ dissection and isolation of single vomeronasal neurons was performed as previously described[20].

Received 14 November 2008; accepted 2 April 2009.
Published online 22 April 2009; corrected 28 May 2009 (see full-text HTML version for details).

1. Kelliher, K. R. The combined role of the main olfactory and vomeronasal systems in social communication in mammals. *Horm. Behav.* **52,** 561–570 (2007).
2. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65,** 175–187 (1991).
3. Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442,** 645–650 (2006).
4. Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83,** 195–206 (1995).
5. Matsunami, H. & Buck, L. B. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90,** 775–784 (1997).
6. Herrada, G. & Dulac, C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90,** 763–773 (1997).
7. Ryba, N. J. & Tirindelli, R. A new multigene family of putative pheromone receptors. *Neuron* **19,** 371–379 (1997).
8. Migeotte, I., Communi, D. & Parmentier, M. Formyl peptide receptors: a promiscuous subfamily of G protein-coupled receptors controlling immune responses. *Cytokine Growth Factor Rev.* **17,** 501–519 (2006).
9. Roppolo, D. *et al.* Gene cluster lock after pheromone receptor gene choice. *EMBO J.* **26,** 3423–3430 (2007).
10. Le, Y. *et al.* Biologically active peptides interacting with the G protein-coupled formylpeptide receptor. *Protein Pept. Lett.* **14,** 846–853 (2007).
11. Le, Y., Murphy, P. M. & Wang, J. M. Formyl-peptide receptors revisited. *Trends Immunol.* **23,** 541–548 (2002).
12. Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405,** 792–796 (2000).
13. Takata, M. *et al.* Detection of amyloid beta protein in the urine of Alzheimer's disease patients and healthy individuals. *Neurosci. Lett.* **435,** 126–130 (2008).
14. Chromek, M. *et al.* The antimicrobial peptide cathelicidin protects the urinary tract against invasive bacterial infection. *Nature Med.* **12,** 636–641 (2006).
15. Casella, R., Shariat, S. F., Monoski, M. A. & Lerner, S. P. Urinary levels of urokinase-type plasminogen activator and its receptor in the detection of bladder carcinoma. *Cancer* **95,** 2494–2499 (2002).
16. Nodari, F. *et al.* Sulfated steroids as natural ligands of mouse pheromone-sensing neurons. *J. Neurosci.* **28,** 6407–6418 (2008).
17. Kavaliers, M., Choleris, E., Agmo, A. & Pfaff, D. W. Olfactory-mediated parasite recognition and avoidance: linking genes to behavior. *Horm. Behav.* **46,** 272–283 (2004).
18. Kavaliers, M. *et al.* Inadvertent social information and the avoidance of parasitized male mice: a role for oxytocin. *Proc. Natl Acad. Sci. USA* **103,** 4293–4298 (2006).
19. Kavaliers, M., Choleris, E. & Pfaff, D. W. Genes, odours and the recognition of parasitized individuals by rodents. *Trends Parasitol.* **21,** 423–429 (2005).
20. Spehr, M., Hatt, H. & Wetzel, C. H. Arachidonic acid plays a role in rat vomeronasal signal transduction. *J. Neurosci.* **22,** 8429–8437 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C.-D. K. Ballif, C. Engelhardt and H. Bartel for technical help, and C. Bauer and P. Descombes for assistance on the NCCR 'Frontiers in Genetics' bioimaging and genomic platforms, respectively. We also thank H. Hatt for unlimited access to departmental infrastructure, and P. Vassalli for comments on the manuscript. This work was supported by the Swiss National Science Foundation, the Claraz, Schmidheiny, Volkswagen and Schlumberger Foundations, and the Emmy Noether program of the Deutsche Forschungsgemeinschaft.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.R. (ivan.rodriguez@unige.ch).

Vol 459|28 May 2009|doi:10.1038/nature07952

LETTERS

Bacteria hijack integrin-linked kinase to stabilize focal adhesions and block cell detachment

Minsoo Kim[1,3], Michinaga Ogawa[2,3], Yukihiro Fujita[2,3], Yuko Yoshikawa[2,3], Takeshi Nagai[2,3], Tomohiro Koyama[4], Shinya Nagai[4], Anika Lange[5], Reinhard Fässler[5] & Chihiro Sasakawa[1,2,3]

The rapid turnover and exfoliation of mucosal epithelial cells provides an innate defence system against bacterial infection[1,2]. Nevertheless, many pathogenic bacteria, including *Shigella*, are able to surmount exfoliation and colonize the epithelium efficiently[3,4]. Here we show that the *Shigella flexneri* effector OspE[5,6] (consisting of OspE1 and OspE2 proteins), which is highly conserved among enteropathogenic *Escherichia coli*, enterohaemorrhagic *E. coli*, *Citrobacter rodentium* and *Salmonella* strains[7], reinforces host cell adherence to the basement membrane by interacting with integrin-linked kinase (ILK)[8]. The number of focal adhesions was augmented along with membrane fraction ILK by ILK–OspE binding. The interaction between ILK and OspE increased cell surface levels of β1 integrin and suppressed phosphorylation of focal adhesion kinase and paxillin, which are required for rapid turnover of focal adhesion in cell motility[9]. Nocodazole-washout-induced focal adhesion disassembly was blocked by expression of OspE. Polarized epithelial cells infected with a *Shigella* mutant lacking the *ospE* gene underwent more rapid cell detachment than cells infected with wild-type *Shigella*. Infection of guinea pig colons with *Shigella* corroborated the pivotal role of the OspE–ILK interaction in suppressing epithelial detachment, increasing bacterial cell-to-cell spreading, and promoting bacterial colonization. These results indicate that *Shigella* sustain their infectious foothold by using special tactics to prevent detachment of infected cells.

The intestinal epithelium acts as an intrinsic barrier against microbial invaders. This barrier is composed of four major functional elements: resident commensal bacteria, rapid epithelial turnover, innate immune responses and epithelial barrier integrity[10,11]. *Shigella* invade and colonize the intestinal barrier and subsequently spread to adjacent cells, which is the essential mechanism that expands replicative niches during infection[12,13]. While establishing colonization, *Shigella* deliver a variety of virulence factors, called effectors, into epithelial cells by means of the type III secretion system[14]. A recent study reported that one of the effectors delivered by intracellular *Shigella sonnei*, OspE2, localizes around focal adhesions (FAs). Furthermore, epithelial cells infected with an Δ*ospE2* mutant show more cell rounding than wild type (WT)-infected cells[15]. Importantly, OspE cognate genes are highly conserved among many enteropathogenic bacteria, including *Shigella*, enteropathogenic *E. coli*, enterohaemorrhagic *E. coli*, *Citrobacter rodentium* and *Salmonella*[7], indicating that OspE has a general role in bacterial infection.

To identify the host cell factor(s) targeted by OspE, we performed a glutathione *S*-transferase (GST)–OspE pull-down assay with HeLa cell lysates. The 50 kDa protein that precipitated with GST–OspE but not GST alone was identified as ILK by a matrix-assisted laser desorption/ionization time-of-flight tandom mass spectrometer (MALDI-TOF/TOF; Fig. 1a and Supplementary Fig. 1). No other focal adhesion components including paxillin, vinculin, talin, β1 integrin or focal adhesion kinase (FAK) were precipitated (Fig. 1a). *In vitro* and

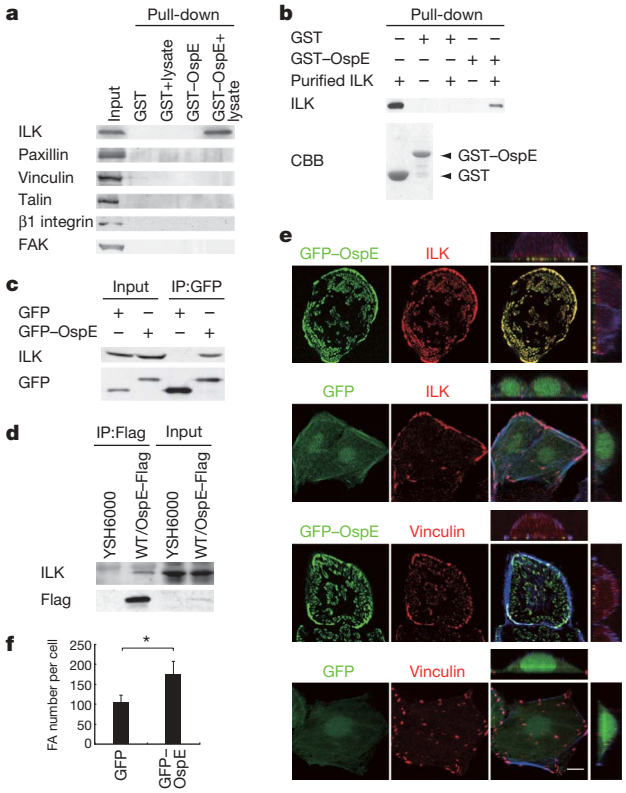


Figure 1 | OspE interacts with ILK at focal adhesions. a, GST–OspE or GST beads were mixed with HeLa cell lysates and the bound proteins were immunoblotted with the antibodies indicated. **b**, GST–OspE or GST beads were mixed with purified ILK from *E. coli* and the bound proteins were immunoblotted with an anti-ILK antibody. **c**, Lysates of HeLa cells expressing GFP–OspE or GFP were immunoprecipitated (IP) with an anti-GFP antibody. The bound proteins were immunoblotted with anti-ILK or anti-GFP antibodies. **d**, Lysates from HeLa cells infected with YSH6000 or WT/OspE–Flag were immunoprecipitated with an anti-Flag antibody. The bound proteins were immunoblotted with an anti-ILK or anti-Flag antibodies. **e**, Co-localization of GFP–OspE with ILK and vinculin at FAs. HeLa cells transfected with pGFP-OspE or pGFP were stained with phalloidin (blue) and an anti-ILK or anti-vinculin antibodies (red). Scale bar, 10 μm. **f**, The number of FAs in **e** was counted by vinculin staining (>20 cells, $n = 3$). Data are mean and s.d. $^*P < 0.001$.

[1]Department of Infectious Disease Control, International Research Center for Infectious Diseases, [2]Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [3]CREST, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan. [4]Nippon Institute for Biological Science, 9-2221-1, Shinmachi, Ome, Tokyo 198-0024, Japan. [5]Department of Molecular Medicine, Max-Planck Institute of Biochemistry, D-82152, Martinsried, Germany.

in vivo binding assays confirmed that ILK interacted with OspE (Fig. 1b–d). Therefore, we examined the subcellular localization of green fluorescent protein (GFP)–OspE and ILK in HeLa cells expressing GFP–OspE. OspE colocalized with ILK at FAs, as visualized by vinculin staining (Fig. 1e). Moreover, the number of FAs increased in OspE-expressing cells (Fig. 1f).

To investigate the functional link between OspE and ILK in FA formation, we retrovirally transduced *Ilk*-deficient mouse fibroblasts[16] ($Ilk^{-/-}$ cells) with WT ILK ($Ilk^{-/-}$ WT-ILK), ILK(K220M) (kinase-domain-inactive mutant[17]; $Ilk^{-/-}$ ILK(K220M)) or ILK(S343D) (constitutively active kinase domain[17]; $Ilk^{-/-}$ ILK(S343D)). These $Ilk^{-/-}$ cells expressing or not expressing wild-type or mutant ILKs were transfected with pGFP-OspE. In the absence of ILK, OspE was dispersed throughout the cytoplasm, whereas OspE colocalized with WT ILK at FAs (Fig. 2a), indicating that the localization of OspE at FAs requires ILK. Colocalization of ILK(K220M) and OspE at FAs was less pronounced, whereas ILK(S343D) colocalized with OspE to a similar extent as WT ILK. To investigate the influence of OspE on FAs, we next retrovirally expressed comparable levels of OspE in rescued $Ilk^{-/-}$ cells ($Ilk^{-/-}$ mock transduced, $Ilk^{-/-}$ OspE, $Ilk^{-/-}$ WT-ILK/mock, $Ilk^{-/-}$ WT-ILK/OspE, $Ilk^{-/-}$ ILK(K220M)/mock, $Ilk^{-/-}$ ILK(K220M)/OspE, $Ilk^{-/-}$ ILK(S343D)/mock and $Ilk^{-/-}$ ILK(S343D)/OspE) and subsequently quantified the number of FAs. $Ilk^{-/-}$ WT-ILK/OspE and $Ilk^{-/-}$ ILK(S343D)/OspE cells contained more FAs than $Ilk^{-/-}$ OspE and $Ilk^{-/-}$ ILK(K220M)/OspE cells (Fig. 2b, c). Because OspE bound to each of the ILK mutants to the same degree as WT ILK (Supplementary Fig. 2a), we tested whether OspE affects ILK kinase activity using an *in vitro* kinase assay. Anti-ILK immunoprecipitates from $Ilk^{-/-}$ mock, $Ilk^{-/-}$ OspE, $Ilk^{-/-}$ WT-ILK or $Ilk^{-/-}$ WT-ILK/OspE cell lysates were incubated with myelin basic protein (MBP). MBP was phosphorylated by WT ILK, regardless of the presence of OspE (Supplementary Fig. 2b). GST–OspE was not phosphorylated by ILK (Supplementary Fig. 2c). Furthermore, the presence of OspE affected the levels of neither phosphorylated AKT nor phosphorylated GSK3β (Supplementary Fig. 2d). OspE did not affect ILK stability (Supplementary Fig. 3a). Moreover, OspE had no effect on the formation of the ILK–PINCH (LIMS1)–parvin (IPP) complex, although OspE bound to the carboxy-terminal region of ILK (Supplementary Fig. 3b–d). These results indicate that OspE does not affect ILK kinase activity or IPP complex formation.

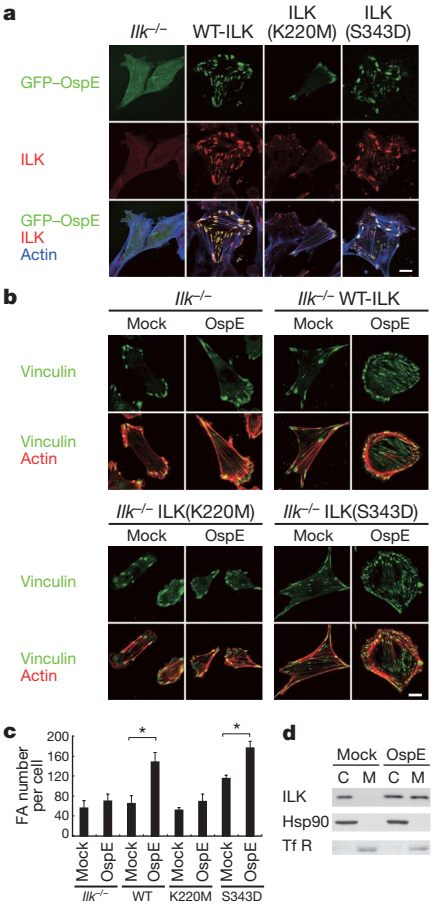


Figure 2 | Interaction between OspE and ILK facilitates FA formation in an ILK kinase domain-dependent manner. a, $Ilk^{-/-}$ cells or the same cells expressing WT ILK, ILK(K220M) or ILK(S343D) were transiently transfected with pGFP-OspE and stained with an anti-ILK antibody (red) and phalloidin (blue). Scale bar, 10 μm. **b**, $Ilk^{-/-}$ cells or the same cells expressing WT ILK, ILK(K220M) or ILK(S343D) with or without OspE were stained with an anti-vinculin antibody (green) and phalloidin (red). Scale bar, 10 μm. **c**, The number of FAs visualized by vinculin staining in **b** was quantified (>10 cells, $n = 3$). Data are mean and s.d. $^*P < 0.001$. **d**, NIH3T3 cells stably expressing mock or OspE were subjected to cytosolic (C) and membrane (M) fractionation. The fractions were immunoblotted with antibodies against ILK, transferrin receptor (Tf R, a membrane marker) and Hsp90 (a cytosolic marker).

Because a membrane-targeted form of ILK strengthens FA formation[18], we performed cell fractionation experiments to determine the effect of OspE on the subcellular localization of ILK in NIH3T3 cells. We found that OspE boosted ILK in the membrane fraction (Fig. 2d). These results indicate that OspE promotes membrane retention of ILK, whereas the ILK–OspE complex stabilizes and strengthens FAs.

ILK has a critical role in coupling the extracellular matrix (ECM) to the actin cytoskeleton as well as signalling complexes, thereby modulating FA assembly and disassembly[19–21]. We therefore investigated whether OspE affects the early cell-matrix adhesion process in FA assembly. $Ilk^{-/-}$, $Ilk^{-/-}$ OspE, $Ilk^{-/-}$ WT-ILK, $Ilk^{-/-}$ WT-ILK/OspE, $Ilk^{-/-}$ ILK(K220M), $Ilk^{-/-}$ ILK(K220M)/OspE, $Ilk^{-/-}$ ILK(S343D) and $Ilk^{-/-}$ ILK(S343D)/OspE cells were seeded on fibronectin-coated coverslips, and then examined for their ability to form FAs. FA formation in $Ilk^{-/-}$ and $Ilk^{-/-}$ ILK(K220M) cells was reduced compared with $Ilk^{-/-}$ WT-ILK and $Ilk^{-/-}$ ILK(S343D) cells (Supplementary Fig. 4a–c), confirming that intact ILK activity contributes to FA assembly[22,23]. Importantly, however, OspE had no effect on FA assembly in any of the examined cells (Supplementary Fig. 4a–c). When we investigated $Ilk^{-/-}$ mock, $Ilk^{-/-}$ OspE, $Ilk^{flox/flox}$ mock and $Ilk^{flox/flox}$ OspE cells for the extent of cell adhesion to defined ECMs, fibronectin, collagen and laminin, we observed that cell adhesion depended on ILK expression but was not further affected by OspE expression (Supplementary Fig. 4d). This finding was confirmed in HeLa cells (Supplementary Fig. 4e). Next, we used the nocodazole-washout method to examine the effect of OspE on FA disassembly (Fig. 3a and Supplementary Fig. 5)[24]. Cells were serum-starved, treated with nocodazole for 30 min to induce FAs, and then washed to remove the nocodazole. After nocodazole was washed out, FAs in $Ilk^{-/-}$, $Ilk^{-/-}$ WT-ILK, $Ilk^{-/-}$ ILK(K220M) and $Ilk^{-/-}$ ILK(S343D) cells disassembled within 30 min. Similar FA disassembly occurred in $Ilk^{-/-}$ OspE and $Ilk^{-/-}$ ILK(K220M)/OspE cells, but not in $Ilk^{-/-}$ WT-ILK/OspE or $Ilk^{-/-}$ ILK(S343D)/OspE cells (Fig. 3a), indicating that the OspE–ILK interaction interferes with FA disassembly.

Because FA disassembly or rapid FA turnover is accompanied by the phosphorylation of FAK and paxillin[9], we next investigated the effect of the OspE–ILK association on FAK phosphorylation. FAK(Y397) phosphorylation levels were higher in $Ilk^{-/-}$ WT-ILK and $Ilk^{-/-}$ ILK(S343D) cells than in $Ilk^{-/-}$ and $Ilk^{-/-}$ ILK(K220M) cells (Fig. 3b), indicating that an intact ILK kinase domain is needed for efficient FAK phosphorylation. In the presence of OspE, however, FAK(Y397) phosphorylation was markedly reduced in $Ilk^{-/-}$ WT-ILK/OspE and $Ilk^{-/-}$ ILK(S343D)/OspE cells. The same was true for tyrosine phosphorylation levels of paxillin and total cellular proteins. These results strongly indicated that the OspE–ILK complex interferes with FA disassembly and thereby dampens FA turnover in an ILK-kinase-domain-dependent manner. Because ILK binds to β1 integrin

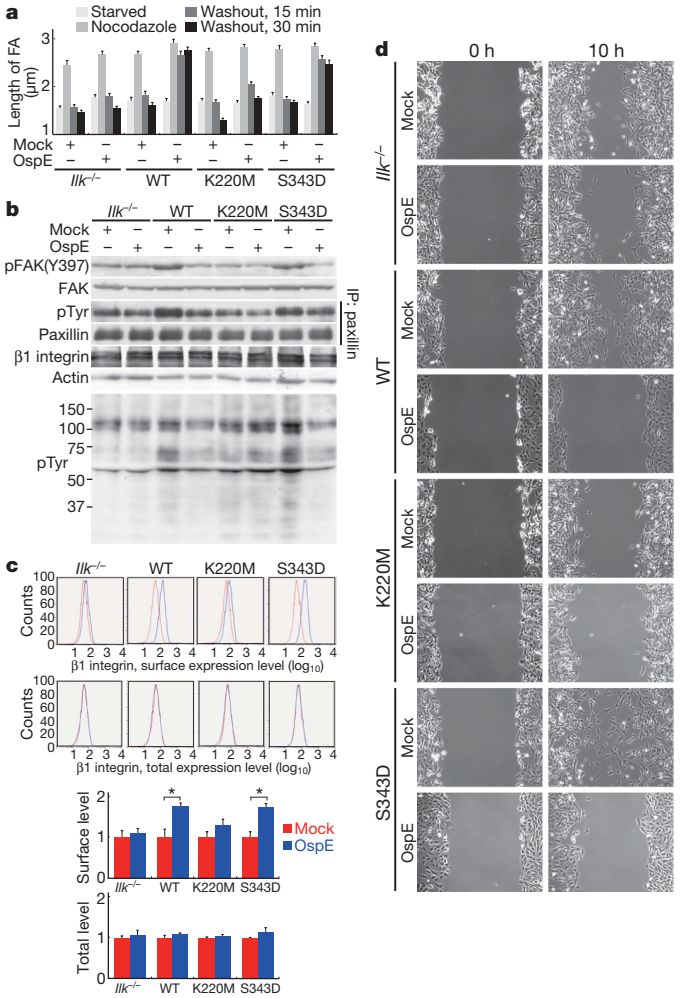


Figure 3 | The interaction between OspE and ILK suppresses FA turnover. **a**, *Ilk*$^{-/-}$ cells or the same cells expressing WT ILK, ILK(K220M) or ILK(S343D) with or without OspE were treated with nocodazole, which was washed out for the indicated periods. The length of the FAs visualized by vinculin staining was quantified (>20 cells, $n = 3$). Data are mean and s.e.m. **b**, Lysates or anti-paxillin immunoprecipitates of the cells were immunoblotted with antibodies as indicated. pTyr, phospho-tyrosine. **c**, The cells were analysed by FACS with an anti-β1 integrin antibody under non-permeabilizing (top) or permeabilizing (bottom) conditions. The bar graphs are representative of three independent experiments. Data are means and s.d. *$P < 0.001$. **d**, The cells were seeded on fibronectin-coated coverslips. The surface was scratched and cells were allowed to migrate into the wound at 37 °C. Wound closure was monitored by time-lapse photography.

and FA turnover is modulated by β1 integrin signalling[25,26], we determined the cell surface levels of β1 integrin by fluorescence-activated cell sorting (FACS; Fig. 3c). The surface levels of β1 integrin were 1.8-fold higher in *Ilk*$^{-/-}$ WT-ILK/OspE and *Ilk*$^{-/-}$ ILK(S343D)/OspE cells than in cells without OspE, whereas OspE did not affect surface β1 integrin levels in *Ilk*$^{-/-}$ or *Ilk*$^{-/-}$ ILK(K220M) cells. ILK did not affect the surface level of β1 integrin. The total β1 integrin levels (cell surface plus cytoplasmic content) were similar in all cells regardless of whether they expressed OspE (Fig. 3c). The surface levels of activated β1 integrin, which was determined by using antibodies that bind to activation-associated epitopes, were increased in *Ilk*$^{flox/flox}$ but not in *Ilk*$^{-/-}$ cells and HeLa cells expressing WT OspE (Supplementary Fig. 6a, b). In line with reduced FA turnover in OspE-expressing WT ILK cells, OspE expression also delayed the turnover of β1 integrin in *Ilk*$^{flox/flox}$ cells (Supplementary Fig. 6c). To demonstrate a functional consequence of the OspE–ILK association on surface β1 integrin levels, *Ilk*$^{-/-}$ and ILK-rescue cells grown at equivalent confluency were scratched and then allowed to migrate into the wounded area in 37 °C (Fig. 3d). Motility was significantly enhanced in cells expressing WT or ILK(S343D)-expressing cells (see *Ilk*$^{-/-}$ WT-ILK and *Ilk*$^{-/-}$ ILK(S343D)) compared with *Ilk*$^{-/-}$ or ILK(K220M)-expressing cells (see *Ilk*$^{-/-}$ and *Ilk*$^{-/-}$ ILK(K220M)). Interestingly, OspE expression blocked the motility of cells expressing WT ILK or ILK with an active kinase domain (see *Ilk*$^{-/-}$ WT-ILK/OspE and *Ilk*$^{-/-}$ ILK(S343D)/OspE). Taken together, these results indicated that OspE–ILK association interferes with FA disassembly and dampens integrin turnover by means of its effect on the phosphorylation of FAK, paxillin and other yet unidentified ILK target proteins.

To pursue the biological significance of the OspE–ILK association, we generated a series of OspE mutants (Supplementary Fig. 7a), and identified OspE(W68A) as a single OspE point mutant that is unable to bind to ILK and localize to FAs (Supplementary Fig. 7a–c). Consistent with the specificity of ILK interaction with OspE, we observed that stable expression of WT OspE but not OspE(W68A) in NIH3T3 cells induces FA formation, suppresses FAK phosphorylation and increases β1 integrin cell surface levels (Supplementary Fig. 7d–g). Interestingly, tryptophan 68 (Trp 68) of OspE is highly conserved among OspE family proteins encoded by *Shigella*, enterohaemorrhagic *E. coli*, enteropathogenic *E. coli*, *C. rodentium*, *S. typhimurium* and *S. enteritidis* (Supplementary Fig. 8a). These OspE homologues were demonstrated to interact with ILK and localize to FAs (Supplementary Fig. 8b, c), implying that a large number of OspE cognates have an important role in bacterial infection.

To establish the role of OspE in *Shigella* infection, we infected HeLa cells with *S. flexneri* (YSH6000, Δ*ospE*, Δ*ospE/ospE* or Δ*ospE/ospE*(W68A)) for 2 h and measured the extent of cell detachment from culture plates. When HeLa cells were infected with Δ*ospE/ospE*, 14% of the cells rounded up, whereby the area of their attachment to the substrate was reduced. In contrast, 42% of cells infected with Δ*ospE/ospE*(W68A) rounded up (Fig. 4a). The ability of OspE to prevent cell detachment during bacterial infection was further confirmed by measuring the rounding up of WT or ILK-depleted HeLa cells infected with YSH6000 or Δ*ospE* (Supplementary Fig. 9a, b). The surface levels of β1 integrin were higher on cells infected with YSH6000 or Δ*ospE/ospE* than on cells infected with Δ*ospE* and Δ*ospE/ospE*(W68A) (Fig. 4b). We also confirmed that membrane fraction ILK increased on infection with *Shigella* bearing WT OspE (Fig. 4c). These data supported the view that OspE promotes membrane retention of ILK, whereas the OspE–ILK complex contributes to inhibition of FA disassembly. Indeed, the effect of OspE on epithelial integrity after bacterial infection was substantial. When polarized MDCK monolayers were infected with a set of bacterial strains for 12 h, cells infected with YSH6000 or Δ*ospE/ospE* maintained basement membrane localization of ILK and normal cell architecture (Fig. 4d). In contrast, cells infected with Δ*ospE* or Δ*ospE/ospE*(W68A) detached from the culture plates, and ILK was dispersed in the cytoplasm (Fig. 4d). The ability of Δ*ospE* and Δ*ospE/ospE*(W68A) to spread intercellularly, as monitored by *Shigella* lipopolysaccharide (LPS) staining or a plaque-forming assay, was greatly impaired compared with YSH6000 or Δ*ospE/ospE* (Fig. 4d and Supplementary Fig. 9c).

To elucidate an *in vivo* role of OspE, guinea pigs were intrarectally inoculated with 1×10^9 cfu of YSH6000, Δ*ospE*, Δ*ospE/ospE* or Δ*ospE/ospE*(W68A). Inoculation of distal colon with YSH6000 or Δ*ospE/ospE* resulted in severe inflammation, internal haemorrhaging and diarrhoea, whereas these pathogenic features were not prominent after Δ*ospE* or Δ*ospE/ospE*(W68A) inoculation (Fig. 4e). Furthermore, 24 h post inoculation, colonization by Δ*ospE* or Δ*ospE/ospE*(W68A) was significantly reduced compared to YSH6000 or Δ*ospE/ospE* (Fig. 4f).

We show that *Shigella* use a previously unknown tactic to strengthen adhesion of infected cells to the underlying extracellular matrix. *Shigella* delivers OspE into the host, which binds to ILK. The ILK–OspE complex stabilizes integrin-containing adhesion sites, reduces adhesion turnover and suppresses the detachment of infected cells from the basement membrane, which ultimately enables the pathogen to maintain an infectious foothold. Our findings uncover

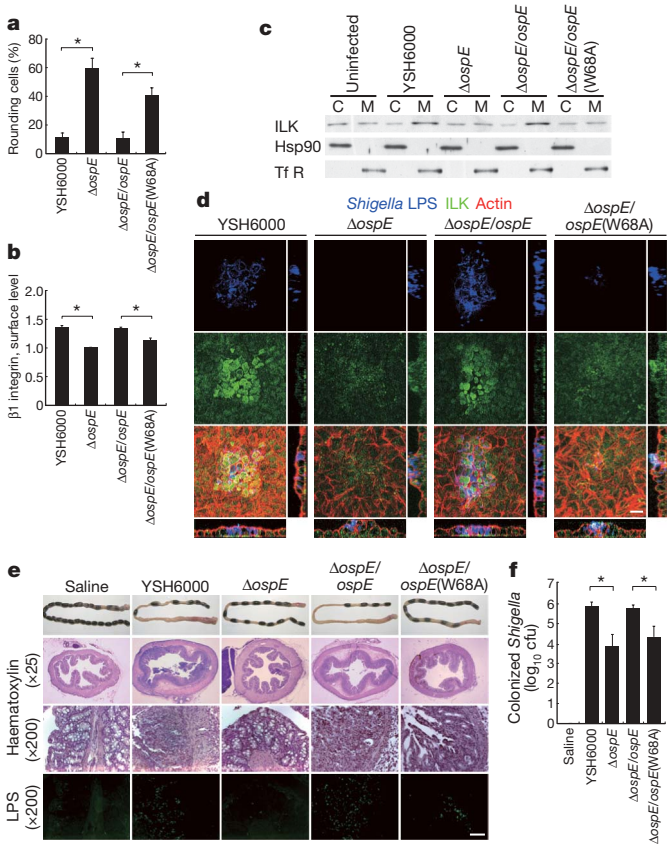


Figure 4 | Establishment of OspE as a *Shigella* virulence factor. a, **b**, HeLa S3 cells were infected with the indicated *Shigella* strains for 2 h. In **a**, the percentage of rounded up cells was calculated (>100 cells, $n = 3$). In **b**, the cells were analysed by FACS with an anti-β1 integrin antibody under non-permeabilizing conditions. The bar graphs are representative of three independent experiments. Data are mean and s.d. $*P < 0.01$. **c**, The cytosolic (C) and membrane (M) fractions of *Shigella* infected HeLa S3 cells were immunoblotted with antibodies as indicated. **d**, MDCK monolayer cells were infected with the indicated *Shigella* strains for 12 h and then stained with an anti-*Shigella* LPS antibody (blue), anti-ILK antibody (green) and phalloidin (red). Confocal *x*–*y*, *x*–*z* and *y*–*z* sectional views are shown. Scale bar, 20 μm. **e**, The large intestines of guinea pigs were dissected 24 h after inoculation and stained with haematoxylin or an anti-LPS antibody. Scale bar, 100 μm. **f**, At 24 h post inoculation, colonized bacteria recovered from the large intestines were determined ($n = 3$). Data are mean and s.d. $*P < 0.01$.

a potential strategy that could be exploited to interfere with infections of *Shigella* and other mucosal pathogens expressing OspE homologues. Discovery of small molecules that inhibit the interaction of OspE with ILK may provide effective therapies for several bacterial infectious diseases.

METHODS SUMMARY

Shigella strains used in this study and procedures for immunoprecipitation, immunostaining, immunoblotting, antibodies, materials, *Shigella* infection and proteins expression are described in Methods. GST pull-down assay, mass spectrometric analysis, cell fractionation, FACS analysis and *in vitro* ILK kinase assay were performed as described in Methods. Retroviral gene transfer and half-life of ILK were performed as described previously[27]. For the wound healing assay, cells grown for 24 h on fibronectin-coated plates were used. Scratches were made in the cell layer with a pipette tip, and then cells were allowed to migrate into the wound at 37 °C for 10 h. Wound closure was monitored with a phase-contrast microscope. FA disassembly was analysed as described previously[24]. In brief, cells were grown on coverslips, serum-starved in DMEM containing 0.5% BSA for 48 h, and then treated with 5 μM of nocodazole at 37 °C for 30 min. The cells were washed to remove the nocodazole, incubated in DMEM containing 0.5% BSA for 15 or 30 min, and then fixed. The size of vinculin-labelled FAs was measured with a confocal laser-scanning microscope. For the spreading assay, cells were trypsinized, replated on fibronectin-pre-coated coverslips in complete medium, and then allowed to spread at 37 °C at the indicated time. Adhesion of mouse fibroblasts to ECM proteins were measured as described previously[28]. An adhesion assay with HeLa S3 cells was performed with the ECM cell adhesion array kit (Chemicon). Guinea-pig infection assays were performed as described previously[29]. The integrin internalization assay was performed as described previously[30]. Statistical analysis was performed using Student's *t*-test or one-way analysis of variance (ANOVA) followed by the Tukey–Kramer post-hoc test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 December 2008; accepted 27 February 2009.
Published online 6 May 2009.

1. Cliffe, L. J. *et al.* Accelerated intestinal epithelial cell turnover: a new mechanism of parasite expulsion. *Science* **308,** 1463–1465 (2005).
2. Mulvey, M. A., Schilling, J. D., Martinez, J. J. & Hultgren, S. J. Bad bugs and beleaguered bladders: interplay between uropathogenic *Escherichia coli* and innate host defenses. *Proc. Natl Acad. Sci. USA* **97,** 8829–8835 (2000).
3. Iwai, H. *et al.* A Bacterial effector targets Mad2L2, an APC inhibitor, to modulate host cell cycling. *Cell* **130,** 611–623 (2007).
4. Muenzner, P., Rohde, M., Kneitz, S. & Hauck, C. R. CEACAM engagement by human pathogens enhances cell adhesion and counteracts bacteria-induced detachment of epithelial cells. *J. Cell Biol.* **170,** 825–836 (2005).
5. Buchrieser, C. *et al.* The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* **38,** 760–771 (2000).
6. Kane, C. D., Schuch, R., Day, W. A. Jr & Maurelli, A. T. MxiE regulates intracellular expression of factors secreted by the *Shigella flexneri* 2a type III secretion system. *J. Bacteriol.* **184,** 4409–4419 (2002).
7. Tobe, T. *et al.* An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc. Natl Acad. Sci. USA* **103,** 14941–14946 (2006).
8. Hannigan, G. E. *et al.* Regulation of cell adhesion and anchorage-dependent growth by a new beta 1-integrin-linked protein kinase. *Nature* **379,** 91–96 (1996).
9. Mitra, S. K., Hanson, D. A. & Schlaepfer, D. D. Focal adhesion kinase: in command and control of cell motility. *Nature Rev. Mol. Cell Biol.* **6,** 56–68 (2005).
10. Radtke, F. & Clevers, H. Self-renewal and cancer of the gut: two sides of a coin. *Science* **307,** 1904–1909 (2005).
11. Macdonald, T. T. & Monteleone, G. Immunity, inflammation, and allergy in the gut. *Science* **307,** 1920–1925 (2005).
12. Ogawa, M. & Sasakawa, C. Intracellular survival of *Shigella*. *Cell. Microbiol.* **8,** 177–184 (2006).
13. Sansonetti, P. J. War and peace at mucosal surfaces. *Nature Rev. Immunol.* **4,** 953–964 (2004).
14. Parsot, C. *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. *FEMS Microbiol. Lett.* **252,** 11–18 (2005).
15. Miura, M. *et al.* OspE2 of *Shigella sonnei* is required for the maintenance of cell architecture of bacterium-infected cells. *Infect. Immun.* **74,** 2587–2595 (2006).
16. Sakai, T. *et al.* Integrin-linked kinase (ILK) is required for polarizing the epiblast, cell adhesion, and controlling actin accumulation. *Genes Dev.* **17,** 926–940 (2003).
17. Lynch, D. K., Ellis, C. A., Edwards, P. A. & Hiles, I. D. Integrin-linked kinase regulates phosphorylation of serine 473 of protein kinase B by an indirect mechanism. *Oncogene* **18,** 8024–8032 (1999).
18. Boulter, E., Grall, D., Cagnol, S. & Van Obberghen-Schilling, E. Regulation of cell-matrix adhesion dynamics and Rac-1 by integrin linked kinase. *FASEB J.* **20,** 1489–1491 (2006).
19. Wu, C. & Dedhar, S. Integrin-linked kinase (ILK) and its interactors: a new paradigm for the coupling of extracellular matrix to actin cytoskeleton and signaling complexes. *J. Cell Biol.* **155,** 505–510 (2001).
20. Hannigan, G., Troussard, A. A. & Dedhar, S. Integrin-linked kinase: a cancer therapeutic target unique among its ILK. *Nature Rev. Cancer* **5,** 51–63 (2005).
21. Legate, K. R., Montanez, E., Kudlacek, O. & Fassler, R. ILK, PINCH and parvin: the tIPP of integrin signalling. *Nature Rev. Mol. Cell Biol.* **7,** 20–31 (2006).
22. Lorenz, K. *et al.* Integrin-linked kinase is required for epidermal and hair follicle morphogenesis. *J. Cell Biol.* **177,** 501–513 (2007).
23. Bendig, G. *et al.* Integrin-linked kinase, a novel component of the cardiac mechanical stretch sensor, controls contractility in the zebrafish heart. *Genes Dev.* **20,** 2361–2372 (2006).
24. Ezratty, E. J., Partridge, M. A. & Gundersen, G. G. Microtubule-induced focal adhesion disassembly is mediated by dynamin and focal adhesion kinase. *Nature Cell Biol.* **7,** 581–590 (2005).
25. van der Flier, A. & Sonnenberg, A. Function and interactions of integrins. *Cell Tissue Res.* **305,** 285–298 (2001).
26. De Arcangelis, A. & Georges-Labouesse, E. Integrin and ECM functions: roles in vertebrate development. *Trends Genet.* **16,** 389–395 (2000).
27. Kim, M. *et al.* A new ubiquitin ligase involved in p57KIP2 proteolysis regulates osteoblast cell differentiation. *EMBO Rep.* **9,** 878–884 (2008).
28. Fassler, R. *et al.* Lack of beta 1 integrin gene in embryonic stem cells affects morphology, adhesion, and migration but not integration into the inner cell mass of blastocysts. *J. Cell Biol.* **128,** 979–988 (1995).

29. Shim, D. H. *et al.* New animal model of shigellosis in the Guinea pig: its usefulness for protective efficacy studies. *J. Immunol.* **178,** 2476–2482 (2007).
30. Roberts, M. *et al.* PDGF-regulated rab4-dependent recycling of αvβ3 integrin from early endosomes is necessary for cell adhesion and spreading. *Curr. Biol.* **11,** 1392–1402 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Ohmi, H. Fukuda and C. Takamura for MALDI-TOF analysis. We thank S. Yamaji and Y. Ishigatsubo for discussion. We thank the members of the Sasakawa laboratory, especially H. Mimuro, M. Suzuki and H. Ashida, for their advice. We are grateful to R. Whittier and T. Tezuka for critical reading of the manuscript. We thank H. Erickson for fibronectin fragment expression vector. This work was supported by the Deutsche Forschungsgemeinschaft (SFB576), the Max Planck Society, a Grant-in-Aid for the Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and the Special Coordination Funds for Promoting Science from Japan Science and Technology Agency (JSTA).

Author Contributions M.K., M.O. and Y.F. designed and performed the experiments. T.N. and Y.Y. assisted the experiments. R.F. and A.L. gave advice regarding the design of the experiments and provided ILK materials. T.K. and S.N. made antibodies. C.S. and R.F. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.S. (sasakawa@ims.u-tokyo.ac.jp).

METHODS

Antibodies, materials, immunoprecipitation and immunostaining. A polyclonal rabbit anti-OspE antibody was generated against recombinant maltose-binding protein–OspE (1–88 amino acids). Anti-*Shigella* lipopolysaccharide (LPS) were prepared as described previously[3]. Anti-ILK (Upstate Biotechnology, BD Bioscience and Cell Signaling), anti-FAK (Upstate Biotechnology), anti-phospho-FAK (Y397) (Biosource), anti-paxillin (Transduction Laboratories), anti-transferrin receptor (Zymed), anti-phospho-tyrosine (4G10) (Upstate Biotechnology), anti-Flag (Sigma), anti-GFP (MBL), anti-Myc (A14) (SantaCruz), anti-Myc (9B11) (Cell Signaling), anti-tubulin (DM1A) (Sigma), anti-vinculin (Sigma), anti-talin (Sigma), anti-phospho-Akt (Ser 473) (Cell Signaling), anti-Akt (Cell Signaling), anti-GSK3β (BD Bioscience), and anti-phospho-GSK3β (Ser 9) (Cell Signaling) antibodies were used as primary antibodies for immunostaining, immunoblotting and immunoprecipitation. Rhodamine-phalloidin (Sigma) and Alexa-633-phalloidin (Invitrogen) were used for F-actin staining. Anti-β1 integrin, anti-α-parvin and anti-β-parvin antibodies were used as described previously[31,32].

Bacterial strains and infections. The WT *Shigella flexneri* 2a YSH6000 strain has been described previously[33]. To construct a non-polar *ospE* deletion mutant, *ospE1* and its isogenic *ospE2* genes were disrupted using the λ red recombinase-mediated recombination system (called Δ*ospE*)[34]. *Shigella* Δ*ospE*/*ospE* and Δ*ospE*/*ospE*(W68A) mutants were created as described previously[3]. To generate *Shigella* expressing OspE–Flag, the original *ospE* genes were replaced with the *ospE–Flag* fusion gene by λ red recombinase-mediated recombination system, allowing the OspE–Flag protein to be produced under same context and secreted by means of the type III secretion system. *Shigella* infection and the plaque-forming assay were performed as described previously[35].

FACS analysis. Cell surface expression of β1 integrin was quantified by FACS analysis as described previously[36]. To determine total integrin levels, cells were permeabilized in 10 mM HEPES (pH 7.6), 150 mM glutamate, 5 mM glucose, 2 mM $MgCl_2$, 0.4 mM EGTA and 0.003% digitonin. For mouse fibroblasts, total and cell surface β1 integrin levels were analysed with an anti-β1 integrin antibody (Ha2/3) (BD Pharmingen). The active state of β1 integrin was analysed with an anti-β1 integrin antibody (9EG7). For HeLa S3 cells, an anti-CD29 antibody was used to detect β1 integrin. The active state integrin was detected using FITC-labelled fibronectin (FN)-III fragments. FN-$III_{7–10}$ RGD and FN-$III_{7–10}$ RGE were expressed and purified as described previously[37]. The levels of active state integrin were calculated as follows: the mean fluorescence of FN-$III_{7–10}$ RGD minus the background obtained with the mean fluorescence of FN-$III_{7–10}$ RGE.

Cell fractionation. Cells were suspended in buffer A (50 mM Tris-HCl, pH 7.6, 0.2% saponin, protease inhibitor mixture (Roche)). After centrifuging at 17,400*g* for 30 min, the supernatant was collected as the cytosolic fraction. The pellet was lysed in buffer B (50 mM Tris-HCl, pH 7.6, 1% Triton, protease inhibitor mixture). After 15 min incubation on ice, the lysate was centrifuged at 17,400*g* for 30 min. The resulting supernatant was used as the membrane fraction.

GST pull-down assay and mass spectrometric analysis. HeLa cells were lysed in lysis buffer (1% Triton X-100 in 50 mM Tris-HCl, pH 7.4, 150 mM NaCl and protease inhibitors), and the cleared lysates were mixed with GST–OspE or GST-bound glutathione sepharose 4B at 4 °C overnight (12 h). Bound proteins were separated by SDS–PAGE and stained with Coomassie blue. The protein band at approximately 50 kDa was excised from the gel. The gel slice was subjected to in-gel digestion with trypsin. The resulting tryptic peptides were injected into a nano-LC system (DiNa, KYA). The eluent was mixed with the matrix solution and the mixture was directly blotted onto a MALDI sample plate with a micro-fractionation system (AccuSpot, Shimadzu) for MALDI-TOF tandem mass spectrometry analysis. Overall peptide identification was performed using a MALDI-TOF/TOF 4700 proteomics analyser (Applied Biosystems), followed by a database search with Mascot version 2.0 (Matrix Science).

***In vitro* ILK kinase assay.** Cells were lysed in 50 mM Tris-HCl (pH 7.4), 1% NP-40, 50 mM NaCl, 5 mM EDTA, 0.1 mM Na_3VO_4, 5 mM β-glycerophosphate, 2.5 mM NaF and a protease inhibitor cocktail, and then immunoprecipitated with an anti-ILK antibody. The immunoprecipitates were washed five times with lysis buffer and then washed twice with kinase reaction buffer (20 mM HEPES, pH 7.0, 10 mM $MgCl_2$, 3 mM $MnCl_2$, 0.1 mM Na_3VO_4, 5 mM β-glycerophosphate, 2.5 mM NaF). The kinase assay was performed with immuno-purified ILK and purified substrates in 40 μl kinase reaction buffer containing 10 μCi of [γ-^{32}P]-ATP for 30 min at 30 °C. The phosphorylated substrates were resolved by SDS–PAGE followed by autoradiography.

Protein expression in mammalian cells. The immunoprecipitation, immunoblotting and immunostaining analyses were performed as described previously[3]. To construct $Ilk^{-/-}$ cells and NIH3T3 cells stably expressing WT ILK, ILK(K220M), ILK(S343D) and OspE2, cDNAs encoding these genes were subcloned into the pMX-neo or pMX-puro retroviral expression vectors. To generate HeLa S3 stable clones, HeLa S3 cells were transfected with pMX-puro, pMX-puro-OspE(WT), and pMX-puro-OspE(W68A), respectively, and then cloned under puromycin selection. The cDNAs for WT, the N-terminal region (amino acids 1–196) and C-terminal region (amino acids 197–452) with or without E359K mutation of ILK were cloned into the pcDNA-6Myc expression vector. A small interfering RNA (CTCAACGAGAATCACTCT) duplex specific for human ILK was transfected into HeLa S3 cells using Oligofectamine (Invitrogen).

31. Aszodi, A., Hunziker, E. B., Brakebusch, C. & Fassler, R. Beta1 integrins regulate chondrocyte rotation, G1 progression, and cytokinesis. *Genes Dev.* **17,** 2465–2479 (2003).
32. Chu, H. *et al.* gamma-Parvin is dispensable for hematopoiesis, leukocyte trafficking, and T-cell-dependent antibody response. *Mol. Cell. Biol.* **26,** 1817–1825 (2006).
33. Sasakawa, C. *et al.* Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and Congo red binding activity in *Shigella flexneri*. *Infect. Immun.* **51,** 470–475 (1986).
34. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97,** 6640–6645 (2000).
35. Ogawa, M. *et al.* IcsB, secreted via the type III secretion system, is chaperoned by IpgA and required at the post-invasion stage of Shigella pathogenicity. *Mol. Microbiol.* **48,** 913–931 (2003).
36. Cowden Dahl, K. D., Robertson, S. E., Weaver, V. M. & Simon, M. C. Hypoxia-inducible factor regulates alphavbeta3 integrin cell surface expression. *Mol. Biol. Cell* **16,** 1901–1912 (2005).
37. Ohashi, T. & Erickson, H. P. Domain unfolding plays a role in superfibronectin formation. *J. Biol. Chem.* **280,** 39143–39151 (2005).

LETTERS

A regulated auxin minimum is required for seed dispersal in *Arabidopsis*

Karim Sorefan[1], Thomas Girin[1], Sarah J. Liljegren[2,3], Karin Ljung[4], Pedro Robles[2,5], Carlos S. Galván-Ampudia[6], Remko Offringa[6], Jiří Friml[7], Martin F. Yanofsky[2] & Lars Østergaard[1]

Local hormone maxima are essential for the development of multicellular structures and organs. For example, steroid hormones accumulate in specific cell types of the animal fetus to induce sexual differentiation[1] and concentration peaks of the plant hormone auxin direct organ initiation and mediate tissue patterning[2–4]. Here we provide an example of a regulated local hormone minimum required during organogenesis. Our results demonstrate that formation of a local auxin minimum is necessary for specification of the valve margin separation layer where *Arabidopsis* fruit opening takes place. Consequently, ectopic production of auxin, specifically in valve margin cells, leads to a complete loss of proper cell fate determination. The valve margin identity factor INDEHISCENT (IND) is responsible for forming the auxin minimum by coordinating auxin efflux in separation-layer cells. We propose that the simplicity of formation and maintenance make local hormone minima particularly well suited to specify a small number of cells such as the stripes at the valve margins.

Control of seed dispersal was crucial for crop domestication during the Neolithic revolution[5] and remains a major problem for young crops of the Brassicaceae family such as oilseed rape[6]. Understanding the mechanism of fruit development in the closely related model species *Arabidopsis thaliana* will probably provide the insight necessary to solve this issue.

The major exterior parts of an *Arabidopsis* fruit include the valves or carpel walls, the replum, a thin structure extending along the length of the fruit, and the valve margins, which form at the valve/replum border (Supplementary Fig. 1a). The valve margins differentiate into narrow stripes of cells consisting of a separation layer and a layer of lignified cells[7]. Immediately before fruit opening, cells in the separation layer secrete cell-wall-degrading enzymes to mediate cell separation in a process known as dehiscence[8].

Auxin is involved in diverse aspects of plant growth and development, including the patterning of Brassicaceae fruit[9–12]. The *DR5rev::GFP* reporter[3] has been used to visualize auxin responses in various *Arabidopsis* tissues. We used this tool to characterize tissue-specific auxin-response levels throughout *Arabidopsis* fruit development (Fig. 1a and Supplementary Fig. 1b). A dynamic tissue-specific green fluorescent protein (GFP) expression pattern was observed (see Supplementary Information and Supplementary Fig. 1b) suggesting that distribution of auxin is under careful regulation, possibly to ensure proper postfertilization growth, and agrees with direct measurements of the naturally occurring auxin, indole-3-acetic acid (IAA; Supplementary Fig. 1c). The separation layer of the valve margins had very low GFP signals especially at later stages (Fig. 1a, developmental stages defined in refs 7 and 13), indicating that these cells have reduced auxin response and/or levels. In the following text, we will refer to this as an auxin minimum.

Valve margin formation requires the activity of IND, and *ind* mutant fruit fail to open on maturation[14]. This defect is known as indehiscence. The *IND* gene encodes a putative basic helix–loop–helix transcription factor and is expressed in stripes where the valve margins will subsequently differentiate (Supplementary Fig. 2). *IND* expression is inhibited in the valves by FRUITFULL (FUL)[15] and in the replum by REPLUMLESS (RPL)[16]; however, little is known about the downstream events that lead to valve margin formation.

In contrast to wild type, no auxin minimum in the separation layer of *ind-2* mutant fruit was observed using the *DR5rev::GFP* reporter (Fig. 1a, b). Conversely, in *rpl* mutants with ectopic *IND* expression in the reduced replum[16], *GFP* signal was greatly decreased, showing that the auxin minimum had expanded to include the replum (Fig. 1c). These results indicate that IND is involved in creating the auxin minimum observed at the separation layer of wild-type fruit (Fig. 1a).

To test whether this spatially restricted auxin minimum is required for valve margin specification, the bacterial auxin biosynthesis gene *iaaM*[17] was used to increase auxin concentration specifically in valve margin cells. For this purpose, we modified the valve-margin-specific *IND::IND:GUS* construct (Supplementary Fig. 2) and created transgenic *IND::IND:iaaM* lines. In these lines, the auxin minimum no longer formed (Fig. 1d). We furthermore found that *IND::IND:iaaM* transgenic fruit (from 22 independent lines out of 23) were indehiscent, but without reduced *IND* expression and with increased auxin levels (Supplementary Fig. 3). Scanning electron microscopy (SEM), lignin staining and marker line analysis revealed that these fruit lacked valve margin differentiation (Fig. 1e–l), similar to what has previously been described for *ind* mutants[14]. The results obtained using the *iaaM* gene demonstrate that formation of the auxin minimum is necessary for separation layer development and indicate that the indehiscent phenotype of the *ind* mutant is at least partially due to a failure to direct a local auxin minimum at the valve margins.

To gain further insight into the role of IND in valve margin differentiation, the *IND* gene was expressed under control of the constitutive CaMV35S promoter (*35S::IND*)[14] or ectopically induced with dexamethasone (DEX)[18] using a translational fusion of IND to the glucocorticoid receptor (*35S::IND:GR*). The developmental defects of ectopic IND activity are similar to those obtained when the *IND*-related *HECATE* genes are over expressed[19], and are reminiscent of mutants with defects in polar auxin transport such as *pin-formed* (*pin1*) and *pinoid* (*pid*) (Fig. 2a–d)[20,21]. The *PIN1* gene encodes a member of the PIN family of auxin efflux carriers[22]. Polar, subcellular localization of PIN proteins determines the direction

[1]Crop Genetics Department, John Innes Centre, Norwich Research Park, Colney Lane, Norwich, Norfolk NR4 7UH, UK. [2]Section of Cell and Developmental Biology, University of California at San Diego, La Jolla, California 92093-0116, USA. [3]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [4]Department of Forest Genetics and Plant Physiology, Umeå Plant Science Centre, S-901 83 Umeå, Sweden. [5]División de Genética e Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202, Alicante, Spain. [6]Department of Molecular and Developmental Genetics, Institute of Biology, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands. [7]Department of Plant Systems Biology, VIB, and Department of Plant Biotechnology and Genetics, Ghent University, 9052 Gent, Belgium.

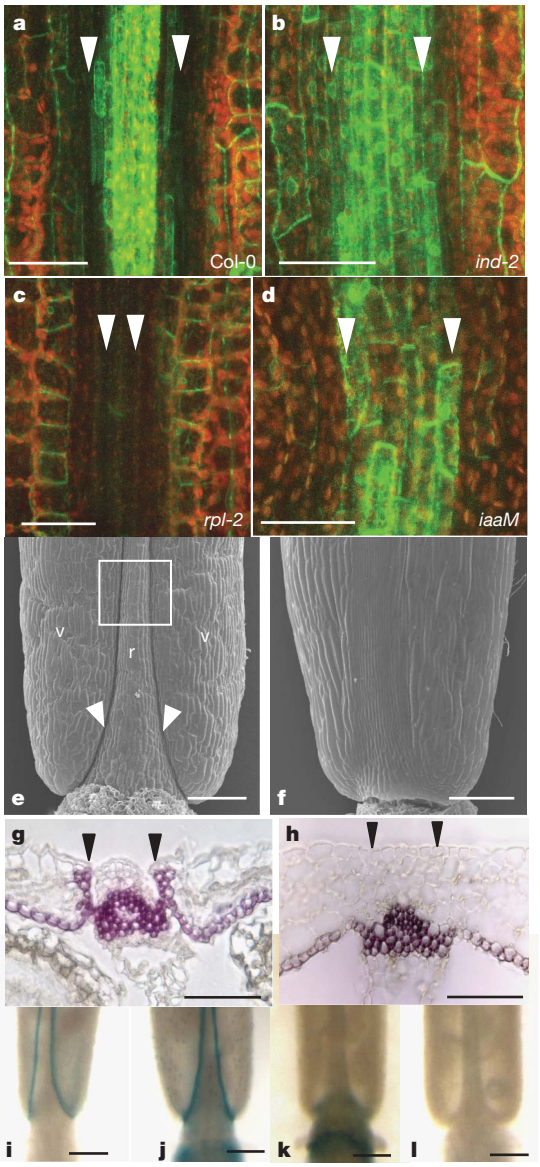


Figure 1 | An auxin minimum forms at the valve margin of *Arabidopsis* fruit. **a–d**, Confocal images of *DR5::GFP* taken at the valve/replum region (white frame in **e**) at stage 17b of Col-0 (**a**), *ind-2* (**b**), *rpl-2* (**c**) and *IND::IND:iaaM* (**d**). GFP signals are in green and plastids are in red. **e**, **f**, SEM images of the basal part of stage-17b fruit from wild-type (**e**) and *IND::IND:iaaM* (**f**) plants. **g**, **h**, Phloroglucinol-stained cross sections of stage-17b wild-type (**g**) and *IND::IND:iaaM* (**h**) fruit. Arrowheads indicate valve margin positions. **i–l**, Expression of valve margin markers in stage-15 wild-type (**i**, **j**) and *IND::IND:iaaM* (**k–l**) fruit. Scale bars: 250 μm (**i–l**), 100 μm (**e**, **f**), 50 μm (**a–d**, **g**, **h**).

of auxin flow[23] and they have been shown to mediate auxin maxima in multiple developmental processes[24]. The *PID* gene is necessary for proper auxin distribution[25] by ensuring polar localization of PIN proteins at the plasma membrane[26]. The observed phenotypes therefore indicate that IND regulates auxin transport, which is highly consistent with the requirement for IND in generating the auxin minimum at the valve margin (Fig. 1b).

To identify components of the auxin transport machinery that IND interacts with, *PIN* reporter lines were analysed for fruit expression and localization. PIN3 is apolarly localized in root columella cells and diverts the auxin flux downwards in response to gravity[27]. Here we find that *PIN3* is also expressed during *Arabidopsis* fruit development. A transcriptional *PIN3::GUS* construct was expressed ubiquitously in mature fruit (Supplementary Fig. 4), whereas a translational *PIN3::PIN3:GFP* fusion only produced signal in the valves and replum

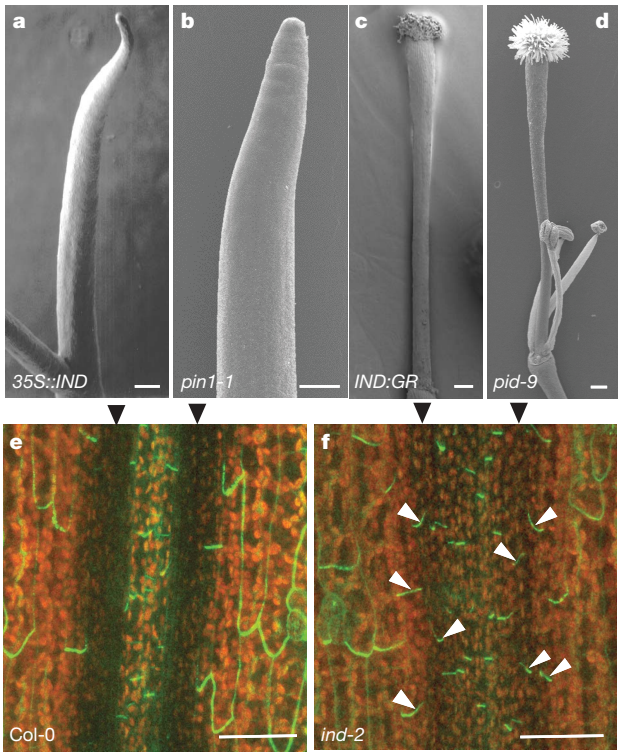


Figure 2 | IND regulates auxin transport. **a–d**, SEM images of *35S::IND* pin-like inflorescence (**a**), a *pin1-1* mutant (**b**), stage-15 fruit from DEX-treated *35S::IND:GR* (**c**) and a *pid-9* fruit at a similar stage (**d**). **e**, **f**, IND prevents polar PIN3 localization at the valve margin. *PIN3::PIN3:GFP* in Col-0 (**e**) and *ind-2* (**f**) stage-17 fruit. PIN3:GFP is basally localized in replum and valve cells, but not observed in the valve margin cells of Col-0 fruit. In *ind-2* fruit, PIN3:GFP is basally localized in valve margin cells (white arrowheads). For confocal images, $n > 10$. Scale bars: 200 μm (**c**, **d**), 100 μm (**a**, **b**), 50 μm (**e**, **f**).

where it is predominantly basally localized (Fig. 2e). At the separation layer, PIN3:GFP is absent in wild-type fruit; however, in *ind-2* mutant fruit, PIN3:GFP becomes basally localized in all cell files spanning across from the valves to the replum (Fig. 2f). These data imply that the *PIN3* gene is post-transcriptionally regulated at the separation layer, and that IND is required to prevent polar PIN3 localization in this region.

The *PIN3::PIN3:GFP* line was crossed to *35S::IND:GR* plants to study the effect of DEX-induced IND activity on PIN3 localization. In the absence of DEX, we observed a strict polar localization of PIN3:GFP in the root endodermis (Fig. 3a, b), whereas 48 h of DEX treatment caused delocalization of PIN3:GFP to all sides of the cells (Fig. 3c, d). To check whether the effect of IND on localization is specific to PIN3, *35S::IND:GR* was crossed to a *PIN1::PIN1:GFP* line[4]. PIN1:GFP is apically localized in valve epidermal cells of stage-10 gynoecia (Fig. 3e); however, 16 h of DEX treatment resulted in a marked PIN1:GFP apolarization (Fig. 3f) and occasional anticlinal cell divisions (see Supplementary Information and Supplementary Fig. 5). These results show that IND controls PIN localization and thereby the direction of auxin transport.

After 24 h of DEX treatment, the PIN1:GFP signal had almost disappeared, indicating that PIN1:GFP proteins in the apolar state are unstable and may be relatively short lived (Fig. 3g). This could explain the absence of PIN3:GFP signal at the separation layer of wild-type fruit despite *PIN3* expression (Supplementary Fig. 4).

As a putative basic helix–loop–helix transcription factor, IND is unlikely to directly regulate PIN polarization. Localization of PIN proteins is partially governed by their phosphorylation status, which depends on the antagonistic action of the PP2AA1/2/3 phosphatases[28], the PID kinase[26] and probably the related kinases WAG1, WAG2 and AGC3–4 (PID2) (refs 29 and 30). To test whether IND transcriptionally regulates the expression of any of these factors, we performed quantitative messenger RNA abundance assays (quantitative polymerase

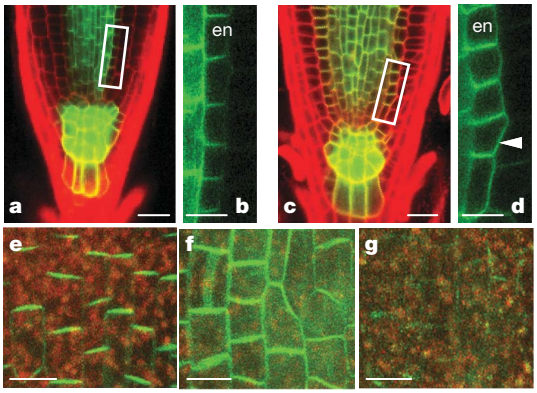


Figure 3 | IND mediates PIN delocalization. Polar PIN3:GFP localization in the root endodermis (en) of two-day-old seedlings (**a**, **b**) is shifted to apolar/lateral localization (**c**, **d**) in *35S::IND:GR* treated with DEX (indicated with arrowheads). **e**, Apical PIN1:GFP localization in valve cells of a stage-10 uninduced *35S::IND:GR* gynoecium. **f**, **g**, Delocalization of PIN1:GFP in a 16-h DEX-treated *35S::IND:GR* gynoecium (**f**). The PIN1:GFP signal significantly decreases after 24 h of DEX treatment (**g**). Control DEX treatments of the *PIN3::PIN3:GFP* and *PIN1::PIN1:GFP* lines did not change PIN polarity (data not shown). For confocal images, $n > 10$. Scale bars: 25 μm (**a**, **c**), 10 μm (**b**, **d**–**g**).

chain reaction, qPCR) using tissue from *35S::IND:GR* seedlings treated with DEX (Fig. 4a). Whereas *PP2AA* expression did not significantly change (data not shown), *PID* expression was repressed by DEX treatment and the induction by auxin was blocked in the presence of DEX. Moreover, *WAG2* expression was unaffected by DEX alone, whereas a modest induction by auxin was notably enhanced in the presence of DEX. These data show that IND can inversely regulate the expression of *PID* and *WAG2* (Fig. 4a). In contrast, IND did not regulate *PIN1* expression, and *PIN3* was only mildly induced (Fig. 4a). Both PID and WAG2 are key regulators of PIN polar localization (ref. 28 and C.S.G.-A. *et al.*, manuscript in preparation) and thus the IND-dependent regulation of these kinases provides a mechanism for how IND modulates PIN localization during margin formation.

To confirm the relevance of the results obtained in seedlings, endogenous *PID* and *WAG2* expression in stage-15 fruit from wild type (Col-0), *ind-2* and *ful-2* were compared (Fig. 4b). *WAG2* expression was significantly decreased in *ind-2* compared to wild type ($P < 0.05$), whereas no detectable difference in *PID* expression was observed. The lack of change in *PID* expression may be due to minor increases that are masked by using the whole fruit. In contrast, in *ful-2* fruit with an expanded *IND* expression domain[15], *PID* expression is markedly reduced and *WAG2* is induced (Fig. 4b). These data support the suggestion that *IND* is involved in the regulation of PIN-dependent auxin transport by inverse regulation of *PID* and *WAG2* transcription.

To investigate whether IND is a direct regulator of *PID* and *WAG2*, we performed chromatin immunoprecipitation (ChIP) assays using a GR antibody followed by qPCR and found significant enrichments of *PID* and *WAG2* fragments in *35S::IND:GR* plants induced by DEX ($P < 0.05$, Fig. 4c). An induction experiment with DEX and the translational inhibitor cycloheximide showed similar results (Supplementary Fig. 6). We have therefore identified *PID* and *WAG2* as two targets of a potential set of intermediate factors directly regulated by IND to control auxin transport during formation of the separation layer. These results also show that IND interprets cellular auxin levels to modify the direction of auxin flow by acting both as a transcriptional repressor and as an inducer. Additional genetic analysis will be necessary to investigate the specific contributions of *PID*, *WAG2* and potentially other factors to separation layer development.

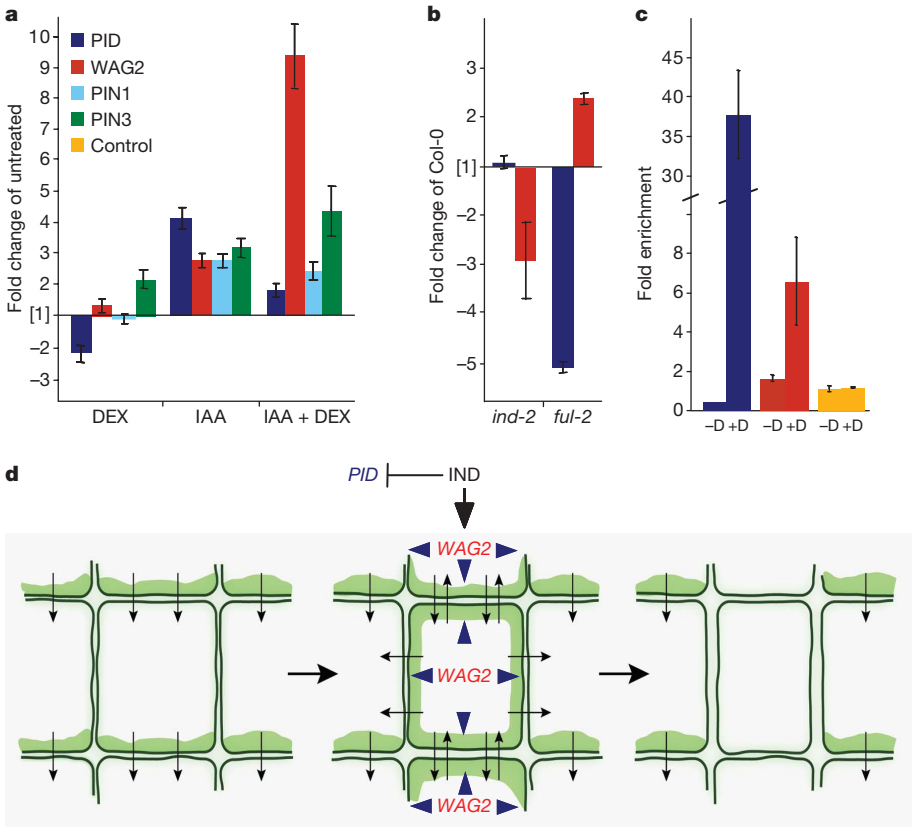


Figure 4 | IND directly regulates *PID* and *WAG2* expression. **a**, **b**, Quantitative RT–PCR in 7-day-old *35S::IND:GR* seedlings ±DEX and ±IAA ($n = 3$) in wild-type (**a**) versus *ind-2* and *ful-2* ($n = 4$) (**b**) stage-15 fruit. **c**, ChIP experiment against the *PID* and *WAG2* genes. –D and +D indicate absence or presence of DEX, respectively. Standard errors ($n = 3$) of biological repeats are indicated. The *NRT2.1* gene (At1g08090) serves as a negative control not affected by DEX treatment. Colour coding for genes in **a**–**c** is as shown in **a**. **d**, Model for IND-mediated auxin minimum formation at the separation layer. PIN localization is shown in green and auxin flow is indicated by arrows from cell to cell.

IND expression is initiated in the presumptive valve margin region of stage-11 flowers just before fertilization (Supplementary Fig. 2). IND is therefore likely to have important functions in setting up fruit patterning from an early stage. Here we have identified a specific role for IND in separation layer formation, which occurs late during fruit development to facilitate seed dispersal. We have summarized our results in a model suggesting that IND is responsible for forming an auxin minimum by regulating polar localization of PIN auxin transporters (Fig. 4d). Our results indicate that IND coordinates auxin transport to specify the separation layer at least partially by repressing *PID* and inducing *WAG2* expression at the valve margins, subsequently leading to PIN relocation from apico-basal to apolar-lateral and resulting in a local auxin minimum.

Great progress has been achieved in recent years on the importance of auxin maxima and gradients in various aspects of plant development[2–4]. Our data show that highly regulated local depletion of auxin can have an important role in determining cell fate as well. We propose that the creation of such an auxin minimum is particularly well suited to ensure proper specification of a small number of cells such as the stripes at the valve margin, because coordination of auxin transport to create and maintain an auxin minimum may be a simpler process than for an auxin maximum. To our knowledge, a regulated minimum of a signalling molecule has not previously been reported in plants or animals, and thus provides a new concept in general hormone biology.

METHODS SUMMARY

Confocal microscopy. Confocal microscopy was performed using a Leica SP laser scanning microscope equipped with an Argon krypton laser (Leica Microsystems). The 488-nm excitation line of an argon ion laser was used to excite GFP. GFP emission spectra were collected between 505 and 584 nm and plastid autofluorescence was collected between 601 and 790 nm. Epidermal surface images were captured using laser reflection between 480 and 499 nm. Floral buds were dissected and mounted in water and ~0.05% Tween-20 ($n > 10$). 1–2-μm *Z* sections were imaged and converted to maximum projections. Images were averaged 2–3 times and were processed using the Leica CONFOCAL software, ImageJ and Adobe Photoshop software (version 7.01; Adobe Systems).

Scanning electron microscopy. Fruit were fixed for approximately 4 h at 25 °C in FAA (50% ethanol, 5% glacial acetic acid and 3.7% formaldehyde). After critical point drying, tissue was coated with gold/palladium and examined in a Quanta 600 microscope using an acceleration voltage of 20 kV.

Quantitative PCR with reverse transcription. Total RNA was extracted using RNeasy Plant mini kit (Qiagen, 74904). Complementary DNA was produced using 2–5 μg total RNA and a polyT(15) primer. Three biological and two technical repeats were performed for the experiments in Fig. 4a and Supplementary Fig. 6a, c. Because the experiments in Fig. 4b, Supplementary Figs 3a and 6b were carried out on cDNA derived from RNA of pooled tissue, three (Fig. 4b and Supplementary Fig. 6b) and two (Supplementary Fig. 3a) technical repeats were performed.

Quantitative PCR was performed with SYBR green jumpstart Taq readymix (Sigma, S4438) on a Chromo4 real-time PCR detector. PCR products were checked by agarose gel electrophoresis. *UBQ10* was used as a normalization control because its expression is not affected by auxin treatment. Student's *t*-test was performed and significant results had P values < 0.05.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 November 2008; accepted 11 February 2009.

1. Wu, Z., Wan, S. & Lee, M. M. Key factors in the regulation of fetal and postnatal leydig cell development. *J. Cell. Physiol.* **213,** 429–433 (2007).
2. Sabatini, S. *et al.* An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* **99,** 463–472 (1999).
3. Friml, J. *et al.* Efflux-dependent auxin gradients establish the apical–basal axis of *Arabidopsis*. *Nature* **426,** 147–153 (2003).
4. Benková, E. *et al.* Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* **115,** 591–602 (2003).
5. Konishi, S. *et al.* An SNP caused loss of shattering during rice domestication. *Science* **312,** 1392–1396 (2006).
6. Østergaard, L., Kempin, S. A., Bies, D., Klee, H. J. & Yanofsky, M. F. Pod shatter-resistant *Brassica* fruit produced by ectopic expression of the *FRUITFULL* gene. *Plant Biotechnol. J.* **4,** 45–51 (2006).
7. Roeder, A. H. K. & Yanofsky, M. F. in *The Arabidopsis Book* (eds Somerville, C. R. & Meyerowitz, E. M.) doi:10.1199/tab.0075 (Am. Soc. Plant Biol., 2006).
8. Petersen, M. *et al.* Isolation and characterisation of a pod dehiscence zone-specific polygalacturonase from *Brassica napus*. *Plant Mol. Biol.* **31,** 517–527 (1996).
9. Nemhauser, J., Feldmann, L. J. & Zambryski, P. C. Auxin and ETTIN in *Arabidopsis* gynoecium morphogenesis. *Development* **127,** 3877–3888 (2000).
10. Østergaard, L. Don't 'leaf' now. The making of a fruit. *Curr. Opin. Plant Biol.* **12,** 36–41 (2009).
11. Chauvaux, N. *et al.* The role of auxin in cell separation in the dehiscence zone of rapeseed pods. *J. Exp. Biol.* **48,** 1423–1429 (1997).
12. Sohlberg, J. J. *et al. STY1* regulates auxin homeostasis and affects apical–basal patterning of the *Arabidopsis* gynoecium. *Plant J.* **47,** 112–123 (2006).
13. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in *Arabidopsis*. *Plant Cell* **2,** 755–767 (1990).
14. Liljegren, S. J. *et al.* Control of fruit patterning in *Arabidopsis* by INDEHISCENT. *Cell* **116,** 843–853 (2004).
15. Ferrándiz, C., Liljegren, S. J. & Yanofsky, M. F. Negative regulation of the *SHATTERPROOF* genes by FRUITFULL during *Arabidopsis* fruit development. *Science* **289,** 436–438 (2000).
16. Roeder, A. H., Ferrándiz, C. & Yanofsky, M. F. The role of the REPLUMLESS homeodomain protein in patterning the *Arabidopsis* fruit. *Curr. Biol.* **13,** 1630–1635 (2003).
17. Romano, C. P., Robson, P. R., Smith, H., Estelle, M. & Klee, H. Transgene-mediated auxin overproduction in *Arabidopsis*: hypocotyl elongation phenotype and interactions with the *hy6–1* hypocotyl elongation and *axr1* auxin-resistant mutants. *Plant Mol. Biol.* **27,** 1071–1083 (1995).
18. Aoyama, T. & Chua, N.-H. A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.* **11,** 605–612 (1997).
19. Gremski, K., Ditta, G. & Yanofsky, M. F. The *HECATE* genes regulate female reproductive tract development in *Arabidopsis thaliana*. *Development* **134,** 3593–3601 (2007).
20. Okada, K., Ueda, J., Komaki, M. K., Bell, C. J. & Shimura, Y. Requirement of the auxin polar transport system in early stages of *Arabidopsis* floral bud formation. *Plant Cell* **3,** 677–684 (1991).
21. Bennett, S. M. R., Alvarez, J., Bossinger, G. & Smyth, D. R. Morphogenesis in *pinoid* mutants of *Arabidopsis thaliana*. *Plant J.* **8,** 505–520 (1995).
22. Petrásek, J. *et al.* PIN proteins perform a rate-limiting function in cellular auxin efflux. *Science* **312,** 914–918 (2006).
23. Wisniewska, J. *et al.* PIN localization directs auxin flow in plants. *Science* **312,** 883 (2006).
24. Tanaka, H., Dhonukshe, P., Brewer, P. B. & Friml, J. Spatiotemporal asymmetric auxin distribution: a means to coordinate plant development. *Cell. Mol. Life Sci.* **63,** 2738–2754 (2006).
25. Benjamins, R., Quint, A., Weijers, D., Hooykaas, P. & Offringa, R. The PINOID protein kinase regulates organ development in *Arabidopsis* by enhancing polar auxin transport. *Development* **128,** 4057–4067 (2001).
26. Friml, J. *et al.* A PINOID-dependent binary switch in apical–basal PIN polar targeting directs auin efflux. *Science* **306,** 862–865 (2004).
27. Friml, J., Wisniewska, J., Benková, E., Mendgen, K. & Palme, K. Lateral relocation of auxin efflux regulator PIN3 mediates tropism in *Arabidopsis*. *Nature* **415,** 806–809 (2002).
28. Michniewicz, M. *et al.* Antagonistic regulation of PIN phosphorylation by PP2A and PINOID directs auxin flow. *Cell* **130,** 1044–1056 (2007).
29. Santner, A. A. & Watson, J. C. The WAG1 and WAG2 protein kinases negatively regulate root waving in *Arabidopsis*. *Plant J.* **45,** 752–764 (2006).
30. Galván-Ampudia, C. S. & Offringa, R. Plant evolution: AGC kinases tell the auxin tale. *Trends Plant Sci.* **12,** 541–547 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Acknowledgements We thank E. York and K. Findlay for assistance on SEM analysis, G. Calder for assistance with confocal microscopy, P. Pople for graphics assistance, and H. F. Klee for the pMON518 plasmid containing the *iaaM* gene. We also wish to thank G. S. Ditta, L. Dolan, S. Fuentes, J. Kleine-Vehn, R. Sablowski, P. Stephenson and T. Wood for carefully reading the manuscript and constructive criticism. P.R was the recipient of a postdoctoral fellowship from the Spanish government. This work was supported by grants from FWO (Odysseus program) to J.F., from The Netherlands Organisation for Scientific Research (ALW-NWO) to R.O., from the National Science Foundation to M.F.Y., and from the Biotechnological and Biological Sciences Research Council as well as core strategic funds from the John Innes Centre to L.Ø.

Author Contributions This project was conceived by K.S., M.F.Y. and L.Ø. Experiments were designed by K.S. and L.Ø. K.S. performed the confocal microscopy and expression analyses, T.G. carried out the chromatin immunoprecipitations, L.Ø. made the *IND::IND:iaaM* lines and performed plant sections and tissue staining, K.L. performed the IAA measurements, P.R. made the *35S::IND:GR* transgenic line, S.J.L. characterized the *35S::IND* lines and took the *35S::IND* SEM image shown, and C.S.G.-A., J.F. and R.O. analysed the effect of WAG2 activity on PIN localization. K.S. and L.Ø. analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.Ø. (lars.ostergaard@bbsrc.ac.uk).

METHODS

Constructs and generation of transgenic plants. Generation of transgenic plants harbouring the *35S::IND* construct was previously reported[5]. Out of 101 T1 lines, 8 were observed to have the *pin*-like phenotype.

To construct the *35S::IND:GR* plasmid, we first cloned the *IND* open reading frame into the pCR2.1 vector (Invitrogen) for sequencing and further into the BamHI site of pGREEN0229-*35S:GR*[31].

Primers INDa and INDb (primer list in Supplementary Information) were used to PCR amplify a 3.2-kb fragment of the *IND* gene containing 2.7 kb of promoter sequence and the 0.5-kb open reading frame. This fragment was cloned into the pCR2.1 vector for sequencing and further into the StuI and AgeI sites of the GFP-containing plant transformation vector pEGAD[32], giving rise to *IND::IND:GFP*. A fragment containing the β-glucoronidase (GUS) open reading frame was amplified. This fragment was cloned into the pCR2.1 for sequencing and further into the AgeI and EcoRI sites of the *IND::IND:GFP* plasmid exchanging the GFP for GUS. Expression of *IND::IND:GUS* at the valve margin was assessed (Supplementary Fig. 2). Out of 100 transgenic lines, 32 were found to cosuppress *IND* expression resulting in indehiscent fruit. This is similar to what was previously reported for an independent *IND::GUS* reporter construct[14].

A fragment containing the entire 2.3-kb open reading frame of the *iaaM* gene was amplified using the vector pMON518 (ref. 33) as a template. This fragment was cloned into the pCR2.1 vector for sequencing and further as an AgeI/EcoRI(blunted) fragment into the AgeI and BamHI (blunted) sites of *IND::IND:GFP*, replacing the GFP open reading frame and giving rise to *IND::IND:iaaM*. Out of 132 T1 lines, 109 showed severe auxin overproduction phenotypes (rosette leaf epinasty, increased apical dominance and senescing shoot apical meristems). These lines never produced fruit. For the remaining 23 lines, 22 were indehiscent. We also constructed a plasmid containing the *IND* promoter directly fused to the *iaaM* gene. Out of 52 T1 lines, all were severely affected by auxin overproduction and only 3 lines produced fruit. These were all indehiscent. On the basis of the higher level of ectopic effects in the *IND::iaaM* plants, we decided to continue our studies using *IND::IND:iaaM* lines with phenotypes restricted to the valve margin.

Generation of the *PIN3::PIN3:GFP* reporter line is described elsewhere (Z. Ding *et al.*, submitted).

Transgenic plants were selected by kanamycin or glufosinate resistance after *Agrobacterium*-mediated transformation into *Arabidopsis* (Col-0 ecotype).

IAA quantification. For each measurement, three independent samples were collected. Frozen samples (containing pooled tissue of 10–20 mg fresh weight) were homogenized in 0.5 ml 50 mM sodium-phosphate buffer, pH 7.0, containing 0.02% diethyldithiocarbamic acid (antioxidant, Sigma) and 2,000 pg [$^{13}C_6$]IAA (Cambridge Isotope Laboratories) internal standard, using the Retsch MM 301 vibration mill (Retsch GmbH & Co. KG) and a 3 mm tungsten carbide bead at a frequency of 30 Hz for 2 min. The samples were then incubated for 15 min at +4°C with continuous shaking. pH was adjusted to 2.7 with 1 M HCl, and the samples were purified by solid phase extraction on a 500 mg Isolute C8 (EC) column (International Sorbent Technology), conditioned with 2 ml methanol and 2 ml 1% acetic acid. After sample application, the column was first washed with 2 ml 10% methanol in 1% acetic acid and then eluted with 2 ml 70% methanol in 1% acetic acid and the sample was evaporated to dryness. The sample was dissolved in 0.2 ml 2-propanol and 1 ml dichloromethane, and 5 µl 2 M trimethylsilyl-diazomethane in hexane (Aldrich) was added to methylate the sample. After incubation at room temperature (22 °C) for 30 min, 5 µl 2 M acetic acid in hexane was added to destroy excess diazomethane and the sample was evaporated to dryness. The samples were then silylated by adding 15 µl acetonitril and 15 µl BSTFA plus 1% TMCS (Pierce). After 30 min incubation at 70 °C, the samples were evaporated to dryness, dissolved in heptane and quantified by GC-SRM-MS (gas chromatography-selected reaction monitoring-mass spectrometry) as previously described[34].

DEX induction. Flowers from *PIN1::PIN1:GFP 35S::IND:GR* double-transgenic line were sprayed with a solution of 10 µM DEX and 0.05% Silwet. A DMSO control was included. For expression assays, *35S::IND:GR* seeds (~20) were germinated in 5 ml of 0.5% glucose (w/v) 0.5× Mureshige and Skoog medium with constant shaking. After 7 days growth under constant light, seedlings were treated and tissue collected at the time points described before snap freezing in liquid nitrogen. Experiments were performed in seedlings because it was not possible to control the chemical dose in fruit. Dexamethasone (Sigma, D1756) and cycloheximide (Sigma, C4859) were dissolved in DMSO, and indole-3-acetic-acid (Sigma, I2886) was dissolved in 70% ethanol. The no-treatment controls had the equivalent concentrations of solvents.

Primer design. Primers for quantitative RT–PCR assays were designed using Primer3 (ref. 35) or AtRTPrimer[36] software. A list of the primers used is provided in Supplementary Information.

ChIP. *35S::IND-GR* seeds were grown for 7 days in 50 ml of 0.5% glucose (w/v) 0.5× Mureshige and Skoog medium with constant shaking. Seedlings were then treated with 50 µM IAA ± 10 µM DEX and ChIP was performed as described in ref. 37 with the following modifications. 2 µg of anti-GR antibody (AB3580, Abcam) was used for approximately 200 mg of tissue. Dynabeads-protein A (100-02D, Invitrogen Ltd) was used. Quantitative PCRs were done using primers Mu-likeF, Mu-likeR, NRT2.1F, NRT2.1R, PID-678F, PID-368R, WAG2F and WAG2R from the list in Supplementary Information. Mu-like retrotransposon was used as an internal control for normalization[38].

Lignin staining. Tissue was fixed and sectioned. Sections 8 µm thick were treated for 2 min with 2% phloroglucinol (in 95% ethanol) and photographed in 50% hydrochloric acid.

Reporter gene activity. Analysis of β-glucoronidase expression in YJ80, YJ161 and *PIN3::GUS* was done as previously described[39].

31. Yu, H., Ito, T., Wellmer, F. & Meyerowitz, E. M. Repression of *AGAMOUS-LIKE 24* is a crucial step in promoting flower development. *Nature Genet.* **36,** 157–161 (2004).
32. Cutler, S. R., Ehrhardt, D. W., Griffitts, J. S. & Somerville, C. R. Random GFP:cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proc. Natl Acad. Sci. USA* **97,** 3718–3723 (2000).
33. Romano, C. P., Hein, M. B. & Klee, H. J. Inactivation of auxin in tobacco transformed with the indoleacetic acid-lysine synthetase gene of *Pseudomonas savastanoi. Genes Dev.* **5,** 438–446 (1991).
34. Edlund, A., Eklöf, S., Sundberg, B., Moritz, T. & Sandberg, G. A microscale technoque for gas chromatography-mass spectrometry measurements of picogram amounts of indole-3-acetic acid in plant tissues. *Plant Physiol.* **108,** 1043–1047 (1995).
35. Rozen, S. & Skaletsky, H. J. in *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz, S. & Misener, S.) 365–386 (Humana, 2000).
36. Han, S. & Kim, D. AtRTPrimer: database for *Arabidopsis* genome-wide homogeneous and specific RT-PCR primer-pairs. *BMC Bioinformatics* **7,** 179 (2006).
37. Morohashi, K. *et al.* Participation of the *Arabidopsis* bHLH factor GL3 in trichome initiation regulatory events. *Plant Physiol.* **145,** 736–746 (2007).
38. Ito, T., Ng, K. H., Lim, T. S., Yu, H. & Meyerowitz, E. M. The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in *Arabidopsis. Plant Cell* **19,** 3516–3529 (2007).
39. Blázquez, M. A., Soowal, L. N., Lee, I. & Weigel, D. LEAFY expression and flower initiation in *Arabidopsis. Development* **124,** 3835–3844 (1997).

LETTERS

Genetic analysis of radiation-induced changes in human gene expression

Denis A. Smirnov[1], Michael Morley[2], Eunice Shin[2], Richard S. Spielman[3] & Vivian G. Cheung[1,2,3,4]

Humans are exposed to radiation through the environment and in medical settings. To deal with radiation-induced damage, cells mount complex responses that rely on changes in gene expression. These gene expression responses differ greatly between individuals[1] and contribute to individual differences in response to radiation[2]. Here we identify regulators that influence expression levels of radiation-responsive genes. We treated radiation-induced changes in gene expression as quantitative phenotypes[3,4], and conducted genetic linkage and association studies to map their regulators. For more than 1,200 of these phenotypes there was significant evidence of linkage to specific chromosomal regions. Nearly all of the regulators act *in trans* to influence the expression of their target genes; there are very few *cis*-acting regulators. Some of the *trans*-acting regulators are transcription factors, but others are genes that were not known to have a regulatory function in radiation response. These results have implications for our basic and clinical understanding of how human cells respond to radiation.

In the past 20 years there has been a large increase in the use of radiation in medical diagnostic procedures and treatment protocols. Radiation is genotoxic and induces DNA damage in human cells. To ensure genomic integrity, cells mount complex responses that depend on changes in gene expression. It has long been known that individuals vary in their sensitivity to radiation. This individual variability is also observed at the gene expression level[1]. We and others have used genetic studies to identify chromosomal regions and genetic variants that influence expression levels of many genes in human cells at the baseline[5–9]. Here we extend our analysis by genetic mapping of regulatory elements that influence radiation-induced changes in gene expression.

We used microarrays to measure the expression levels of genes in irradiated immortalized B cells from members of 15 Centre d'Étude du Polymorphisme Humain (CEPH) Utah pedigrees[10]. Data were collected for cells at baseline and at 2 and 6 h after exposure to 10 Gy of ionizing radiation. Of the 10,174 genes on the microarrays that are expressed in immortalized B cells, we focused on 3,280 'ionizing-radiation-responsive' genes that showed at least a 1.5-fold change in gene expression levels at 2 h and/or 6 h after irradiation relative to the baseline. For each of these 3,280 genes we calculated the ratio of expression level at 2 h after irradiation relative to the expression level at the baseline, and the same for the 6 h time point. The $\log_2$ of these ratios is the quantitative trait that we mapped to chromosomal locations by genome scans. From here onwards we will refer to the log-transformed expression ratios as the '2-h-after-irradiation' and '6-h-after-irradiation' expression phenotypes. As shown in Fig. 1, the fold change at 2 h and/or 6 h for some of these radiation-responsive genes varies greatly between individuals.

Because cellular responses rely in part on changes in gene expression, the extent to which the radiation-responsive genes are induced or repressed influences how cells deal with radiation exposure[2]. For instance, one of the variable radiation-induced expression phenotypes is the expression level of *JUN* (Fig. 1). We compared the cellular survival in individuals with low and high *JUN* induction after exposure to 10 Gy of irradiation. The results showed that in individuals with low *JUN* induction, cell survival is higher and cell death is lower than in individuals with high *JUN* induction (Supplementary Fig. 1). To identify the regulators that influence these individual differences in radiation-induced levels of *JUN* and other responsive genes, we carried out genetic analyses. Supplementary Fig. 2 shows a flow chart of our analyses.

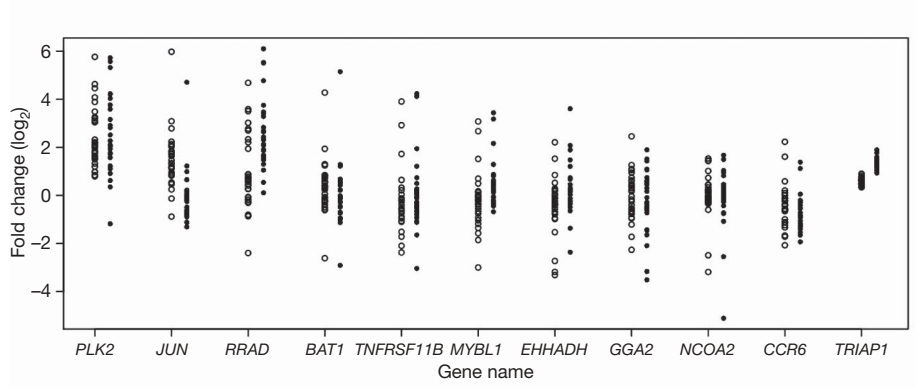


Figure 1 | Individual variation in gene expression response to ionizing radiation. Fold change in 11 radiation-responsive genes; data for each individual are shown as a circle (open circles for 2-h-after-irradiation expression phenotypes; filled circles for 6-h-after-irradiation expression phenotypes). *TRIAP1* (TP53-regulated inhibitor of apoptosis 1) is shown as an example of a gene that showed little individual variation in response to ionizing radiation. Other genes show extensive individual variation in gene expression.

[1]Howard Hughes Medical Institute, [2]The Children's Hospital of Philadelphia, [3]Department of Genetics and [4]Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Genotypes for 4,600 single nucleotide polymorphism (SNP) markers were obtained with a standard SNP-based linkage panel. We used the computer program S.A.G.E. v. 5.4 (http://darwin.cwru.edu/) to carry out genome-wide linkage analysis for each of the 3,280 2-h-after-irradiation and 6-h-after-irradiation expression phenotypes in 15 CEPH families. The analysis gives the strength of the evidence for linkage at each map position in the form of a t value, with an associated pointwise significance level[11]. We selected expression phenotypes for further analysis by using a threshold of $t = 4$ from the S.A.G.E. analysis; in our sample of families, this corresponds to a P value of 4×10^{-5} (lod score about 3.4) and about $P = 0.05$ genome-wide[12]. We found 1,275 (39%) 2-h-after-irradiation phenotypes and 1,298 (40%) 6-h-after-irradiation phenotypes that exceeded this threshold. With a genome-wide threshold of 0.05, among the 3,280 phenotypes we expect 164 at each time point to show linkage evidence anywhere in the genome with a P value this extreme by chance. We found more than 1,250 phenotypes with linkage significant at this level, so we concluded that false positive findings are at most a small fraction of the results. Some of the expression phenotypes have significant evidence of linkage far beyond the $t = 4$ threshold. In Table 1 we show the expression phenotypes with the most significant linkage results. These include *FAM57A* and *GADD45B*, which are known to influence radiation response through the regulation of cell cycle and apoptosis[13–15]. Figure 2 shows examples of genome scan results for several expression phenotypes.

In comparison with baseline expression, the *cis*-regulatory and *trans*-regulatory landscape for radiation-induced gene expression is quite different. We considered *cis* regulators to be those that were mapped within 5 megabases (Mb) of the target gene[5], and all other significant linkage findings to represent *trans* regulators. Of the 1,275 2-h-after-irradiation phenotypes with significant linkage anywhere ($t > 4$), only 9 (less than 1%) were *cis* regulated. Similarly, among the 1,298 6-h-after-irradiation phenotypes, 12 (less than 1%) were *cis* regulated. The remaining phenotypes were *trans* regulated. In contrast, for the baseline gene expression phenotypes, we found that about 20% of the phenotypes had a *cis*-acting regulator and about 80% had a *trans*-acting regulator[5,16].

Hotspots are genomic regions that probably contain regulators influencing the expression levels of many genes[4,5,17]. To identify these regions we divided the autosomal genome into 554 windows of 5 Mb each and determined the number of regulators mapping to each window. We examined the regulators for the 1,275 2-h-after-irradiation phenotypes and the 1,298 6-h-after-irradiation phenotypes with $P < 4 \times 10^{-5}$. We found several windows that contained many more 'hits' than would be expected by chance. If the regulators were randomly distributed across the genome, the probability of 18 or more 'hits' within a 5-Mb window would be less than 3×10^{-5}. Instead, we found four hotspots with 18 or more hits for the 2-h-after-irradiation phenotypes, and two such hotspots for the 6-h-after-irradiation phenotypes. Table 2 shows the phenotypes that mapped to each of these regions. Because these hotspots are 5 Mb in size, it is possible that they contain more than one regulator of gene expression. The target genes whose regulators mapped to the same hotspots seem to have similar functions or are located very close to each other. For example, among the 19 phenotypes whose expression levels map to the regulatory region on chromosome 2 (the 35–40-Mb window) are three genes (*RASL11B*, *RAP1GDS1* and *RASA1*) that encode proteins belonging to the RAS family. Two of the target genes in the regulatory region on chromosome 9 (the 0–5-Mb window), *CD70* and *TNFSF14*, are adjacent to each other on chromosome 19. These observations suggest co-regulation of genes that have similar functions or are closely linked.

Table 1 | Phenotypes with the most significant evidence of linkage

P	Expression phenotype	Location	*Cis/trans*
	2-h-after-irradiation phenotypes		
$<10^{-11}$	*ZNF493*	11q14–q22	*Trans*
$<10^{-10}$	*FAM57A*	17q25	*Trans*
$<10^{-10}$	*OCLN*	9q34	*Trans*
$<10^{-10}$	*HDGF2*	18q21–22	*Trans*
$<10^{-9}$	*PHYH*	15q14–q25	*Trans*
	6-h-after-irradiation phenotypes		
$<10^{-11}$	*EDG1*	3p21–p14	*Trans*
$<10^{-10}$	*GADD45B*	17q12–q21	*Trans*
$<10^{-10}$	*IFNAR2*	6q25–q27	*Trans*
$<10^{-10}$	*CP110*	16p13–p12	*Cis*
$<10^{-9}$	*SLC25A15*	13q13–14	*Cis*

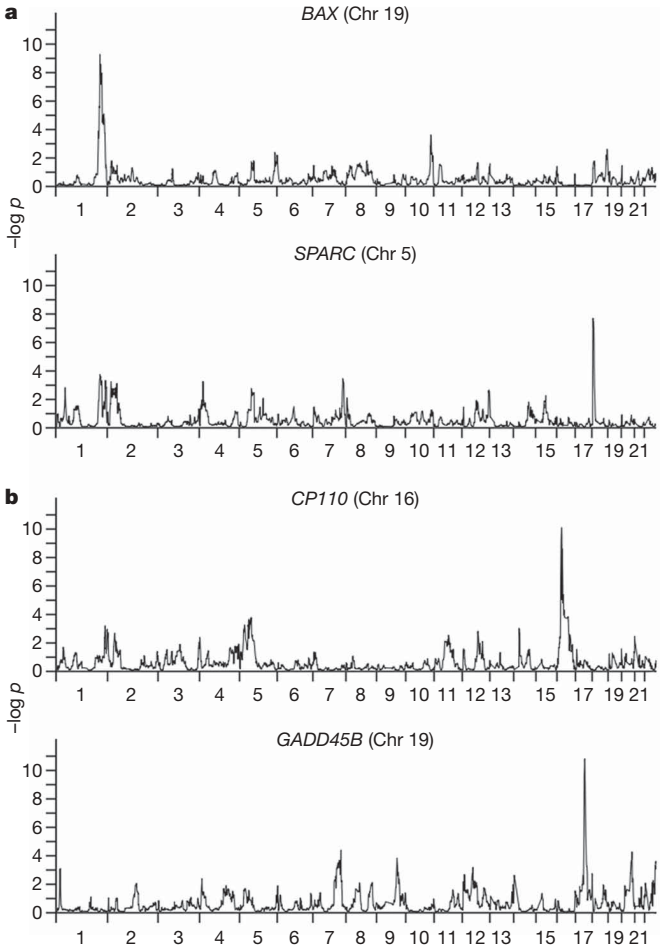


Figure 2 | Genome scans for four expression phenotypes. The chromosomal location of the target gene is given in parentheses. **a**, Example of 2-h-after-irradiation expression phenotypes. **b**, Example of 6-h-after-irradiation expression phenotypes.

Next we followed up results of the linkage scans by family-based association analysis, to confirm the linkage findings and obtain much finer resolution within the regions of linkage. Because there are many significant linkage findings, we focused on phenotypes that showed the largest fold induction or repression in response to ionizing radiation. We selected the expression phenotypes that showed at least a twofold change in expression at 2 h and/or 6 h after irradiation compared to the baseline and had at least one marker with significant evidence ($t = 4$, $P < 4 \times 10^{-5}$) of linkage. There were 182 2-h-after-irradiation and 164 6-h-after-irradiation expression phenotypes that met these criteria.

Of these 346 (182 + 164) phenotypes, 6 were *cis* regulated. We tested each of these *cis*-regulated genes for association by the quantitative transmission disequilibrium test (QTDT)[18], using SNP markers within the target genes and 5 kilobases (kb) upstream and downstream. The QTDT results showed evidence (nominal $P \leq 0.05$) for association (and linkage) for five of these six expression phenotypes, thus supporting the linkage findings that these phenotypes are *cis* regulated. The five phenotypes with *cis* regulation are the

Table 2 | Hotspots and the phenotypes whose regulators mapped to these regions

Region	Phenotypes
	2-h-after-irradiation phenotypes
Chromosome 9: 0–5 Mb	*ASPM*, *STK17B*, *DOCK10*, C4ORF29, *PIK3R1*, *MEF2C*, *BAT3*, *RFX3*, *RXRA*, *CSTF3*, *SLC37A4*, LOC55652, *BCL7A*, *PFAAP5*, *TMCO3*, *ACTN1*, *TIPIN*, *SKIP*, *MYH10*, *CD70*, *TNFSF14*
Chromosome 20: 55–60 Mb	*HOOK1*, *DNAJC6*, *PDE4B*, *RETSAT*, *RAPGEF2*, *BXDC2*, *PRPF4B*, C6ORF32, *MLLT3*, *CTSC*, *DRAM*, *CKAP4*, *LHFP*, *MEIS2*, *EFTUD1*, *PBEF1*, *MGC14376*, *CDK5R1*, *TADA2L*, *KLHL12*
Chromosome 2: 35–40 Mb	*FUBP1*, *CXCR4*, *SLC4A7*, *WDR48*, *RASL11B*, *RAP1GDS1*, *RASA1*, *MAN1A1*, *TWISTNB*, *STAU2*, *NACAP1*, *REXO2*, *FAM62A*, *RNASE6*, C14ORF111, *CENPB*, *RPS21*, *SLC5A3*, *ATP6AP2*
Chromosome 2: 225–230 Mb	*MR1*, *ZFP36L2*, *RY1*, *LAMP3*, *LETM1*, *KIAA0922*, *POLR3G*, *PRPF4B*, *RRS1*, *STIP1*, *PDGFD*, *MLL*, *ITPR2*, *TNFSF11*, *TRAC*, *NQO1*, *ARHGDIA*, C18ORF24, *APOBEC3G*
	6-h-after-irradiation phenotypes
Chromosome 20: 5–10 Mb	C2ORF3, *NCAPG*, *CENTD1*, *IL8*, *HLA-DPA1*, *ASCC3*, *ZMIZ2*, *CSTB*, *RAD54B*, *ADM*, *KLRA1*, *SMUG1*, *PTPRU*, *CPM*, *ING1*, *XYLT1*, *TADA2L*, *KNT2C*, *CYP4F3*
Chromosome 10: 0–5 Mb	KIAA0492, *STK17B*, *NEK1*, EST, *MCM4*, *GTPBP4*, *TSPAN14*, *ARHGAP19*, *DEAF1*, *EP400*, *USP10*, *NPTN*, *ST3GAL2*, *WIPF2*, *ZNP133*

2-h-after-irradiation phenotypes of *NDUT15*, *PMAIP1* and *PTGER4*, and the 6-h-after-irradiation phenotypes of *CP110* and *PHLDA3*.

For the remaining 340 phenotypes (178 at 2 h and 162 at 6 h), the linkage evidence suggests that their expression levels are regulated by *trans*-acting regulators, giving rise to '*trans*-peaks'. One of these phenotypes is the radiation-induced expression level of *BAX* (Fig. 2a), which has a role in apoptosis. We found a highly significant linkage peak on chromosome 1 ($P < 10^{-9}$). This candidate region contains the gene *TP53BP2*, a known regulator of *BAX*[19]. We confirmed the linkage finding by family-based association analysis, which showed significant evidence ($P < 0.02$) for the combined presence of linkage and association at several markers within and near *TP53BP2*. These results illustrate that genetic analysis allows the identification of polymorphic *trans*-acting regulators of gene expression. This is important because few *trans*-acting regulators have previously been identified in studies of the genetics of human gene expression.

Under most of the *trans*-peaks, unlike that for *BAX*, there are no obvious candidates. Often the linkage peaks are megabases in size and include several genes that are possible candidates. To limit the number of potential regulators for follow-up analyses, we used Chilibot[20] to look for co-occurrence of the names of the putative regulator and target genes in the literature and for interactive relationships between them. We used this text-mining program to search for interactive relationships in any cell types and conditions; we did not limit the searches to immortalized B cells or to radiation response. The literature search resulted in 73 potential regulators for 32 expression phenotypes. We performed QTDT analysis using SNP markers within the putative regulators and 5 kb upstream and downstream of these 'regulatory' genes. We further limited our analysis to genes with at least one 'informative' SNP (defined as 30 informative probands for a SNP). This reduced our QTDT analysis from 73 to 58 regulators for 29 expression phenotypes. The results showed evidence (nominal $P < 0.05$) for combined linkage and association for 13 of these 29 phenotypes (2 unlinked regulators for expression of *FAS* 6 h after irradiation). Table 3 shows the linkage and association results for these 13 phenotypes and their corresponding *trans* regulators. We also regressed the expression levels of these 13 expression phenotypes onto SNP markers in their corresponding regulators. Despite the small sample size (30 individuals), significant evidence of association was found between some radiation-induced expression phenotypes and SNP alleles in their *trans* regulators (see Supplementary Table 1). These results show that polymorphisms in the *trans* regulators contribute to individual differences in radiation-induced gene expression.

To confirm the *cis*-linkage and association findings, we conducted a differential allelic gene expression analysis. One of the *cis*-regulated phenotypes is expression of *CP110* (Table 1), which encodes a protein involved in centrosome duplication. Disruption of CP110 leads to unscheduled centrosome duplication[21], radiation-induced chromosomal instability and cell death[22]. To confirm that the radiation-induced level of *CP110* is *cis* regulated, we used quantitative PCR to measure allele-specific changes in *CP110* expression levels in irradiated cells from 12 unrelated CEPH individuals who are heterozygous at an exonic SNP (rs179050) in *CP110*. We found that on average the radiation-induced expression of the T allele at rs179050 was 12% (range 5–40%) more than that of the C allele.

To confirm the findings of *trans*-acting regulators, we used short interfering RNA (siRNA) to knock down the potential regulators and then measured the expression of the corresponding target genes. We assessed the effect of the knockdown by measuring the expression of the regulators and the corresponding target genes in irradiated cells. Results from our association studies had identified 14 potential regulators for 13 radiation-induced expression phenotypes. Successful knockdown was achieved for 11 of the 14 regulators. The expression levels of the regulators decreased by about 30% to about 70% after the genes were silenced, whereas no changes in expression of the regulators were observed when siRNAs with no sequence homology to the regulators were used. We then measured the expression levels of the target genes and found corresponding changes in expression levels of five of

Table 3 | Family-based association analysis in candidate regions identified by linkage scans

Radiation-induced gene expression phenotype	Time after irradiation (h)	Potential regulator	GO function	Linkage (*P*)	Chromosome	QTDT SNP ID	QTDT (*P*)
ARHGDIA	2	*SERPINE2*	Serine peptidase inhibitor	3.21×10^{-7}	2	rs6734100	0.0002
HRAS	2	*RB1*	Transcription factor	5.40×10^{-6}	13	rs198584	0.0033
JUN	2	*LCP2*	Transmembrane receptor signalling pathway	1.46×10^{-9}	5	rs10068619	0.0213
BIRC4	2	*CDK9*	Cyclin-dependent kinase	6.80×10^{-6}	9	rs1002095	0.0305
TNFSF9	2	*CD44*	Receptor	1.38×10^{-5}	11	rs353615	0.0321
HMMR	2	*CDC2*	Cyclin-dependent kinase	1.10×10^{-6}	10	rs1871445	0.0385
RFX3	2	*HIVEP2*	Regulation of transcription	6.00×10^{-7}	6	rs533173	0.0442
TRAF4	6	*FAS*	Signal transduction	4.00×10^{-7}	10	rs9658786	0.0044
TNFSF4	6	*LTA4H*	Zinc ion binding	3.38×10^{-7}	12	rs7296106	0.0057
SPARC	6	*VDR*	Transcription factor activity	1.85×10^{-7}	12	rs2254210	0.0077
RFX3	6	*DYNC2LI1*	Motor activity	4.00×10^{-7}	2	rs2278356	0.01
NT5E	6	*NDUFB6*	Mitochondrial electron transport	5.48×10^{-9}	9	rs628425	0.0179
FAS	6	*SSB*	mRNA binding	1.20×10^{-6}	2	rs7589845	0.0215
FAS	6	*UBA52*	Ribosome/protein modification	1.13×10^{-8}	19	rs2057649	0.0296

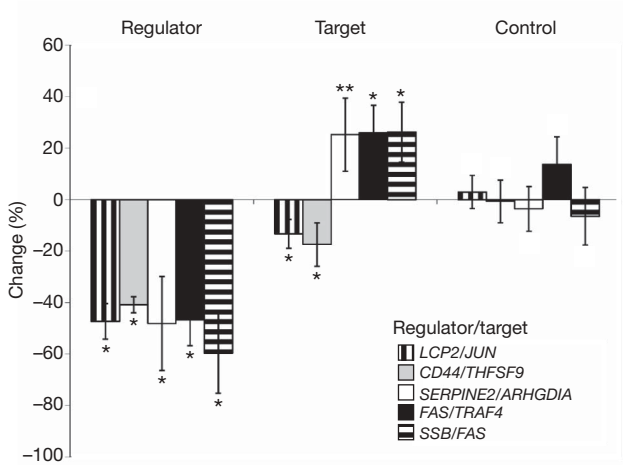


Figure 3 | Results of knockdown of five regulators of radiation-expression phenotypes. Regulators of expression levels of radiation-responsive genes were knocked down by siRNAs. Changes in expression levels of the regulators and their corresponding target genes after knockdown of the regulators are shown, as are changes in expression levels of a control gene (encoding glyceraldehyde-3-phosphate dehydrogenase) after knockdown of the regulators. Data are shown as means ± s.e.m. for four or more independent transfections. Asterisk, $P < 0.05$; two asterisks, $P = 0.07$; no symbol, $P > 0.05$.

them (Fig. 3). Expression levels of the target genes did not change after knockdown of *GAPDH* (Supplementary Fig. 3a), indicating that the changes in expression levels of the target genes were specific effects of silencing their regulators. These results provide molecular support for the regulator–target-gene relationships identified by our genetic analyses.

The five regulator–target-gene pairs that we validated molecularly were: *LCP2* (also known as *SLP-76*) as regulator of expression of *JUN*, *CD44* as regulator of *TNFSF9*, *FAS* as regulator of *TRAF4*, *SERPINE2* as regulator of *ARHGDIA*, and *SSB* as regulator of *FAS* (Fig. 3). Of these five regulators, only *FAS* is a known regulator of signal transduction in response to ionizing radiation. Others such as *SERPINE2* and *SSB* are not previously known to regulate cellular response to radiation.

Changes in expression of the target genes after knockdown of the regulators support the regulator–target-gene relationships identified by genetic mapping. However, the lack of changes in target gene expression does not argue against the regulatory relationship. The expression levels of the regulators were only partly decreased by the siRNAs; depending on the regulatory mechanism, partial expression may be sufficient to influence the response of a target gene to radiation. In addition, because gene expression response to radiation exposure is critical for survival, there are probably backup mechanisms that will influence the target gene expression even after the main regulator is silenced.

We began with gene expression response to radiation because cellular responses rely on changes in gene expression. We also performed an association analysis to test for allelic association with radiation-induced cell death. We focused on the regulators that were confirmed by QTDT; these included 5 *cis* regulators, and the 14 *trans* regulators in Table 3. We measured cell death in irradiated cells from 30 unrelated individuals (parents of the 15 CEPH families) using cytotoxicity and caspase assays, and then tested for evidence of association of these measurements with SNPs within and 5 kb upstream and downstream of the regulators by linear regression. Of the 19 regulators, there were 5 with SNP alleles that were significantly ($P < 0.05$) associated with cytotoxicity and caspase levels (Table 4). This shows that DNA variants in these polymorphic regulators not only affect gene expression but also cellular response to radiation.

In this study we used genetic linkage and association analyses to identify DNA variants that influence the expression levels of genes in irradiated cells. The results allowed us to characterize the global regulatory landscape of gene expression response to radiation and will facilitate the development of genetic tools for radiosensitivity. First, we found that the regulatory landscape of gene expression in irradiated cells is quite different from that of cells at the baseline. Of the gene expression phenotypes in cells at the baseline, at least 20% are *cis* regulated[5,16,23]. In contrast, less than 1% of the radiation-induced expression phenotypes are *cis* regulated; almost all of the phenotypes are regulated by *trans*-acting factors. The large number of *trans*-acting regulators provides cells with multiple mechanisms to mount responses to different types of stressors, and because *trans* regulators can influence the expression of several genes, it also permits coordinated responses. In *C. elegans*[24] and yeast[25], *trans* regulators are also found more frequently than *cis* regulators in influencing gene expression response to external stimuli, suggesting that this may be a phenomenon that is seen across different organisms.

Second, by combining results from genetic mapping and molecular validation studies, we have identified regulatory regions and regulators that influence radiation-induced changes in gene expression. The results allowed us to uncover polymorphic regulators of gene expression response to radiation. One might expect many of the *trans*-acting regulators to be transcription factors; indeed, three of them (*RB1*, *HIVEP2* and *VDR*) are. The remaining regulators include a cell-surface receptor (*CD44*) and genes that have few known functions (*DYNC2LI1* and *UBA52*). These results indicate that many genes other than transcription factors have a function in regulating gene expression. Functions of a large number of human genes remain unknown; the genetics of gene expression is a powerful approach for identifying those that have a regulatory role.

Last, our results have medical implications. Individuals differ in their response to radiation. The regulatory variants identified in this study will enable the genetic prediction of individual sensitivity to radiation. In addition, the identification of genes involved in regulating radiation response will enable the development of radiosensitizers that increase the sensitivity of tumours to radiation. Together, information on patients' sensitivity to radiation and methods of sensitizing

Table 4 | SNPs significantly associated with radiation-induced cytotoxity and caspase levels

Regulator	*Cis/trans*	SNP	Major allele	Phenotype	Average activity*			R^2	*P*
					AA	AB	BB		
PHLDA3	*Cis*	rs3888929; GA	G	Caspase	2.7	3.0	4.4	0.15	0.033
				Cytotoxicity	2.2	2.7	2.6	0.16	0.028
LCP2	*Trans*	rs4867592; CA	C	Caspase	2.4	3.1	4.0	0.33	0.001
				Cytotoxicity	2.9	4.4	4.3	0.23	0.008
LTA4H	*Trans*	rs7970524; TC	T	Caspase	3.2	2.4	NA	0.18	0.018
				Cytotoxicity	2.8	1.7	NA	0.18	0.021
NDUFB6	*Trans*	rs12003093; AG	A	Caspase	2.6	3.2	4.3	0.23	0.007
				Cytotoxicity	3.1	4.1	5.5	0.26	0.004
VDR	*Trans*	rs4760658; AG	A	Caspase	2.8	2.9	4.5	0.18	0.018
				Cytotoxicity	1.9	2.8	3.2	0.16	0.027

* Average activity compared with unirradiated cells; A represents the major allele, and B the minor allele.

malignant cells to radiation will improve outcomes of radiotherapies. Thus, the results of studies of genetics of human gene expression expand our understanding of gene regulation and have clinical implications.

METHODS SUMMARY

The data were from parents and a mean of 8 offspring per sibship (range 7–9) of 15 CEPH families (CEPH 1333, 1341, 1346, 1362, 1408, 1416, 1420, 1421, 1423, 1424, 1444, 1447, 1451, 1454 and 1582). For the expression analysis, immortalized B cells were grown at a density of 5×10^5 cells ml^{-1} and irradiated at 10 Gy in a ^{137}Cs irradiator. Cells were harvested before irradiation and at 2 h and 6 h after exposure to ionizing radiation. RNA was extracted from the cells, labelled and hybridized onto Affymetrix Human U133A 2.0 arrays. For the genetic studies, low-density genotypes for 4,600 autosomal SNP markers were obtained with the Illumina Linkage Panel (v. 3) and high-density genotypes were obtained from the HapMap Project[26] and by inference[27]. Multipoint genome-wide linkage analysis was done by SIBPAL in S.A.G.E. (http://darwin.cwru.edu/) using the 'W4' option[28]. Family-based association analysis with SNP markers within, and 5 kb upstream and downstream of the target genes (*cis*) or candidate regulators (*trans*) was carried out with QTDT software, using the orthogonal (ao) model and variance component options (wega)[18].

Various cellular and molecular analyses were performed. Cell death was measured 24 h after irradiation by using the multitox-fluor multiplex cytotoxicity assay (Promega) and caspase-glo 3/7 assay (Promega). Allele-specific quantitative PCR was performed with complementary DNA samples as templates and TaqMan SNP genotyping assay for rs179050 (Applied Biosystems). For knockdown assays, immortalized B cells were transfected with Accell siRNA pools (Dharmacon) against candidate regulators or non-target control in accordance with the manufacturer's instructions. After transfection, the cells were irradiated at 10 Gy in a ^{137}Cs irradiator and RNA was harvested at the indicated times after irradiation. The effect of siRNA on gene expression was analysed by quantitative PCR.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 September 2008; accepted 24 February 2009.
Published online 6 April 2009.

1. Correa, C. R. & Cheung, V. G. Genetic variation in radiation-induced expression phenotypes. *Am. J. Hum. Genet.* **75,** 885–890 (2004).
2. Amundson, S. A. *et al.* Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. *Cancer Res.* **68,** 415–424 (2008).
3. Cheung, V. G. *et al.* Natural variation in human gene expression assessed in lymphoblastoid cells. *Nature Genet.* **33,** 422–425 (2003).
4. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296,** 752–755 (2002).
5. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430,** 743–747 (2004).
6. Cheung, V. G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437,** 1365–1369 (2005).
7. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422,** 297–302 (2003).
8. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nature Genet.* **39,** 1202–1207 (2007).
9. Stranger, B. E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1,** e78 (2005).
10. Dausset, J. *et al.* Centre d'Étude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6,** 575–577 (1990).
11. Haseman, J. K. & Elston, R. C. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2,** 3–19 (1972).
12. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11,** 241–247 (1995).
13. Pan, D. *et al.* Down-regulation of *CT120A* by RNA interference suppresses lung cancer cells growth and sensitizes to ultraviolet-induced apoptosis. *Cancer Lett.* **235,** 26–33 (2006).
14. Vairapandi, M., Balliet, A. G., Hoffman, B. & Liebermann, D. A. GADD45b and GADD45g are cdc2/cyclinB1 kinase inhibitors with a role in S and G2/M cell cycle checkpoints induced by genotoxic stress. *J. Cell. Physiol.* **192,** 327–338 (2002).
15. Selvakumaran, M. *et al.* The novel primary response gene *MyD118* and the proto-oncogenes *myb*, *myc*, and *bcl-2* modulate transforming growth factor β1-induced apoptosis of myeloid leukemia cells. *Mol. Cell. Biol.* **14,** 2352–2360 (1994).
16. Price, A. L. *et al.* Effects of *cis* and *trans* genetic ancestry on gene expression in African Americans. *PLoS Genet.* **4,** e1000294 (2008).
17. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nature Rev. Genet.* **7,** 862–872 (2006).
18. Abecasis, G. R., Cardon, L. R. & Cookson, W. O. A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66,** 279–292 (2000).
19. Samuels-Lev, Y. *et al.* ASPP proteins specifically stimulate the apoptotic function of p53. *Mol. Cell* **8,** 781–794 (2001).
20. Chen, H. & Sharp, B. M. Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics* **5,** 147 (2004).
21. Chen, Z. *et al.* CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells. *Dev. Cell* **3,** 339–350 (2002).
22. Sato, N., Mizumoto, K., Nakamura, M. & Tanaka, M. Radiation-induced centrosome overduplication and multiple mitotic spindles in human tumor cells. *Exp. Cell Res.* **255,** 321–326 (2000).
23. Pant, P. V. *et al.* Analysis of allelic differential expression in human white blood cells. *Genome Res.* **16,** 331–339 (2006).
24. Li, Y. *et al.* Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* **2,** e222 (2006).
25. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* **6,** e83 (2008).
26. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
27. Burdick, J. T., Chen, W. M., Abecasis, G. R. & Cheung, V. G. *In silico* method for inferring genotypes in pedigrees. *Nature Genet.* **38,** 1002–1004 (2006).
28. Shete, S., Jacobs, K. B. & Elston, R. C. Adding further power to the Haseman and Elston method for detecting linkage in larger sibships: weighting sums and differences. *Hum. Hered.* **55,** 79–85 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. George and W. Ewens for advice and discussion, A. Bruzel, S. Solomon, T. Weber and K. Halasa for technical help, and C. McGarry for manuscript preparation. Some analyses for this paper were performed by using the program package S.A.G.E., which is supported by a grant from the National Center for Research Resources. This work is supported by grants from the National Institutes of Health (to V.G.C. and R.S.S.), by seed grants from the University of Pennsylvania Center for Excellence in Environmental Toxicology (to V.G.C.), by the W. W. Smith Endowed Chair (to V.G.C.) and the Howard Hughes Medical Institute (to V.G.C.).

Author Information The microarray data are deposited in the GEO database under accession number GSE12626. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to V.G.C. (vcheung@mail.med.upenn.edu).

METHODS

CEPH samples and expression phenotyping. The data were from parents and a mean of eight offspring per sibship (range seven to nine) of 15 CEPH families (CEPH 1333, 1341, 1346, 1362, 1408, 1416, 1420, 1421, 1423, 1424, 1444, 1447, 1451, 1454 and 1582). For the expression analysis, immortalized B cells (lymphoblastoid cells) were grown at a density of 5×10^5 cells ml^{-1} in RPMI 1640 with 15% fetal bovine serum, 2 mM L-glutamine and 100 U ml^{-1} penicillin/streptomycin and irradiated at 10 Gy in a ^{137}Cs irradiator. Cells were harvested before irradiation and at 2 h and 6 h after exposure to ionizing radiation. RNA was extracted from the cells and hybridized onto Affymetrix Human U133A 2.0 arrays. We used a random number generator to determine the order in which the cells were grown, and array hybridizations were performed; cells from family members were not processed together except by chance. The baseline, 2-h-after-irradiation and 6-h-after-irradiation samples for each individual were processed together. The cRNA samples were prepared in a total of four batches (about 96 samples per batch). Hybridizations were performed in batches of 48 samples. Expression intensity was scaled to 500 using the global scaling method implemented in the Expression Console software (Affymetrix) and transformed by $\log_2$. We defined 'radiation-responsive' genes as those that showed at least a 1.5-fold change in gene expression levels in 6 or more of 30 unrelated individuals. We focused on these 3,280 radiation-responsive genes in this study.

Viability, cytotoxicity and caspase activity measurements. Immortalized B-cell lines from 30 CEPH parents and individuals with high (GM07048, GM10834 and GM10861) or low (GM10842, GM10843 and GM12752) radiation induction of *JUN* expression were irradiated at 10 Gy in a ^{137}Cs irradiator. The cells were then analysed 24 h after irradiation with a multitox-fluor multiplex cytotoxicity assay (Promega) and a caspase-glo 3/7 assay (Promega).

Genotypes. Low-density genotypes for 4,600 autosomal SNP markers were obtained with the Linkage Panel, v. 3 (Illumina). We used PEDSTATS[29] to check for Mendelian inconsistencies. This resulted in the removal of 48 genotypes at four distinct SNP markers. High-density genotypes for family-based association (QTDT) were obtained by inference using the low-density genotypes and high-density genotypes on selected individuals[27].

Analysis of linkage and association. Multipoint genome-wide linkage analysis was done by SIBPAL in S.A.G.E. v. 5.4 (http://darwin.cwru.edu/) using the 'W4' option[28]. SIBPAL determines evidence for linkage at each SNP from regression of the phenotype difference between siblings on the estimated proportion of marker alleles shared identical-by-descent between siblings; the result is reported as a *t* value with corresponding significance, as given in the text. Pointwise significance was converted to genome-wide significance[12]. In permutation analysis by S.A.G.E. of the results for 14 phenotypes (6 from Table 1, and 8 phenotypes with *t* values close to 5: 4.99–5.4), we found one phenotype with two *t* values over 5 and one phenotype with seven *t* values over 5 among 100,000 replicates. For the other 12 phenotypes (including the 6 from Table 1), we did not find any *t* values over 5 in 100,000 permutations for each phenotype.

Family-based association analysis with SNPs near and within the target genes or candidate regulators was performed with QTDT. We used the orthogonal (ao) model[18] and variance component options (wega). For association analysis, after irradiation the expression phenotype, cytotoxicity or caspase level, as dependent variable, was regressed on SNP genotypes (coded 0, 1 or 2). R^2 was estimated for each phenotype–SNP combination as the ratio of the regression sum of squares to the total sum of squares.

Hotspots. The autosomal genome was divided into 554 windows of 5 Mb each (the last window of the q arm of a chromosome can be smaller). For each of the 1,275 2-h-after-irradiation phenotypes and 1,298 6-h-after-irradiation phenotypes, we considered all SNPs with $t > 4$, $P < 4 \times 10^{-5}$. Any window with one or more such SNPs was counted as having one 'hit' for that phenotype. Some phenotypes had more than one hit in the genome because some had multiple linkage peaks, or because peaks for some phenotypes were broad and spanned adjacent 5-Mb windows; in this case, each window was counted as having one hit. There were 3,151 such hits for the 2-h-after-irradiation phenotypes and 3,120 hits for the 6-h-after-irradiation phenotypes. We assumed that if the hits were distributed randomly across the genome, their distribution over windows would be approximately Poisson, with a mean of 5.69 and 5.63 for the 2-h and 6-h phenotypes, respectively (3,151/554 and 3,120/554).

SNP genotyping and allele-specific RT–PCR. Allele-specific quantitative PCR was performed with cDNA samples as templates, using TaqMan primer for rs179050 (Applied Biosystems). A standard curve for the quantitative PCR was produced with genomic DNA from individuals homozygous for rs179050. The standard curve showed that genomic DNA from the individuals heterozygous at that SNP contained about 50% of the T and C alleles as expected. The standard curve was then used to quantify the T:C allele in the 12 cDNA samples at the baseline and 6 h after irradiation.

Knockdown of candidate regulators. Immortalized B cells were transfected with Accell siRNAs (Dharmacon) against candidate regulators or non-target control in accordance with the manufacturer's instructions. For each regulator we used a pool of siRNAs that targeted the regulators, to minimize off-target effects[30]. After transfection, the cells were irradiated at 10 Gy in a ^{137}Cs irradiator, and RNA was harvested after irradiation. In addition to the siRNAs, we also used short hairpin RNA (shRNA) to knock down two regulators, LCP2 and CD44. Although the shRNAs were not as effective as the siRNAs in the knockdown, similar trends were seen, further supporting the regulator–target-gene relationships (Supplementary Fig. 3b). MISSION shRNA plasmid (Sigma-Aldrich) transfections were performed with the nucleofector cell line nucleofector kit V (Amaxa), in accordance with the manufacturer's instructions. The effect of siRNA (or shRNA) on gene expression was analysed by quantitative PCR. Expression of β-actin was used as a control for normalization, and changes in expression were calculated relative to cells transfected with non-target control siRNA or shRNA. Sequences of PCR primers, siRNAs and shRNAs are presented in Supplementary Tables 2–4.

29. Wigginton, J. E. & Abecasis, G. R. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* **21**, 3445–3447 (2005).
30. Myers, A. J. *et al.* Minimizing off-target effects by using diced siRNAs for RNA interference. *J RNAi Gene Silencing* **2**, 181–194 (2006).

LETTERS

Irreversibility of mitotic exit is the consequence of systems-level feedback

Sandra López-Avilés[1], Orsolya Kapuy[2,3], Béla Novák[2] & Frank Uhlmann[1]

The eukaryotic cell cycle comprises an ordered series of events, orchestrated by the activity of cyclin-dependent kinases (Cdks), leading from chromosome replication during S phase to their segregation in mitosis. The unidirectionality of cell-cycle transitions is fundamental for the successful completion of this cycle. It is thought that irrevocable proteolytic degradation of key cell-cycle regulators makes cell-cycle transitions irreversible, thereby enforcing directionality[1–3]. Here we have experimentally examined the contribution of cyclin proteolysis to the irreversibility of mitotic exit, the transition from high mitotic Cdk activity back to low activity in G1. We show that forced cyclin destruction in mitotic budding yeast cells efficiently drives mitotic exit events. However, these remain reversible after termination of cyclin proteolysis, with recovery of the mitotic state and cyclin levels. Mitotic exit becomes irreversible only after longer periods of cyclin degradation, owing to activation of a double-negative feedback loop involving the Cdk inhibitor Sic1 (refs 4, 5). Quantitative modelling suggests that feedback is required to maintain low Cdk activity and to prevent cyclin resynthesis. Our findings demonstrate that the unidirectionality of mitotic exit is not the consequence of proteolysis but of systems-level feedback required to maintain the cell cycle in a new stable state.

After completion of chromosome segregation during mitosis, the activity of the key cell-cycle kinase Cdk is downregulated to promote mitotic exit and the return of cells to G1. This involves ubiquitin-mediated degradation of mitotic cyclins under control of the anaphase promoting complex (APC), a multisubunit ubiquitin ligase. Mitotic cyclins are initially targeted for degradation by the APC in association with its activating subunit Cdc20 (APC^{Cdc20}). Later, declining Cdk levels and activation of the Cdk-counteracting phosphatase Cdc14 allow a second APC activator, Cdh1, to associate with the APC (APC^{Cdh1})[6–8]. Cyclin proteolysis, a thermodynamically irreversible reaction, is thought to be responsible for the irreversibility of mitotic exit[1–3]. However, *de novo* protein synthesis can counteract degradation and constitutes a similar thermodynamically irreversible process, driven by ATP hydrolysis. In a cellular setting, therefore, protein levels are defined by reversible changes to the rates of two individually irreversible reactions: protein synthesis and degradation. These considerations have led to the hypothesis that it is not proteolysis itself, but systems-level feedback that affects synthesis and degradation rates, making cell-cycle transitions irreversible[9].

To test this hypothesis, we investigated the contribution of cyclin proteolysis to the irreversibility of budding yeast mitotic exit (Fig. 1a). We arrested budding yeast cells in mitosis with high levels of mitotic cyclins by depleting Cdc20 under control of the *MET3* promoter. In these cells, we induced Cdh1 expression from the galactose-inducible *GALL* promoter (an attenuated version of the *GAL1* promoter) for 30 min[6]. We expressed a Cdh1 variant, Cdh1(m11), that activates the APC even in the presence of high Cdk activity owing to mutation of 11 Cdk phosphorylation sites. This led to efficient degradation of the major budding yeast mitotic cyclin Clb2 (Fig. 1b). The mitotic Polo-like kinase, another APC^{Cdh1} target[10], was also efficiently degraded, whereas levels of the S phase cyclin Clb5, a preferential substrate for APC^{Cdc20} (ref. 11), remained largely unaffected (Supplementary Fig. 1). Clb2 destruction was accompanied by dephosphorylation of known mitotic Cdk substrates, seen by their change in electrophoretic mobility (Fig. 1b). These included three proteins whose dephosphorylation contributes to spindle elongation and chromosome segregation, Sli15, Ase1 and Ask1 (refs 12–14). Their dephosphorylation depended on the activity of the mitotic exit phosphatase Cdc14 (Supplementary Fig. 2). Mitotic spindles that were present in the metaphase-arrested cells disassembled as Clb2 levels declined, accompanied by outgrowth of pronounced astral microtubules (Fig. 1c and Supplementary Fig. 3), reminiscent of spindle breakdown at the end of mitosis. $APC^{Cdh1(m11)}$-mediated destruction of the spindle stabilizing factor Ase1 (Fig. 1b), in addition to its dephosphorylation, may contribute to this phenotype[15].

After 50 min, when Clb2 levels became almost undetectable, we turned off $APC^{Cdh1(m11)}$ by inactivating a temperature-sensitive APC core subunit encoded by the *cdc16-123* allele[16]. Notably, in response to $APC^{cdc16\text{-}123}$ inactivation, Clb2 levels recovered and Sli15, Ase1 and Ask1 reappeared in their mitotic hyperphosphorylated forms. Mitotic spindles formed again, suggesting that cells had re-entered a mitotic state. Spindles appeared morphologically intact, but were longer after cyclin reaccumulation ($3.9 \pm 0.8\,\mu m$; mean ± s.d.) compared to metaphase spindles at the beginning of the experiment ($2.1 \pm 0.6\,\mu m$). A probable reason for this lies in compromised sister chromatid cohesion after some, albeit inefficient, inactivation of the anaphase inhibitor securin by $APC^{Cdh1(m11)}$ (Supplementary Fig. 4)[17]. Fluorescence-activated cell sorting (FACS) analysis of DNA content confirmed that cells maintained a 2C DNA content throughout the experiment (Fig. 1d). This demonstrates that cyclin destruction promotes mitotic exit events, but it is not sufficient to render them irreversible. Cyclin resynthesis can reverse mitotic exit. Note that reversibility of mitotic exit under these conditions did not depend on the persistence of the S phase cyclin Clb5 (Supplementary Fig. 1).

We next addressed what makes mitotic exit irreversible, if not cyclin destruction. When we repeated the experiment, but continued Cdh1(m11) induction and inactivated $APC^{cdc16\text{-}123}$ only after 60 min, Clb2 did not reaccumulate and over half of the cells subsequently completed cytokinesis and entered G1 (Fig. 2a). This suggests that after longer periods of Clb2 destruction mitotic exit becomes irreversible. Western blotting showed that around the time when Clb2 loss turned irreversible, the Cdk inhibitor Sic1 accumulated[4,5]. We therefore asked whether Sic1 accumulation was responsible for the irreversibility of mitotic exit. When we repeated the experiment using a *sic1Δ* strain, Clb2 reappeared after $APC^{cdc16\text{-}123}$ inactivation, and only a minority of

[1]Chromosome Segregation Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. [2]Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [3]Budapest University of Technology and Economics, Gellért tér 4, 1521 Budapest, Hungary.

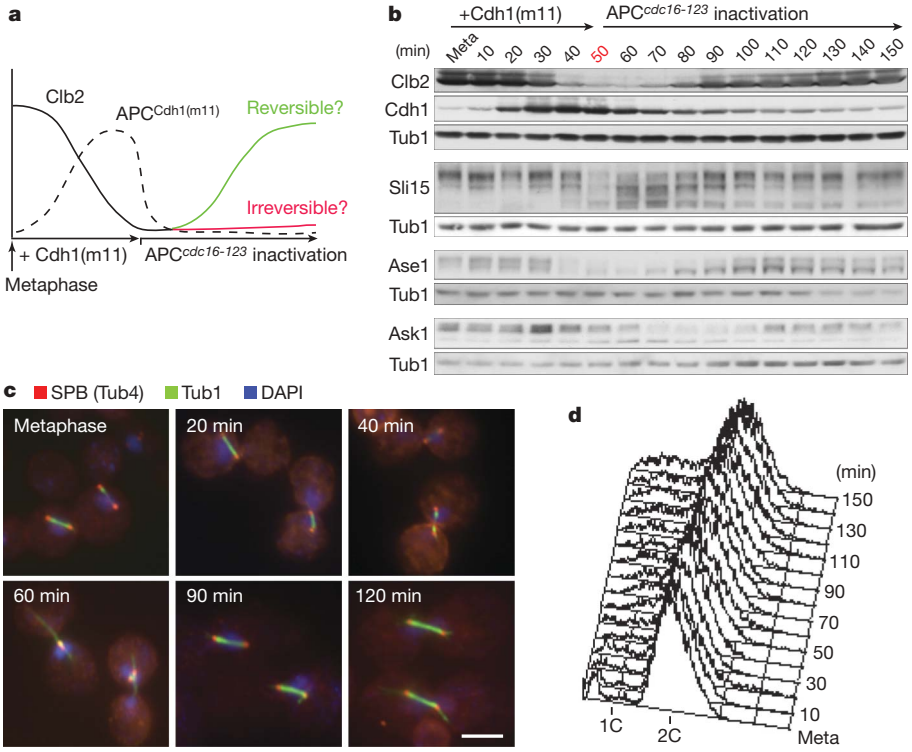


Figure 1 | Clb2 destruction promotes reversible mitotic exit events. **a**, Scheme depicting the experimental design and the predicted outcomes if cyclin proteolysis does, or does not, make mitotic exit irreversible. **b**, $APC^{Cdh1(m11)}$-driven Clb2 destruction is reversible and leads to reversible Cdk substrate dephosphorylation. Cdh1(m11) was induced in metaphase-arrested cells for 30 min, and $APC^{Cdh1(m11)}$ activity was terminated after 50 min by inactivation of the *cdc16-123* allele at 37 °C. Cdh1(m11) was detected by western blotting against its amino-terminal haemagglutinin (HA) epitope; Sli15, Ask1 and Ase1 were detected via carboxy-terminal paramyxovirus SV5 P-k (Pk) epitopes. Tub1 served as a loading control. **c**, Clb2 degradation and reaccumulation are accompanied by spindle breakdown and reassembly. As in **b**, but cells were processed for indirect immunofluorescence to visualize the spindle pole body (SPB) component γ-tubulin (Tub4), mitotic spindles (tubulin, Tub1) and nuclear DNA (stained with 4,6-diamidino-2-phenylindole, DAPI). Scale bar, 5 μm. **d**, FACS analysis of the DNA content of the cells in **c** confirms their mitotic arrest throughout the time course.

cells proceeded to cytokinesis (Fig. 2a). In the absence of Sic1, mitotic exit remained reversible even after prolonged Clb2 destruction for 90 min (Supplementary Fig. 5). Sic1 is part of a double-negative feedback loop in which Cdk downregulation allows the Cdc14 phosphatase to dephosphorylate Sic1, as well as its transcription factor Swi5, increasing the expression and stability of the Cdk inhibitor[5,9]. Because Clb2 positively regulates its own synthesis[18,19], Clb2 inhibition by Sic1 may be required to prevent Clb2 resynthesis.

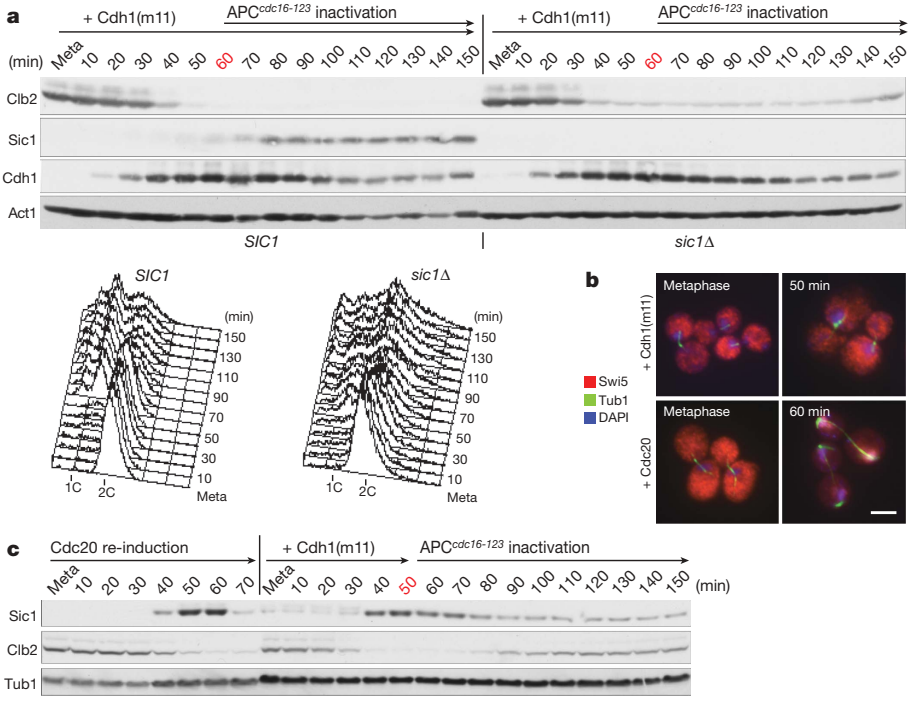


Figure 2 | Irreversibility of mitotic exit requires feedback loop activation. **a**, Irreversible mitotic exit after longer periods of $APC^{Cdh1(m11)}$ activity depends on Sic1. As in Fig. 1, but Cdh1(m11) expression was not terminated and *cdc16-123* was inactivated after 60 min. Mating pheromone α-factor (5 μg ml^{-1}) was added to prevent progression to the next cell cycle. FACS analysis of DNA content (bottom panels) shows completion of cytokinesis in cells containing Sic1. **b**, **c**, Limited activation of feedback loop components during reversible mitotic exit. Swi5 retains cytoplasmic localization (**b**), and Sic1 accumulation is incomplete (**c**) during reversible $APC^{Cdh1(m11)}$-driven mitotic exit. For comparison, cells were released from metaphase arrest into synchronous mitotic exit by Cdc20 reinduction. Swi5 was visualized by indirect immunofluorescence. Levels of Clb2 and Sic1 were analysed by western blotting. Scale bar, 5 μm.

These findings indicate that mitotic exit remained reversible for up to 50 min because the Sic1 feedback loop had not yet been sufficiently activated. Consistent with this possibility, dephosphorylation-dependent translocation of Swi5 from the cytoplasm to the nucleus, indicative of Swi5 activation during mitotic exit[5,20], was inefficient under these conditions (Fig. 2b). Similarly, the levels of Sic1 that became detectable remained below the levels observed in cells undergoing mitotic exit after release from the metaphase block by Cdc20 reinduction (Fig. 2c). If irreversibility of mitotic exit is due to activation of the Sic1-dependent feedback loop, irreversibility should be advanced by increasing Cdc14 phosphatase activity, or by directly enhancing Sic1 levels. As predicted, ectopic expression of either Cdc14 or a version of Sic1 that is stable in the presence of high Cdk activity owing to mutation of three Cdk phosphorylation sites[21], Sic1(m3), made Clb2 destruction irreversible under conditions that otherwise permit Clb2 resynthesis (Fig. 3a). This confirms that activation of the Sic1-dependent feedback loop limits the irreversibility of mitotic exit.

It has been shown that mammalian mitotic exit can be driven in the absence of cyclin proteolysis by chemical inhibition of Cdk activity. After short, but not longer periods, of Cdk inhibition mitotic exit remained reversible[3]. We repeated these experiments in budding yeast cells carrying the ATP analogue (1NM-PP1)-sensitive Cdk allele *cdc28-as1* (ref. 22). As in mammalian cells, we observed reversible mitotic exit events, exemplified by dephosphorylation of the Cdk substrate Orc6 (ref. 23), after transient Cdk inhibition for 10 min. After 50 min of inhibitor treatment, Orc6 dephosphorylation turned irreversible. Irreversibility again correlated with, and depended on, the accumulation of Sic1 (Fig. 3b). This suggests that feedback-loop activation is responsible, and that cyclin destruction is not required, for irreversible mitotic exit. In the presence of cyclin destruction, a shorter period of Cdk inhibition was sufficient to render mammalian mitotic exit irreversible[3]. Although this was taken to demonstrate a requirement for cyclin destruction, we suggest that activation of a feedback loop correlated with cyclin destruction that made mitotic exit irreversible.

To theoretically investigate the contribution of feedback to the irreversibility of budding yeast mitotic exit, we used a mathematical model[24] to describe the cell-cycle control network that operates during mitotic exit (Fig. 4a, a detailed description is found in Supplementary Fig. 6). In our experiments, Cdk downregulation begins by Cdh1(m11)-mediated Clb2 degradation, or 1NM-PP1 addition, which causes Sic1 accumulation because of the double-negative feedback loops. If Clb2 proteolysis is terminated by inactivating $APC^{cdc16\text{-}123}$, or if 1NM-PP1 is removed, before Sic1 reaches a threshold, Sic1 accumulation becomes transient and Cdk activity will recover (Fig. 4b, e). In the absence of Sic1, mitotic exit will therefore always remain reversible (Fig. 4d, g). Clb2 destruction becomes irreversible only if Sic1 levels have reached a threshold that maintains Cdk activity low enough to prevent Clb2 resynthesis (Fig. 4c). Mitotic exit after chemical Cdk inhibition turns irreversible when Sic1 levels are sufficient to maintain Cdk inhibition independently of 1NM-PP1 (Fig. 4f). Note that during normal mitotic exit, Cdk downregulation is initiated by APC^{Cdc20}-mediated Clb2 destruction, and APC^{Cdh1} activation by the decreasing Cdk/Cdc14 ratio forms an additional double-negative feedback loop that maintains low Cdk activity redundantly with Sic1. In mammalian cells, the antagonistic relationship between Cdk and Cdh1, and between Cdk and its inhibitory tyrosine phosphorylation[25,26], create double-negative feedback loops of Cdk inactivation that probably contribute to irreversibility of mitotic exit.

Protein destruction is a commonly used mechanism controlling key cell-cycle transitions. The rise in Cdk activity at the G1/S transition is accompanied by the degradation of Cdk inhibitors. The irreversibility of this transition, however, probably stems from positive feedback during Cdk activation[27]. Likewise, even though the G2/M transition in the vertebrate cell cycle involves proteolysis of the Cdk inhibitory

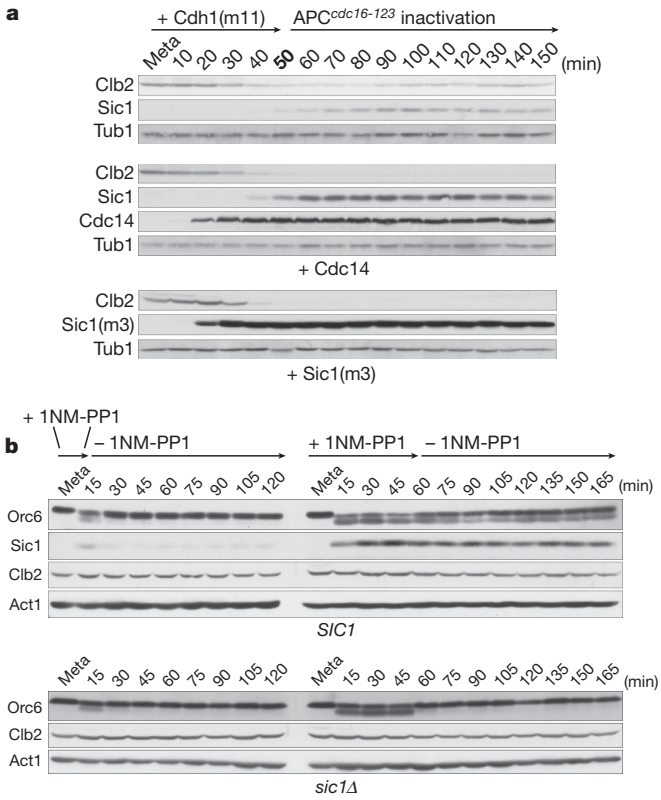


Figure 3 | Sic1 turns mitotic exit irreversible. a, Ectopic expression of Cdc14 or Sic1(m3) advances irreversibility of mitotic exit. Cdh1(m11) was expressed without or together with Cdc14 or Sic1(m3) for 30 min, before $APC^{cdc16\text{-}123}$ was inactivated after 50 min. α-factor was added as in Fig. 2a. **b**, Sic1 promotes irreversible mitotic exit in the absence of APC activity after chemical Cdk inhibition. Cdk (*cdc28-as1*) was inhibited by addition of 5 μM 1NM-PP1 in metaphase-arrested *SIC1* and *sic1Δ* cells, depleted for Cdc20 and $APC^{cdc16\text{-}123}$-inactivated at 37 °C. After 10 or 50 min, 1NM-PP1 was washed out while $APC^{cdc16\text{-}123}$ remained inactive. Levels and gel mobility of the indicated proteins were analysed by western blotting. Tub1 (**a**) or Act1 (**b**) served as loading controls.

kinase Wee1, this transition shows the characteristics of a bistable switch, driven by feedback between Cdk, Wee1 and the Cdk-activating phosphatase Cdc25 (refs 28, 29). Although not irreversible in a cellular context, protein degradation, and in particular protein resynthesis, occur at a slower timescale than the addition or removal of posttranslational modifications. Proteolysis thereby introduces an element of distinct dynamic nature into the cell-cycle control network, the consequences of which merit further investigation.

METHODS SUMMARY

Yeast strains. A list of strains used in this study is found in Supplementary Table 1. Epitope tagging of endogenous genes and gene deletions were performed by gene targeting using PCR products. Integrative expression vectors for Cdh1(m11) and Cdc14 under control of the *GALL* and *GAL1* promoters, respectively, were as described[6,13]; the YIplac211GAL1-SIC1(m3)-HA vector was a gift from E. Schwob.

Experimental procedures. Cells were grown at 25 °C in synthetic complete (SC) medium lacking methionine, with 2% raffinose as carbon source, and arrested in metaphase by depletion of Cdc20 under control of the *MET3* promoter by addition of 2 mM methionine for 5 h. Expression of Cdh1(m11), Sic1(m3) or Cdc14 was induced by addition of 2% galactose. Induction was terminated by addition of 2% glucose, and $APC^{cdc16\text{-}123}$ was inactivated by shifting the culture to a waterbath at 37 °C. Alternatively, metaphase-arrested cells were released into synchronous mitotic progression by Cdc20 reinduction after filtration and resuspension in methionine-free medium. Protein extracts were prepared using the TCA method[30]. Antibodies used for western blotting and immunostaining were: anti-HA 12CA5, anti-myc 9E10, anti-Pk SV5-Pk1 (Serotec), anti-Orc6 SB49 (a gift from B. Stillman), anti-Clb2 (y-180), anti-Clb5 (yN-19), and anti-Sic1 antisera (FL-284, all Santa Cruz Biotechnology), anti-Tub1 YOL1/34

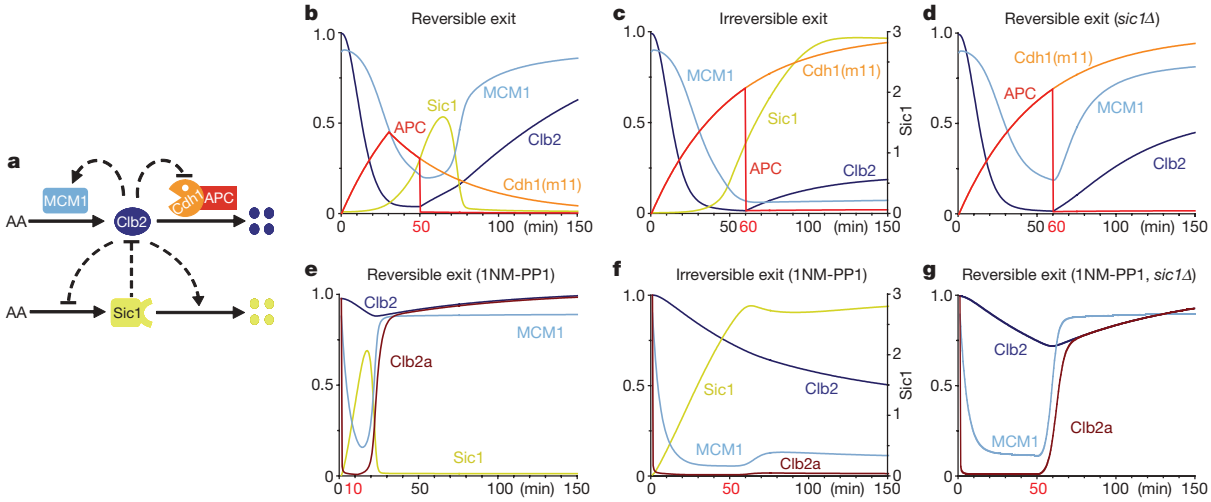


Figure 4 | Computational analysis of mitotic exit. **a**, Wiring diagram for Clb2/Cdk (abbreviated Clb2) and Sic1 regulation during budding yeast mitotic exit. AA, amino acids. **b**–**g**, Numerical simulations of protein levels and activities with a mathematical model (see Supplementary Fig. 6) during Cdh1(m11)-induced mitotic exit (**b**–**d**) and after chemical Cdk inhibition (**e**–**g**). Clb2 represents the total level of Clb2/Cdk complexes, including inactive complexes bound to Sic1; Clb2a denotes its associated kinase activity; MCM1 denotes the active form of the Clb2 transcription factor complex Fkh2/Ndd1/Mcm1; and APC represents the level of active $APC^{Cdh1(m11)}$. In reversible exit (**b**, **e**), $APC^{Cdh1(m11)}$ activity is terminated after 50 min, or Cdk inhibition by 1NM-PP1 is released after 10 min. In irreversible exit (**c**, **f**), $APC^{Cdh1(m11)}$ activity continues for 60 min, or Cdk inhibition for 50 min. In reversible exit in *sic1Δ* cells (**d**, **g**), as in **c** and **f** but Sic1 synthesis is zero.

(Serotec), anti-actin N350 (Amersham), anti-Tub4 serum (a gift from J. Kilmartin), and an antiserum raised against recombinant Cdc14 purified after overexpression in *Escherichia coli*.

Received 22 January; accepted 12 March 2009.
Published online 22 April 2009.

1. King, R. W., Deshaies, R. J., Peters, J.-M. & Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274,** 1652–1659 (1996).
2. Reed, S. I. Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover. *Nature Rev. Mol. Cell Biol.* **4,** 855–864 (2003).
3. Potapova, T. A. *et al.* The reversibility of mitotic exit in vertebrate cells. *Nature* **440,** 954–958 (2006).
4. Donovan, J. D., Toyn, J. H., Johnson, A. L. & Johnston, L. H. $P40^{SDB25}$, a putative CDK inhibitor, has a role in the M/G_1 transition in *Saccharomyces cerevisiae*. *Genes Dev.* **8,** 1640–1653 (1994).
5. Visintin, R. *et al.* The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Mol. Cell* **2,** 709–718 (1998).
6. Zachariae, W., Schwab, M., Nasmyth, K. & Seufert, W. Control of cyclin ubiquitination by CDK-regulated binding of Hct1 to the anaphase promoting complex. *Science* **282,** 1721–1724 (1998).
7. Jaspersen, S. L., Charles, J. F. & Morgan, D. O. Inhibitory phosphorylation of the APC regulator Hct1 is controlled by the kinase Cdc28 and the phosphatase Cdc14. *Curr. Biol.* **9,** 227–236 (1999).
8. Yeong, F. M., Lim, H. H., Padmashree, C. G. & Surana, U. Exit from mitosis in budding yeast: Biphasic inactivation of the Cdc28-Clb2 mitotic kinase and the role of Cdc20. *Mol. Cell* **5,** 501–511 (2000).
9. Novak, B., Tyson, J. J., Gyorffy, B. & Csikasz-Nagy, A. Irreversible cell-cycle transitions are due to systems-level feedback. *Nature Cell Biol.* **9,** 724–728 (2007).
10. Shirayama, M., Zachariae, W., Ciosk, R. & Nasmyth, K. The Polo-like kinase Cdc5p and the WD-repeat protein Cdc20p/fizzy are regulators and substrates of the anaphase promoting complex in *Saccharomyces cerevisiae*. *EMBO J.* **17,** 1336–1349 (1998).
11. Shirayama, M., Toth, A., Galova, M. & Nasmyth, K. APC^{Cdc20} promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. *Nature* **402,** 203–207 (1999).
12. Pereira, G. & Schiebel, E. Separase regulates INCENP-Aurora B anaphase spindle function through Cdc14. *Science* **302,** 2120–2124 (2003).
13. Higuchi, T. & Uhlmann, F. Stabilization of microtubule dynamics at anaphase onset promotes chromosome segregation. *Nature* **433,** 171–176 (2005).
14. Khmelinskii, A., Lawrence, C., Roostalu, J. & Schiebel, E. Cdc14-regulated midzone assembly controls anaphase B. *J. Cell Biol.* **177,** 981–993 (2007).
15. Juang, Y.-L. *et al.* APC-mediated proteolysis of Ase1 and the morphogenesis of the mitotic spindle. *Science* **275,** 1311–1314 (1997).
16. Irniger, S., Piatti, S., Michaelis, C. & Nasmyth, K. Genes involved in sister chromatid separation are needed for B-type cyclin proteolysis in budding yeast. *Cell* **81,** 269–277 (1995).
17. Schwab, M., Neutzner, M., Möcker, D. & Seufert, W. Yeast Hct1 recognizes the mitotic cyclin Clb2 and other substrates of the ubiquitin ligase APC. *EMBO J.* **20,** 5165–5175 (2001).
18. Reynolds, D. *et al.* Recruitment of Thr 319-phosphorylated Ndd1p to the FHA domain of Fkh2p requires Clb kinase activity: a mechanism for CLB cluster gene activation. *Genes Dev.* **17,** 1789–1802 (2003).
19. Pic-Taylor, A., Darieva, Z., Morgan, B. A. & Sharrocks, A. D. Regulation of cell cycle-specific gene expression through cyclin-dependent kinase-mediated phosphorylation of the forkhead transcription factor Fkh2p. *Mol. Cell. Biol.* **24,** 10036–10046 (2004).
20. Moll, T., Tebb, G., Surana, U., Robitsch, H. & Nasmyth, K. The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SWI5. *Cell* **66,** 743–758 (1991).
21. Verma, R. *et al.* Phosphoregulation of Sic1p by G_1 Cdk required for its degradation and entry into S phase. *Science* **278,** 455–460 (1997).
22. Bishop, A. C. *et al.* A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407,** 395–401 (2000).
23. Nguyen, V. Q., Co, C. & Li, J. J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411,** 1068–1073 (2001).
24. Queralt, E., Lehane, C., Novak, B. & Uhlmann, F. Downregulation of $PP2A^{Cdc55}$ phosphatase by separase initiates mitotic exit in budding yeast. *Cell* **125,** 719–732 (2006).
25. Kramer, E. R., Scheuringer, N., Podtelejnikov, A. V., Mann, M. & Peters, J.-M. Mitotic regulation of the APC activator proteins CDC20 and CDH1. *Mol. Biol. Cell* **11,** 1555–1569 (2000).
26. Pomerening, J. R., Ubersax, J. A. & Ferrell, J. E. Rapid cycling and precocious termination of G1 phase in cells expressing CDK1AF. *Mol. Biol. Cell* **19,** 3426–3441 (2008).
27. Cross, F. R., Archambault, V., Miller, M. & Kolvstad, M. Testing a mathematical model of the yeast cell cycle. *Mol. Biol. Cell* **13,** 52–70 (2002).
28. Novak, B. & Tyson, J. J. Numerical analysis of a comprehensive model of M-phase control in *Xenopus* oocyte extracts and intact embryos. *J. Cell Sci.* **106,** 1153–1168 (1993).
29. Pomerening, J. R., Sontag, E. D. & Ferrell, J. E. Jr. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biol.* **5,** 346–351 (2003).
30. Foiani, M., Marini, F., Gamba, D., Lucchini, G. & Plevani, P. The B subunit of the DNA polymerase α-primase complex in *Saccharomyces cerevisiae* executes an essential function at the initial stage of DNA replication. *Mol. Cell. Biol.* **14,** 923–933 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Bouchoux, J. Kilmartin, E. Schwob and W. Zachariae for antibodies and constructs, and members of our laboratory for discussion and critical reading of the manuscript. This work was supported by a European Commission Marie Curie Individual Fellowship to S.L.-A., and the BBSRC and EC FP7 (O.K.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.U. (frank.uhlmann@cancer.org.uk).

LETTERS

NAADP mobilizes calcium from acidic organelles through two-pore channels

Peter J. Calcraft[1]*, Margarida Ruas[2]*, Zui Pan[3]*, Xiaotong Cheng[2], Abdelilah Arredouani[2], Xuemei Hao[4,5], Jisen Tang[5], Katja Rietdorf[2], Lydia Teboul[6], Kai-Ting Chuang[2], Peihui Lin[3], Rui Xiao[5], Chunbo Wang[5], Yingmin Zhu[5], Yakang Lin[5], Christopher N. Wyatt[1], John Parrington[2], Jianjie Ma[3], A. Mark Evans[1], Antony Galione[2] & Michael X. Zhu[5]

Ca^{2+} mobilization from intracellular stores represents an important cell signalling process[1] that is regulated, in mammalian cells, by inositol-1,4,5-trisphosphate ($InsP_3$), cyclic ADP ribose and nicotinic acid adenine dinucleotide phosphate (NAADP). $InsP_3$ and cyclic ADP ribose cause the release of Ca^{2+} from sarcoplasmic/endoplasmic reticulum stores by the activation of $InsP_3$ and ryanodine receptors ($InsP_3$Rs and RyRs). In contrast, the nature of the intracellular stores targeted by NAADP and the molecular identity of the NAADP receptors remain controversial[1,2], although evidence indicates that NAADP mobilizes Ca^{2+} from lysosome-related acidic compartments[3,4]. Here we show that two-pore channels (TPCs) comprise a family of NAADP receptors, with human TPC1 (also known as TPCN1) and chicken TPC3 (TPCN3) being expressed on endosomal membranes, and human TPC2 (TPCN2) on lysosomal membranes when expressed in HEK293 cells. Membranes enriched with TPC2 show high affinity NAADP binding, and TPC2 underpins NAADP-induced Ca^{2+} release from lysosome-related stores that is subsequently amplified by Ca^{2+}-induced Ca^{2+} release by $InsP_3$Rs. Responses to NAADP were abolished by disrupting the lysosomal proton gradient and by ablating TPC2 expression, but were only attenuated by depleting endoplasmic reticulum Ca^{2+} stores or by blocking $InsP_3$Rs. Thus, TPCs form NAADP receptors that release Ca^{2+} from acidic organelles, which can trigger further Ca^{2+} signals via sarcoplasmic/endoplasmic reticulum. TPCs therefore provide new insights into the regulation and organization of Ca^{2+} signals in animal cells, and will advance our understanding of the physiological role of NAADP.

NAADP was first identified as a potent intracellular Ca^{2+} mobilizing agent in sea urchin eggs[5], and was later confirmed as such in various mammalian preparations[6–8]. The Ca^{2+} stores mobilized by NAADP seem to be distinct from sarcoplasmic/endoplasmic reticulum (S/ER)[9–11], and accumulating evidence now suggests that NAADP targets lysosome-like acidic compartments[3,4,12–14]. However, it remains unclear whether NAADP can also release the S/ER Ca^{2+} stores in certain cell types, perhaps by acting directly on RyRs[15–17]. Furthermore, cross-talk between NAADP signalling and that mediated by $InsP_3$, Ca^{2+} and cADP ribose exists in many cell types[11,18] (Supplementary Fig. 1), complicating the interpretation of experimental results.

TPCs are new members of the superfamily of voltage-gated ion channels[19,20]. Their predicted structures indicate two-fold symmetry with a total of 12 putative transmembrane α-helices (Fig. 1a). Three non-allelic *TPC* genes are present in sea urchins and most vertebrate species, with TPC3 absent in most primates and some rodent species (Supplementary Fig. 2). The three TPCs are equally distant from each other, and from plant TPC1, with <30% amino acid identity in the conserved transmembrane regions (Fig. 1b and Supplementary Fig. 2). Although *Arabidopsis* TPC1 has been shown to mediate Ca^{2+} release from plant intracellular vacuoles[21], functional data are lacking for animal TPCs.

Similar to the reported widespread expression of *TPC1* messenger RNA[19], northern analysis shows that human *TPC2* mRNA is expressed in most human tissues, with higher levels in the liver and kidneys (Fig. 1c). Immunofluorescence labelling of HEK293 cells using an anti-TPC2 antibody (see Supplementary Fig. 3) showed punctate staining in the cytoplasm, which was blocked by the antigenic peptides (data not shown) and overlaps with that of lysosome-membrane associated protein 2 (LAMP2; Fig. 1d, Pearson's coefficient = 0.92). Similarly, in a stable cell line expressing haemagglutinin (HA)-tagged TPC2 (hereafter termed TPC2 cells), the HA-tagged protein colocalizes with LAMP2 (Fig. 1e, Supplementary Movie 1, Supplementary Table 1, and Supplementary Fig. 4) but not with markers for early and late endosomes, or with markers for ER, Golgi or mitochondria (Supplementary Fig. 5a–e). Moreover, LysoTracker (Fig. 1f), but not MitoTracker or fluorescent transferrin (data not shown), accumulated in intracellular vesicles surrounded by HA–TPC2. Similar results were obtained for heterologously expressed mouse TPC2 (Supplementary Fig. 6).

In contrast, heterologously expressed TPC1 and TPC3 show only sparse colocalization with TPC2 or LAMP2, and are instead predominantly expressed in endosomes and other unidentified intracellular compartments (Supplementary Figs 5f, g, 7 and Supplementary Table 1). Therefore, all mammalian TPCs are expressed intracellularly on endolysosomes, with TPC2 being specifically targeted to lysosomal membranes.

The lysosomal localization and homology to Ca^{2+} channels prompted us to test whether TPC2 forms a binding site for NAADP. Membranes from the TPC2 and wild-type HEK293 cells were incubated with 0.2 nM [^{32}P]NAADP in the absence and presence of unlabelled NAADP (100 μM). TPC2 membranes showed more than a threefold increase in specific binding compared to wild-type membranes (Fig. 2a). To confirm that the binding is associated with expressed TPC2 proteins, we depleted HA–TPC2 from the membranes using an anti-HA antibody, and tested [^{32}P]NAADP binding

[1]Centre for Integrative Physiology, College of Medicine and Veterinary Medicine, University of Edinburgh, Hugh Robson Building, Edinburgh EH8 9XD, Scotland, UK. [2]Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK. [3]Department of Physiology and Biophysics, UMDNJ-Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854, USA. [4]College of Life Sciences, Peking University, Beijing 100871, China. [5]Department of Neuroscience, Biochemistry, and Center for Molecular Neurobiology, The Ohio State University, 1060 Carmack Road, Columbus, Ohio 43210, USA. [6]The Mary Lyon Centre, MRC Harwell, Oxfordshire OX11 0RD, UK.
*These authors contributed equally to this work.

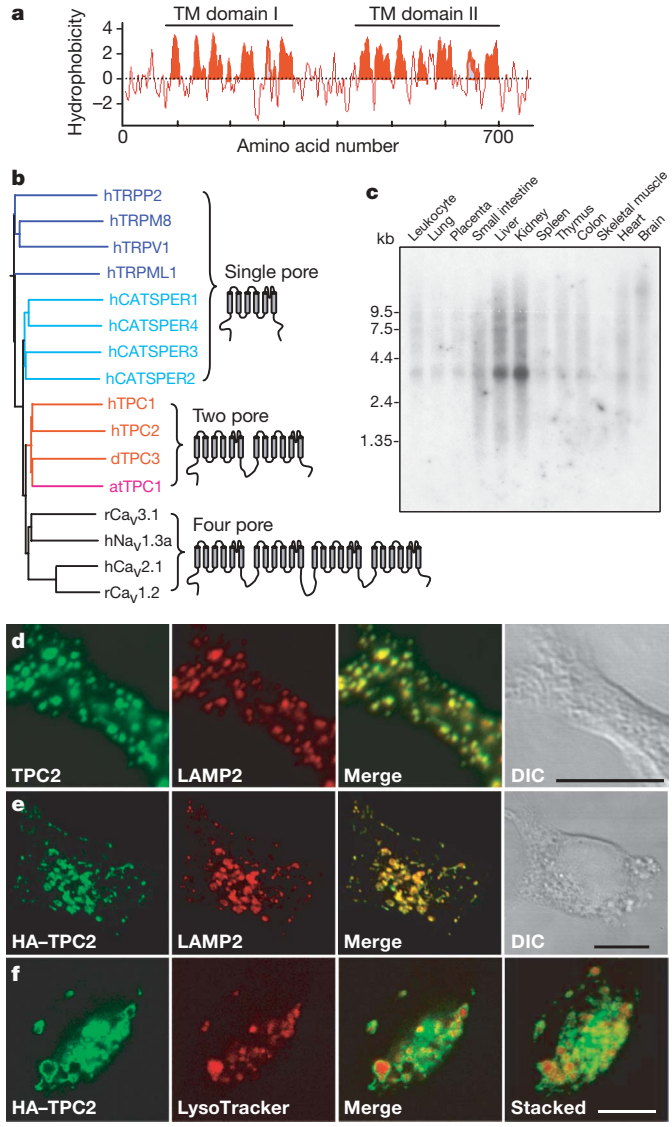


Figure 1 | Tissue and subcellular expression of mammalian TPC2. **a**, Hydropathy plot of human TPC2. Window size = 9 amino acids. **b**, Evolutionary relationships of TPCs with single-pore and four-pore domain channels. at, *Arabidopsis thaliana*; Ca_v, voltage-gated Ca^{2+} channel; d, dog; h, human; Na_v, voltage-gated Na^+ channel; r, rat. **c**, Northern blot analysis of TPC2 expression in human tissues. **d**, Colocalization of endogenous TPC2 with LAMP2 in untransfected HEK293 cells. DIC, differential interference contrast. **e**, **f**, Colocalization of HA–TPC2 with LAMP2 (**e**) or LysoTracker (**f**) in the TPC2 stable cell line. The right panel in **f** shows a stacked image of overall LysoTracker accumulation relative to expressed HA–TPC2 (three-dimensional projections, Supplementary Movie 1). Scale bars, 10 μm.

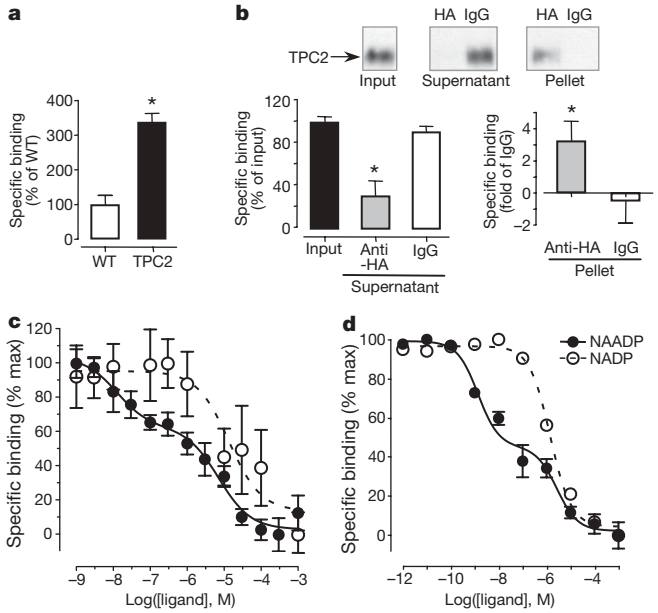


Figure 2 | [^{32}P]NAADP binding to TPC2. **a**, Specific binding for membranes from wild-type (WT) HEK293 and TPC2 cells ($n = 3$). $^*P < 0.05$. **b**, Depletion of TPC2 by immunoprecipitation abolished [^{32}P]NAADP binding. The supernatant was depleted of TPC2 (top panel) and [^{32}P]NAADP binding (bottom panel, left) by anti-HA antibody but not by rat IgG (control) antibody. TPC2 (top panel) and [^{32}P]NAADP binding (bottom panel, right) appeared in the pellet with anti-HA antibody only. **c**, **d**, Representative ligand competition assay for membranes prepared from TPC2 cells (**c**) and from mouse liver (**d**). The maximal specific binding for TPC2 membranes ranged from 167.6 to 300 d.p.m. (disintegrations per minute), and for mouse liver from 1,000 to 1,600 d.p.m.

to the resulting supernatant and pellet. With the anti-HA antibody the binding was mainly associated with the pellet, whereas with a control antibody it remained in the supernatant (Fig. 2b).

A ligand competition assay showed that the TPC2-containing membranes had two affinities to NAADP with dissociation constant (K_d) values of 5.0 ± 4.2 nM (mean ± s.e.m.) and 7.2 ± 0.8 μM ($n = 3$) (Fig. 2c). This binding curve closely resembles that of mouse liver membranes (Fig. 2d), which had affinity values of 6.6 ± 3.5 nM and 4.6 ± 2.4 μM ($n = 4$), and these K_d values for the high affinity binding site compare well with results reported for other mammalian preparations[22–24].

As expected[22,25], NADP—the precursor of NAADP that is unable to mobilize Ca^{2+}—only showed low-affinity binding to TPC2 and mouse liver membranes, with K_d values of 10.3 ± 3.1 μM ($n = 3$) and 4.5 ± 2.3 μM ($n = 4$), respectively. This could also arise from contamination by trace amounts of NAADP in NADP preparations[5]. Moreover, although wild-type membranes showed specific binding to NAADP, the ligand competition assay could only reveal low affinity binding, indicating that the fraction for high affinity binding must be very low. This is supported by quantitative PCR with reverse transcription (qRT–PCR) which showed a >250-fold increase in *TPC2* mRNA in TPC2 cells compared to wild-type cells (Supplementary Information). Therefore, TPC2 expression confers the high affinity NAADP binding. Although we cannot exclude the possibility that interactions with accessory proteins may be necessary for NAADP binding to TPC2, such proteins would have to associate with TPC2 tightly to explain these binding results.

To test whether TPC2 mediates Ca^{2+} release from lysosomes, we studied the effect of flash photolysis of caged NAADP on intracellular Ca^{2+} concentration ($[Ca^{2+}]_i$) in wild-type HEK293 and TPC2 cells by Fluo3 fluorescence. As shown in part a of Fig. 3A, all TPC2 cells responded to photorelease of NAADP with a biphasic $[Ca^{2+}]_i$ transient comprising an initial slow pacemaker-like ramp (10–180 s) and a subsequent large Ca^{2+} transient. No fluorescence increase occurred after ultraviolet flashes if caged NAADP was not included ($n = 6$, data not shown). Furthermore, wild-type cells showed only small and short-lived $[Ca^{2+}]_i$ rises, and lacked both the ramp-like phase and the secondary transient (Fig. 3B).

Consistent with a role for lysosomes in this process, the vacuolar H^+-ATPase inhibitor bafilomycin A1 (1 μM) abolished both phases of the response to NAADP (see d in Fig. 3A), but failed to affect the $[Ca^{2+}]_i$ rise induced by extracellular application of 100 μM ATP, which activates ER Ca^{2+} release (data not shown). In contrast, inclusion of heparin (200 μg ml^{-1}), a competitive inhibitor of $InsP_3Rs$, in the patch pipette only blocked the secondary phase of the Ca^{2+} transient and thereby revealed in its entirety the initial $[Ca^{2+}]_i$ signal triggered by NAADP (see b in Fig. 3A). Consistent with the lack of RyR expression in HEK293 cells[26], both phases of the response to

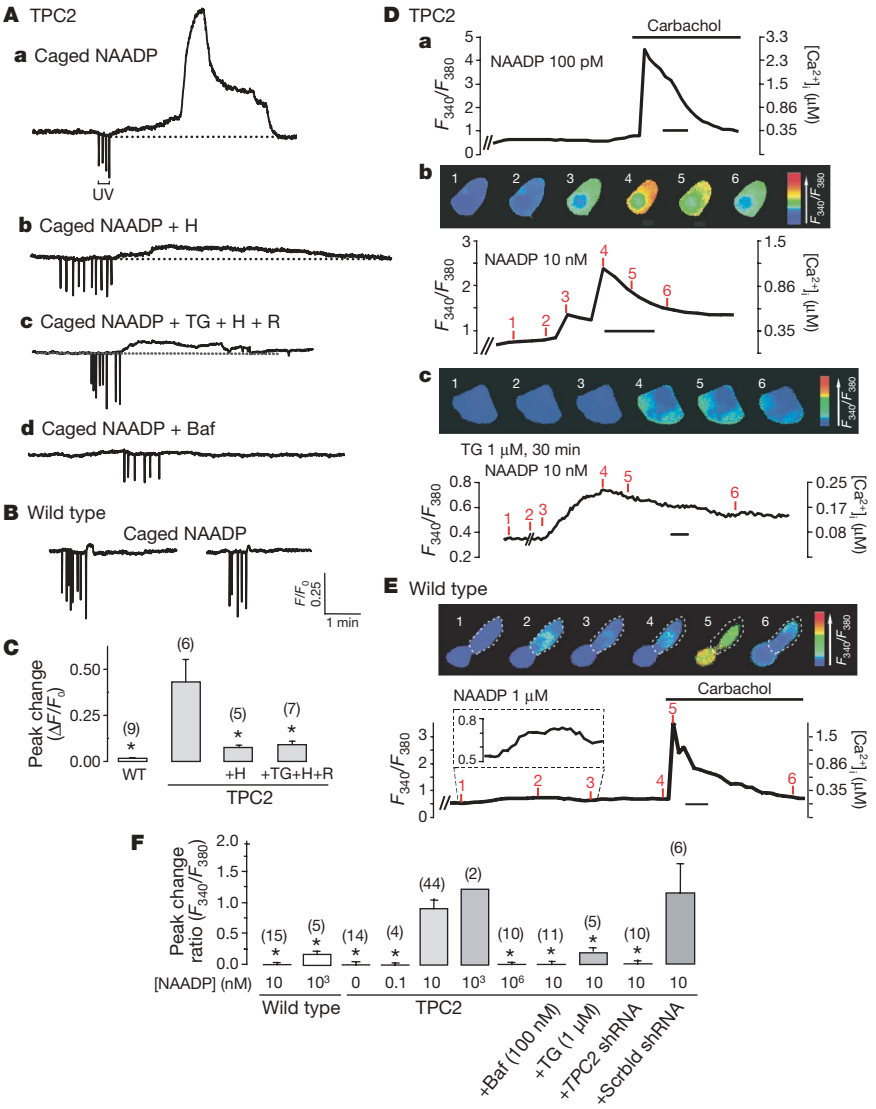


Figure 3 | TPC2 expression and NAADP-evoked Ca^{2+} signalling. **A–C**, Effect of photoreleased NAADP on Fluo3 fluorescence in TPC2 (**A**) and wild-type HEK293 cells (**B**). UV, ultraviolet. **C**, The mean and s.e.m. of the peak response are shown. Baf, bafilomycin; H, heparin; R, ryanodine; TG, thapsigargin.*$P < 0.05$, different from TPC2 only. **D–F**, Effect of intracellular dialysis of NAADP on the Fura2 ratio in TPC2 (**D**) and wild-type (**E**, the cell within dashed lines) cells. Top panels (**b**, **c** in **D**, **E**) show pseudocolour images, bottom panels (**b**, **c** in **D**, **E**) show the ratio against time. Time scale bars, 20 s. **F**, The mean and s.e.m. (except $n = 2$) of the peak response are shown. *$P < 0.05$, different from TPC2, 10 nM NAADP. Scrbld, scrambled. The numbers of cells are shown in parentheses.

photorelease of NAADP persisted in TPC2 cells preincubated with 10 μM ryanodine (data not shown). Furthermore, the combined effects of depleting the ER store by pretreatment with thapsigargin (1 μM), blocking $InsP_3Rs$ with heparin and RyRs with ryanodine caused no further inhibition of the NAADP-induced response than blocking $InsP_3Rs$ with heparin alone (Fig. 3A, c, and C). Therefore, the initial $[Ca^{2+}]_i$ rise is dependent on acidic organelles but independent of the ER, whereas the secondary phase is due to ER Ca^{2+} release by $InsP_3Rs$, presumably through Ca^{2+}-induced Ca^{2+} release together with resting $InsP_3$ levels.

To determine the concentration-response relationship for NAADP, we dialysed known concentrations of NAADP into single cells using patch pipettes, and monitored $[Ca^{2+}]_i$ changes using Fura2 (refs 13,18). In TPC2 cells, whereas 100 pM NAADP did not cause any appreciable $[Ca^{2+}]_i$ rise, 10 nM NAADP elicited a biphasic response reminiscent of that evoked by the photolysis of caged NAADP (a and b in Fig. 3D). Pretreatment with thapsigargin abolished the second but not the first phase (c in Fig. 3D). The response was also seen with 1 μM NAADP but not with much higher concentrations (1 mM; Fig. 3F), consistent with the notion that NAADP-induced Ca^{2+} release desensitizes at high ligand concentrations[6]. In contrast, 10 nM NAADP had no effect in wild-type cells, and 1 μM only induced a small, perhaps more localized, Ca^{2+} transient in three out of five cells (Fig. 3E, F).

Again, the Ca^{2+} transient induced by intracellular dialysis of 10 nM NAADP in TPC2 cells was blocked by bafilomycin. More importantly, the response was abolished by transfection into TPC2 cells of a short hairpin RNA (shRNA) against human *TPC2* (Supplementary Fig. 8), but not that of a scrambled control shRNA (Fig. 3F), demonstrating the essential role of TPC2 in mediating NAADP-induced Ca^{2+} release. All cells responded to extracellularly applied carbachol, which triggers ER Ca^{2+} release, indicating that the cells were viable[27].

To examine the role of endogenous TPC2 in NAADP signalling in a native system, we generated *Tpc2* knockout mice using a gene trap strategy (Fig. 4a–c and Supplementary Information)[28] and isolated pancreatic β-cells in which previous studies have established that NAADP-dependent Ca^{2+} mobilization from a thapsigargin-insensitive acidic store[12,24] underpins the gating of a Ca^{2+}-activated plasma membrane cation current. Figure 4d shows that under the whole-cell configuration, intracellular dialysis of 100 nM NAADP elicited oscillatory

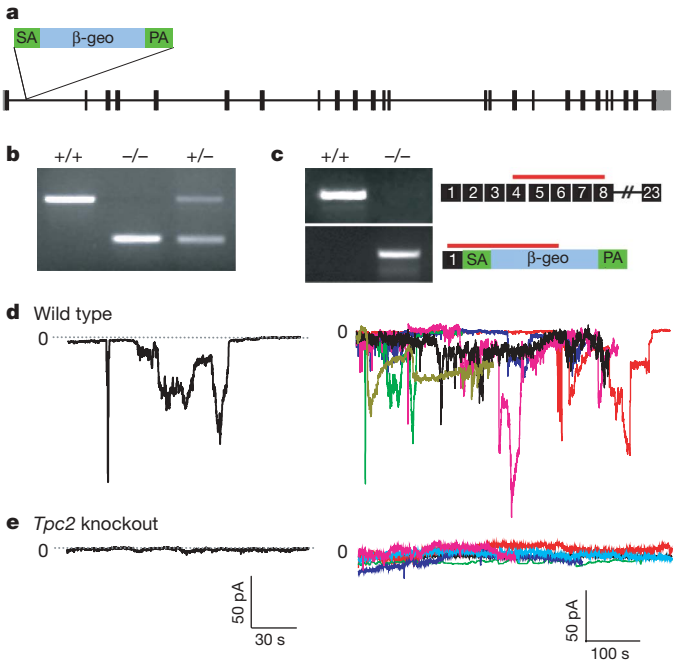


Figure 4 | Pancreatic β-cells from *Tpc2* knockout mice are NAADP-insensitive. a, Approximate position of the gene trap vector in the *Tpc2* gene. PA, polyadenylation signal; SA, splice acceptor. **b**, PCR analysis of genomic DNA. +/+, wild type; +/−, heterozygote; −/−, homozygote. **c**, RT–PCR products from wild-type and mutant *Tpc2* mRNAs with approximate positions of amplicons indicated by red bars; numbers indicate exons (see Supplementary Information for details). **d**, **e**, Cation currents at −70 mV evoked by intracellular dialysis of 100 nM NAADP in pancreatic β-cells isolated from wild-type (**d**) and *Tpc2* knockout (**e**) mice. Left,

inward currents in wild-type β-cells held at −70 mV. No such currents were detected if NAADP was omitted from the pipette solution, if intracellular Ca^{2+} was strongly buffered by 10 mM BAPTA, or if the extracellular cations were replaced by *N*-methyl-D-glucamine (data not shown). Notably, NAADP failed to activate the cation currents in β-cells from the *Tpc2* knockout mice (Fig. 4e), strongly suggesting that TPC2 has a critical role in native NAADP-evoked Ca^{2+} signalling in β-cells.

These results are best explained if TPC2 is a lysosomal Ca^{2+} release channel targeted by NAADP. Although we have focused on TPC2 because of its predominant lysosomal localization, TPC1 and TPC3 may also mediate NAADP-induced Ca^{2+} release from distinct subsets of acidic organelles, such as the distinguished endosome populations suggested by their subcellular distributions. Indeed, we have observed significant but highly localized Ca^{2+} transients in response to 10 nM NAADP in cells that overexpress human TPC1 as opposed to the global $[Ca^{2+}]_i$ changes seen in TPC2 cells (Supplementary Fig. 9 and Supplementary Movie 2). This distinction is consistent with the more restricted subcellular distribution of TPC1 compared to TPC2 (Supplementary Fig. 7).

The biphasic Ca^{2+} response to NAADP in TPC2-overexpressing cells, and the dependence of the later phase on $InsP_3Rs$ and the ER are consistent with the idea that NAADP-induced Ca^{2+} signals are small, and perhaps localized, but able to act, at least via TPC2, as discrete triggers for large global $[Ca^{2+}]_i$ changes through coupling to $InsP_3R$/RyR-S/ER systems. This adds an intriguing possibility for signal diversification, given that the pure NAADP-evoked Ca^{2+} signal is small and highly localized once the ER Ca^{2+} store is depleted by thapsigargin (as shown in parts c of Fig. 3A and D) for heterologously expressed TPC2 as well as the endogenous channels in the human hepatoblastoma cell line HepG2 (Supplementary Fig. 10). However, the localized Ca^{2+} signals will probably reach high levels, particularly at lysosome–ER junctions with unheralded versatility supplied by the fact that TPC-containing vesicles undergo rapid movement within the cytoplasm (Supplementary Movie 3). In this respect it is important to note that NAADP-sensitive Ca^{2+} signals can have several coupling targets. In sea urchin eggs and pancreatic acinar cells, NAADP-induced Ca^{2+} signals are coupled to ER Ca^{2+} release through $InsP_3Rs$ and RyRs[6,11,29]; in pulmonary arterial smooth muscle cells, they seem to selectively target RyRs[13,18]; in pancreatic β-cells, they are coupled to Ca^{2+}-activated cation channels. Thus, the graded local and mobile endolysosome-derived Ca^{2+} signals released via TPCs, through coupling to other systems, are dynamic and versatile. Future investigations on the role of TPCs as NAADP receptors will therefore provide important advances in our understanding of the mechanisms of regulation, spatial organization and diverse functional roles of Ca^{2+} signals in mammalian cells.

METHODS SUMMARY

The complementary DNA for human *TPC2* was cloned from HEK293 cells by RACE–PCR. Northern hybridization was performed using a multi-tissue human mRNA blot (BD Biosciences). For stable expression, the amino-terminal HA-tagged TPC2 was placed in pIRESneo (BD Biosciences), transfected in HEK293 cells, and stable clones were selected and maintained using G418. [^{32}P]NAADP synthesis, membrane purification and radioligand binding studies were carried out as previously described[23,25]. Caged NAADP was synthesized as described[30]. *Tpc2* knockout mice were developed from an embryonic stem cell line (YHD437) containing a gene trap mutation in the *Tpc2* gene. Details for flash photolysis of caged NAADP, intracellular NAADP dialysis, Ca^{2+} imaging, and measurement of Ca^{2+}-activated cation currents are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 November 2008; accepted 2 April 2009.
Published online 22 April 2009.

1. Berridge, M. J., Bootman, M. D. & Roderick, H. L. Calcium signalling: dynamics, homeostasis and remodelling. *Nature Rev. Mol. Cell Biol.* **4,** 517–529 (2003).
2. Galione, A. & Churchill, G. C. Interactions between calcium release pathways: multiple messengers and multiple stores. *Cell Calcium* **32,** 343–354 (2002).
3. Churchill, G. C. *et al.* NAADP mobilizes Ca^{2+} from reserve granules, a lysosome-related organelle, in sea urchin eggs. *Cell* **111,** 703–708 (2002).
4. Yamasaki, M. *et al.* Organelle selection determines agonist-specific Ca^{2+} signals in pancreatic acinar and beta cells. *J. Biol. Chem.* **279,** 7234–7240 (2004).
5. Lee, H. C. & Aarhus, R. A derivative of NADP mobilizes calcium stores insensitive to inositol trisphosphate and cyclic ADP-ribose. *J. Biol. Chem.* **270,** 2152–2157 (1995).
6. Cancela, J. M., Churchill, G. C. & Galione, A. Coordination of agonist-induced Ca^{2+}-signalling patterns by NAADP in pancreatic acinar cells. *Nature* **398,** 74–76 (1999).
7. Bak, J. *et al.* Nicotinic acid adenine dinucleotide phosphate triggers Ca^{2+} release from brain microsomes. *Curr. Biol.* **9,** 751–754 (1999).
8. Berg, I., Potter, B. V., Mayr, G. W. & Guse, A. H. Nicotinic acid adenine dinucleotide phosphate ($NAADP^+$) is an essential regulator of T-lymphocyte Ca^{2+}-signaling. *J. Cell Biol.* **150,** 581–588 (2000).
9. Genazzani, A. A. & Galione, A. Nicotinic acid-adenine dinucleotide phosphate mobilizes Ca^{2+} from a thapsigargin-insensitive pool. *Biochem. J.* **315,** 721–725 (1996).
10. Lee, H. C. & Aarhus, R. Functional visualization of the separate but interacting calcium stores sensitive to NAADP and cyclic ADP-ribose. *J. Cell Sci.* **113,** 4413–4420 (2000).
11. Churchill, G. C. & Galione, A. NAADP induces Ca^{2+} oscillations via a two-pool mechanism by priming IP_3- and cADPR-sensitive Ca^{2+} stores. *EMBO J.* **20,** 2666–2671 (2001).
12. Mitchell, K. J., Lai, F. A. & Rutter, G. A. Ryanodine receptor type I and nicotinic acid adenine dinucleotide phosphate receptors mediate Ca^{2+} release from insulin-containing vesicles in living pancreatic beta-cells (MIN6). *J. Biol. Chem.* **278,** 11057–11064 (2003).
13. Kinnear, N. P., Boittin, F. X., Thomas, J. M., Galione, A. & Evans, A. M. Lysosome-sarcoplasmic reticulum junctions. A trigger zone for calcium signaling by nicotinic acid adenine dinucleotide phosphate and endothelin-1. *J. Biol. Chem.* **279,** 54319–54326 (2004).
14. Macgregor, A. *et al.* NAADP controls cross-talk between distinct Ca^{2+} stores in the heart. *J. Biol. Chem.* **282,** 15302–15311 (2007).
15. Mojžišová, A., Križanová, O., Žáčiková, L., Komínková, V. & Ondriaš, K. Effect of nicotinic acid adenine dinucleotide phosphate on ryanodine calcium release channel in heart. *Pflugers Arch.* **441,** 674–677 (2001).
16. Gerasimenko, J. V. *et al.* NAADP mobilizes Ca^{2+} from a thapsigargin-sensitive store in the nuclear envelope by activating ryanodine receptors. *J. Cell Biol.* **163,** 271–282 (2003).

17. Dammermann, W. & Guse, A. H. Functional ryanodine receptor expression is required for NAADP-mediated local Ca^{2+} signaling in T-lymphocytes. *J. Biol. Chem.* **280,** 21394–21399 (2005).
18. Boittin, F. X., Galione, A. & Evans, A. M. Nicotinic acid adenine dinucleotide phosphate mediates Ca^{2+} signals and contraction in arterial smooth muscle via a two-pool mechanism. *Circ. Res.* **91,** 1168–1175 (2002).
19. Ishibashi, K., Suzuki, M. & Imai, M. Molecular cloning of a novel form (two-repeat) protein related to voltage-gated sodium and calcium channels. *Biochem. Biophys. Res. Commun.* **270,** 370–376 (2000).
20. Anderson, P. A. & Greenberg, R. M. Phylogeny of ion channels: clues to structure and function. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **129,** 17–28 (2001).
21. Peiter, E. *et al.* The vacuolar Ca^{2+}-activated channel TPC1 regulates germination and stomatal movement. *Nature* **434,** 404–408 (2005).
22. Bak, J., Billington, R. A., Timar, G., Dutton, A. C. & Genazzani, A. A. NAADP receptors are present and functional in the heart. *Curr. Biol.* **11,** 987–990 (2001).
23. Patel, S., Churchill, G. C., Sharp, T. & Galione, A. Widespread distribution of binding sites for the novel Ca^{2+}-mobilizing messenger, nicotinic acid adenine dinucleotide phosphate, in the brain. *J. Biol. Chem.* **275,** 36495–36497 (2000).
24. Masgrau, R., Churchill, G. C., Morgan, A. J., Ashcroft, S. J. & Galione, A. NAADP: a new second messenger for glucose-induced Ca^{2+} responses in clonal pancreatic beta cells. *Curr. Biol.* **13,** 247–251 (2003).
25. Patel, S., Churchill, G. C. & Galione, A. Unique kinetics of nicotinic acid-adenine dinucleotide phosphate (NAADP) binding enhance the sensitivity of NAADP receptors for their ligand. *Biochem. J.* **352,** 725–729 (2000).
26. Aoyama, M. *et al.* Requirement of ryanodine receptors for pacemaker Ca^{2+} activity in ICC and HEK293 cells. *J. Cell Sci.* **117,** 2813–2825 (2004).
27. van Koppen, C. J., Meyer zu Heringdorf, D., Alemany, R. & Jakobs, K. H. Sphingosine kinase-mediated calcium signaling by muscarinic acetylcholine receptors. *Life Sci.* **68,** 2535–2540 (2001).
28. Stryke, D. *et al.* BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. *Nucleic Acids Res.* **31,** 278–281 (2003).
29. Cancela, J. M., Gerasimenko, O. V., Gerasimenko, J. V., Tepikin, A. V. & Petersen, O. H. Two different but converging messenger pathways to intracellular Ca^{2+} release: the roles of nicotinic acid adenine dinucleotide phosphate, cyclic ADP-ribose and inositol trisphosphate. *EMBO J.* **19,** 2549–2557 (2000).
30. Morgan, A. J. *et al.* in *Methods in Calcium Signalling* 2nd edn (ed. Putney, J. W. Jr.) 267–334 (CRC, 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the UK Wellcome Trust and the British Heart Foundation to A.G., J.P. and A.M.E., the US National Institutes of Health to M.X.Z. and J.M., and the American Heart Association to M.X.Z. A.G. was a Wellcome Senior Research Fellow in basic Biomedical Science. The work of A.M.E. was funded by a Wellcome Trust Project Grant (CNW, reference number 070772) and a British Heart Foundation Studentship (PJC, reference number FS/05/050). Part of the work of M.X.Z. was made possible by US National Institutes of Health grant P30-NS045758. We thank T. Kong and W. Li for the HepG2 cell line, O. Ogunbayo for technical assistance, and F. Platt, A. Morgan and M. Viapiano for help with the manuscript.

Author Contributions J.T., R.X. and M.X.Z. cloned TPC1, TPC2 and TPC3. J.T., Y.L., C.W., X.H. and M.X.Z. produced stable cell lines and performed immunostaining and confocal microscopy. X.C., M.R., J.P. and A.G. performed radioligand binding. A.A. and A.G. performed flash photolysis experiments. P.J.C., C.N.W. and A.M.E. performed NAADP dialysis, Ca^{2+} imaging, shRNA knockdown, western blotting and associated immunocytochemistry and deconvolution microscopy. M.R., K.R., J.P. and A.G. produced and characterized the TPC2 antibody and performed immunostaining and confocal microscopy. K.R. and Y.Z. performed qRT–PCR, Z.P., P.L. and J.M. produced shRNA constructs and studied the effect of NAADP on HepG2 cells. M.R., A.A., L.T., K.-T.C., J.P. and A.G. produced and characterized the *Tpc2* knockout mice and carried out β-cell studies. M.X.Z, J.M., A.M.E., J.P. and A.G. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The following sequences have been deposited in the GenBank database: AY029200 (human TPC2), AY083666 (human TPC1), EU344154 (chicken TPC3), EU344155 (rabbit TPC3), BK006366 (dog TPC3), BK006367 (zebrafish TPC3), BK006368 (horse TPC3) and BK006573 (bovine TPC3). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.X.Z. (zhu.55@osu.edu).

METHODS

Sequence analysis. For sequence alignment (Fig. 1b), the second transmembrane domain of TPCs, and the fourth transmembrane domain of Ca_v and Na_v were aligned with the transmembrane domains (including all six transmembrane segments and the pore-loop) of selected TRP members and CATSPER proteins. GenBank accession numbers are given in Supplementary Fig. 2. The phylogenetic tree was generated by ClustalW (http://align.genome.jp) and plotted using a Neighbour-Joining algorithm.

DNA constructs and stable cell lines. The cDNA for *TPC2* was cloned from HEK293 cells by RACE–PCR using primers designed against the expressed sequence tag AA309878. Northern hybridization was performed using a multi-tissue human mRNA blot (BD Biosciences) and random-labelled cDNA probes against a BglII fragment (nucleotides 1547–2012) of *TPC2* mRNA. The cDNA for *TPC1* was obtained from Research Genetics (clone ID: CS0DB002YA14) and confirmed by DNA sequencing. For stable expression, the HA epitope was added to the N terminus of TPC2, and the cDNA was placed in the pIRESneo vector (BD Biosciences). A His_6-tag was added to the carboxy terminus of TPC1, and the cDNA was placed in the pIREShyg2 vector (BD Biosciences). Transfection into HEK293 cells was performed using Lipofectamine 2000 (Invitrogen) following the manufacturer's protocol. Stable clones were selected and maintained in regular culture medium containing 400 μg ml^{-1} G418 (for TPC2) or 100 μg ml^{-1} hygromycin B (for TPC1). Monoclonal cell lines were established through limiting dilution.

Immunocytochemistry. A polyclonal anti-TPC2 antibody was made by immunizing rabbits with two synthetic peptides (GGKQDDGQDRERLTY and VKEHPPRPEYQSPFL) conjugated to keyhole limpet hemocyanin and affinity purified using the antigenic peptides. Cells were seeded onto poly-ornithine coated coverslips and grown overnight. After washing in PBS (pH 7.4), the cells were fixed using ice-cold methanol (−20 °C, 15 min, Fig. 1d) or 4% paraformaldehyde (22 °C, 1 h, Fig. 1e) and blocked using 5% dry milk in PBS. Immunostaining followed standard protocols. The rabbit anti-TPC2 antibody was diluted 1:200 and the secondary antibody (1:500) was an Alexa 488-conjugated anti-rabbit IgG (Invitrogen). The rat anti-HA antibody (Roche) was diluted 1:500. The secondary antibody (1:500) was an Alexa 488-conjugated anti-rat IgG (Invitrogen). Mouse anti-LAMP2 monoclonal antibody (Developmental Studies Hybridoma Bank) was diluted 1:500. The secondary anti-mouse antibodies (1:500) were conjugated with Alexa 594. Incubations were at 4 °C overnight. All washes were carried out in the dark at room temperature using PBS containing 0.05% Tween 20. The washed slides were mounted using Gel/Mount (BioMeda). Images were acquired using a Leica TCS SL laser scanning confocal system on a Leica DMIRE2 inverted microscope with a ×100 1.4 numerical aperture oil immersion objective and a pinhole of 182 μm. DIC images were acquired simultaneously to show the shape of corresponding cells.

LysoTracker labelling and imaging. TPC2 cells were incubated with 200 ng ml^{-1} LysoTracker Red DND-99 (Invitrogen) for 30 min and then fixed with 4% paraformaldehyde and immunostained using the rat anti-HA antibody as described above. Images were taken using a Bio-Rad MRC 1000 confocal system attached to a Nikon Optiphoto-2 microscope equipped with a ×60 1.4 numerical aperture oil immersion objective, at iris 2.0 and a *z* step size at 0.5 μm.

NAADP-binding experiments. [^{32}P]NAADP synthesis, membrane purification and radioligand binding studies were carried out as previously described[23,25]. In brief, membranes were incubated with 0.2 nM [^{32}P]NAADP, with or without 100 μM unlabelled NAADP, at room temperature for 60 min and terminated by rapid vacuum filtration through GF/B filters. For ligand competition assays, membranes were pre-incubated with desired concentrations of unlabelled NAADP or NADP at room temperature for 10 min before 0.2 nM [^{32}P]NAADP was added and the incubation continued for another 60 min. Immunodepletion experiments were performed by immunoprecipitation of microsomal membranes of TPC2 cells solubilized for 60 min at 4 °C in immunoprecipitation buffer containing (mM) 150 NaCl, 20 HEPES, 1 EDTA (pH 7.2) with 1% CHAPS and 1× proteinase inhibitor. The same amount of solubilized membrane was incubated with the indicated antibodies at 4 °C overnight and then bound to Protein G sepharose beads (GE Healthcare, 120 min, 4 °C). Beads were centrifuged at 1,000*g* for 1 min and supernatants were collected. Beads were washed three times with immunoprecipitation buffer and bound proteins were either eluted with Laemmli buffer for western blotting or used for [^{32}P]NAADP binding. For solubilized immunoprecipitation samples, extracts (same amount of total protein) were incubated in Glu-IM (in mM: 250 potassium gluconate, 250 *N*-methylglucamine, 20 HEPES, 1 MgCl, pH 7.2) supplemented with 0.2 nM [^{32}P]NAADP with or without 10 μM of unlabelled NAADP for 60 min at 22 °C. Then, 500 μg γ-globulin (Sigma) and polyethylene glycol (15% (w/v), Sigma) were added and the incubation continued for 30 min. After centrifugation at 13,000*g* for 5 min, the pellets were washed with 15% (w/v) polyethylene glycol and dissolved in H_2O for scintillation counting. For proteins coupled to agarose beads, the beads were incubated with Glu-IM as described above and then washed three times with Glu-IM. Proteins were eluted by the addition of 2% SDS, collected and spotted on GF/B filters for determining radioactivity by phosphorimaging.

Flash photolysis of caged NAADP and measurement of $[Ca^{2+}]_i$ changes. Wild-type HEK293 and TPC2 cells were seeded at low density (10% confluency) onto poly-D-lysine coated tissue culture dishes (FluoroDish; World Precision Instruments) one day before the experiments. The intracellular solution contained (mM): 140 KCl, 1 $MgCl_2$, 10 HEPES and 0.1 Fluo3 pentapotassium (pH 7.2 with KOH). The bath was a physiological salt solution (PSS) that contained (mM): 130 NaCl, 5.2 KCl, 1$MgCl_2$, 1.7 $CaCl_2$, 10 glucose, 10 HEPES (pH 7.4 with NaOH). Cells were held at −40 mV in the whole-cell configuration with the pipette solution containing 10 μM caged NAADP. When needed, heparin (200 μg ml^{-1}) was added to the intracellular solution; thapsigargin (1 μM), ryanodine (10 μM), and bafilomycin (1 μM) were included in the bath for pre-incubation. Fluo3 was excited at 490 nm and fluorescence emission at 530 nm was monitored using a photomultiplier tube. The caged NAADP was photolysed with an XF-10 arc lamp (HI-TECH Scientific). To avoid prolonged ultraviolet exposure, several 10-ms ultraviolet flashes, shown as downward strokes in Fig. 3A, B, were applied to yield sufficient photolysis of caged NAADP, which has a much lower uncaging efficiency than most other caged compounds, for example, caged $InsP_3$. Ten micromolar caged NAADP was found to be optimal because higher concentrations yielded no response, presumably owing to desensitization caused by the presence of trace amounts of free NAADP in the caged preparation and/or the bell-shaped dose response relation to NAADP[6].

Intracellular dialysis of NAADP and Ca^{2+} imaging. NAADP was applied intracellularly in the whole-cell configuration of the patch-clamp technique, adapted from an established protocol described previously[13]. In brief, cells grown on poly-D-lysine coated glass-bottom dishes overnight were incubated for 30 min with Fura2-AM (5 μM) in PSS. Dishes were placed on a Leica DMIRBE inverted microscope and superfused with Fura2-free PSS for at least 30 min before experimentation. Control and NAADP pipette solutions containing (mM): 140 KCl, 10 HEPES, 1 $MgCl_2$ and 5 μM Fura2, pH 7.4, and the desired concentration of NAADP, were applied by intracellular dialysis from a patch pipette while the cell was held at −40 mV in the whole-cell configuration. Intracellular Ca^{2+} concentration was reported by Fura2 fluorescence ratio (F_{340}/F_{380} excitation; emission 510 nm). Emitted fluorescence was recorded at 22 °C with a sampling frequency of 0.2 Hz, using a Hamamatsu 4880 CCD camera via a Zeiss Fluar ×40, 1.3 numerical aperture oil immersion lens and Leica DMIRBE microscope. The seal resistance, as measured using an Axopatch 200B amplifier (Axon instruments), was ≥2 GΩ throughout each experiment. Series resistance and pipette resistance were ≤10 MΩ and ≤3 MΩ, respectively. Background subtraction was performed online. Analysis was performed using Openlab imaging software (Improvision). When needed, the TPC2 cells were preincubated with thapsigargin (1 μM) for 30 min or bafilomycin (100 nM) for 45 min. The method for shRNA transfection is described in Supplementary Information.

Measurement of NAADP-sensitive cation currents in pancreatic β-cells. Islets of Langerhans were obtained by collagenase digestion of the pancreas, and single cells were prepared by dispersion of islet cells in a Ca^{2+}-free medium. Cells were cultured for 1–4 days in RPMI 1640 medium containing 10 mM glucose.

NAADP was infused through a patch pipette in the whole-cell configuration. The pipette solution contained (mM): 125 potassium-gluconate, 10 KCl, 10 NaCl, 10 KCl, 1 $MgCl_2$, 3 Mg-ATP, 0.1 Na-GTP, 10 EGTA and 5 HEPES (pH 7.2 with KOH). NAADP was added at a final concentration of 100 nM. The bath solution contained (mM): 120 NaCl, 4.8 KCl, 2.5 $CaCl_2$, 1.2 $MgCl_2$, 24 $NaHCO_3$, 10 glucose, 5 HEPES (pH 7.4 with NaOH). The holding potential was −70 mV. The experiments were carried out at room temperature using Multiclamp 700B and the software pClamp 9 (Axon Instruments). Patch pipettes were pulled from borosilicate glass capillaries (World Precision Instruments) to give a resistance of 3–5 MΩ when filled with the pipette solution.

PROSPECTS

Do women have less success in peer review?

An extensive collaborative analysis concludes that the perception is unwarranted, say Herbert Marsh and Lutz Bornmann.

Peer review assesses what is of value in science, yet it has been widely criticized for biases. One such perceived bias is gender. But evidence for such a bias has been contradictory. A 2007 meta-analysis (L. Bornmann *et al. J. Informet.* **1,** 226–238; 2007; see also *Nature* **445,** 566; 2007) concluded that women are at a disadvantage in peer review. As this study incorporated all known research on this issue, it seemed a definitive answer.

However, a study published last year (H. W. Marsh *et al. Am. Psychol.* **63,** 160–168; 2008) presented conflicting results. It was the most comprehensive primary-research study, based on data from the Australian Research Council (10,023 reviews by 6,233 external assessors of 2,331 proposals from all disciplines). The study found that the gender of the applicant had no effect on

IMAGES.COM/CORBIS

the outcomes of peer review, irrespective of the discipline, the gender and nationality of the reviewers, and whether reviewers were selected by a funding panel or chosen by the applicants.

Why should these two studies have conflicting results? To investigate, both research teams worked together to reanalyse the data and extend the original meta-analysis. We applied new, stronger statistical approaches to 66 sets of results representing 353,725 proposals from 8 countries. In this extended study, which will be published in *Review of Educational Research*, we found no effect of the applicant's gender on the peer review of their grant proposals. This lack of effect held across country, year of publication of the studies included in the meta-analysis, and disciplines ranging from physical sciences to the humanities.

The study did, however, reveal very small — but statistically significant — gender differences in favour of men for the 26 sets of results that were for fellowship applications. However, these fellowship results varied greatly between the individual studies within the analysis, indicating that they are not generalizable. We suggest that the differences might have arisen because fellowship applicants tend not to have established a solid track record in their research. In the absence of sound evidence on which to base their judgements, peer reviewers might therefore have been influenced by irrelevant characteristics such as gender.

At least for grant applications, all of the co-authors from each of the research teams agree that the weight of evidence suggests that the applicant's gender has no effect on the outcome of peer review, and that these findings are robust and broadly generalizable. ■

Herbert Marsh is a professor of education at the University of Oxford, UK, and Lutz Bornmann is a PhD student at the ETH University in Zurich, Switzerland.

NEWS

Australian budget bolsters innovation

Australia's Labor government is countering fiscal challenges with a budget increase of nearly 25% to Aus$8.6 billion (US$6.5 billion) for science and innovation.

The inherent job prospects will come primarily from research opportunities associated with infrastructure spending. Most of the Aus$3.1 billion earmarked in the budget for research and development over the next four years will be spent on building the world-class facilities needed to fuel future research enterprises and industries.

"We are going to transform our economy by creating jobs for today, while at the same time building capacity and infrastructure for tomorrow," says Kim Carr, Australia's minister for innovation, industry, science and research.

The Super Science Initiative designates more than Aus$1 billion for increasing capacity in three key research areas to make Australia more competitive internationally: astronomy, marine and climate change, and future industries. For example, the budget allocates Aus$161 million to boost astronomy, including the establishment of an Australian National Centre of Square Kilometre Array Science to help its bid to site the giant telescope in Australia. Of Aus$388 million to boost marine and climate change research, Aus$120 million will buy a new research vessel. More than Aus$500 million is intended to encourage such industries as nanotechnology and biotechnology.

Kim Carr: aiming to transform the Australian economy.

Kurt Lambeck, president of the Australian Academy of Science, says that the increased spending on research and development is a welcome surprise, but the lack of job opportunities has been a disappointment. "All the emphasis is on infrastructure, but what I call the 'human infrastructure' is missing," says Lambeck.

The budget's commitment to human capital is focused primarily on training, aimed at attracting students to scientific fields. The budget increases the total number of PhD students receiving government stipends from 20% to 42%, and increases their tax-free pay by some Aus$2,000 annually to Aus$22,500 per year. The Super Science Initiative also provides for 100 new fellowships for early-career researchers during the next four years. Changes to laureate schemes should attract international scholars.

Mid-career opportunities have been overlooked, says Lambeck, and funding pressures prevent many universities replacing retired researchers. As the population is ageing, this is causing a gap to form at the mid-career stage.

Lambeck believes that the most important part of the budget is the change in government attitude. Providing for new research centres and equipment, and seeking new research frontiers, show a long-term commitment, he says.

Carr says this is a 10-year plan to accelerate economic recovery by harnessing research. "We're turning ideas into jobs," he says. ■

Virginia Gewin

Q&A

Cherry Murray, soon to be dean of the School of Engineering and Applied Sciences at Harvard University, discusses 'Renaissance engineers'.

When did you know you would be a scientist?
I figured from an early age I would become an artist because my parents are both artists. In high school, however, I got interested in science when I had a wonderful chemistry teacher. At the same time, my older brother was studying physics at MIT [Massachusetts Institute of Technology]. On a visit home, he told me I would never succeed in physics and definitely not at MIT. His challenge spurred me to get accepted in physics at MIT.

Was Bell Labs, where you worked for a while, really a scientific Shangri-La?
Yes, for a couple of reasons. First, no one was allowed to build an empire. Nobody had a huge workforce, so we had to collaborate. Second, we were simply focused on doing the best science. If an exciting discovery in physics was made someplace in the world, you could expect 25 people at Bell Labs to drop what they were doing to work day and night on the new subject. It was all self-assembling. No one organized these responses.

Do you think physics has suffered in recent years without a creative entity like Bell Labs?
I think, in the United States, physics has suffered from lack of funding. The actual amount of funding for fundamental physics, in real dollars accounting for inflation, has decreased. It does a disservice to the field. Obviously you want to spend federal money wisely and be transparent, but Congress increasingly wants to control exactly what scientists can study. It's hard to fund good science that way. Scientists have to have some flexibility.

Do you think your science has suffered or benefited from your work as an administrator?
I know some people see management as paper pushing, but it doesn't have to be. My management experience gave me a new perspective on different areas of science and made me a better scientist as a result.

What motivated your move to Harvard?
This is a career opportunity of a lifetime. The School of Engineering and Applied Sciences is exceptional. Even though it is small, it is full of excellent people who work well in interdisciplinary teams. In its efforts to strengthen three core areas — bioengineering, information science, and applied materials and physics for device engineering — Harvard plans to better connect the sciences to the professional schools. With these goals driving the science, I want to forge collaborations that will build on a broad base of knowledge and produce a new class of Renaissance engineers rooted in good science.

What is the secret of scientific success, in your opinion?
Humility. I think success comes from a combination of hard work and realizing that you don't understand something.

Interview by Virginia Gewin.

POSTDOC JOURNAL

Postdoc, you'd better network!

This imperative phrase rang in my brain as I attended a major conference in my field recently. The meeting-abstract deadline had passed me by as I was ensconced in new motherhood. But mindful of a modest travel allowance that I must spend this year, I packed a bag, left my son with his capable dad and flew off for some excellent networking adventures.

I looked forward to reuniting with long-time colleagues and mentors, as well as meeting new researchers with whom collaborations might be fruitful. And, for better or for worse, I would give prospective search-committee members a face and a personality to match my CV, should it cross their desks. As a fledgling scientist, I feel I must ply all the tools of the trade at my disposal to find that right tenure-track opportunity. That said, I'm not a natural networker.

To motivate my keen self, I aimed to meet at least five new colleagues a day during the four-day conference. As I weaved among symposia, coffee breaks and receptions, I swapped business cards, clinked pints and mentally sketched hypotheses (not always best done alongside beer, as the hypotheses seldom look as pretty the next morning).

An excess of late nights and early mornings led quite literally to 'conference-itis', as I came down with a cold when I arrived home. Yet my trip was worthwhile. I achieved most of my networking goals, which I can only hope will translate into job opportunities in the near future. ■

Julia Boughner is a postdoc in evolutionary developmental biology at the University of Calgary, Canada.

IN BRIEF

Supervising productivity

Women engineers at at least one company enjoy their jobs more and feel more productive when they have a strong, positive relationship with a supervisor, a study has found. The study (C.-P. Lin *et al. Soc. Sci. J.* 46, 192–200; 2009) looked at a manufacturer of liquid-crystal displays based in Taiwan. The authors found women's job enjoyment and perception of productivity depend on how they view their relationships with supervisors. "It is more critical for females than for males to have an enthusiastic supervisor who treats subordinates as good friends," the authors write. "Strengthening social ties may be an appropriate technique for supervisors to use to significantly increase perceived job productivity among female staff."

US to dominate biotech?

An international annual survey has found that the United States is expected to increase its lead in biotechnology. UK law firm Marks & Clerk polled 365 executives in the drug, biotechnology, higher-education and venture-capital sectors. President Barack Obama's initiatives will boost the US position as a centre for the industry, said 85%. This threatens the role of Europe, despite problems in US biotech (see *Nature* 459, 467; 2009). Co-author Paul Chapman, a partner in the firm, says the United States is showing new support for regenerative medicine and potential acceptance of stem-cell research. "Europe and the UK cannot afford to watch from the sidelines," he says in the report.

Connected universities

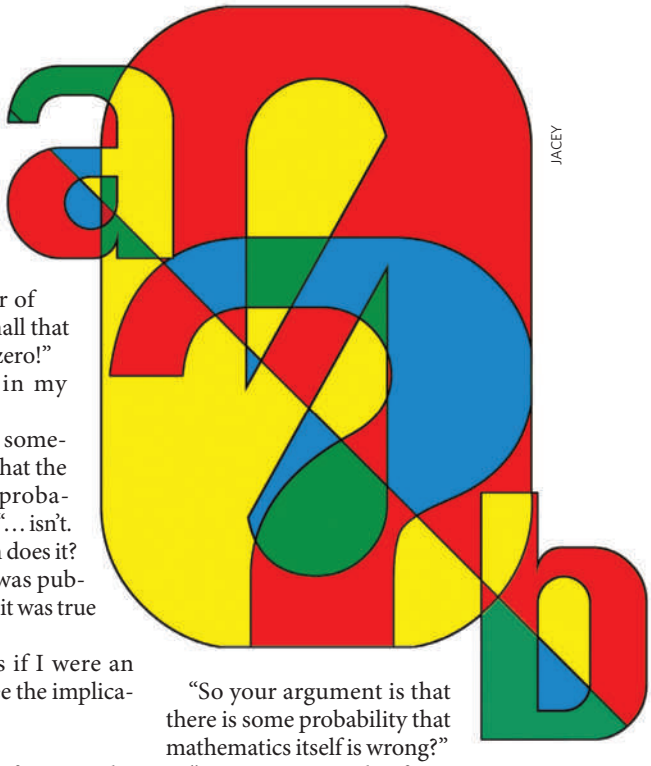
A UK report calls for more recognition of the contributions made by university scientific and other research to national economic prosperity. The report, by the National Endowment for Science, Technology and the Arts, says that the University of Dundee's excellence in life sciences, for example, led to the creation in the Scottish city of one of Britain's leading life-sciences clusters. Policy-makers have had difficulty putting the excellence of UK universities to economic use, says 'The Connected University', but now "this issue has become urgent". Noting that "the innovative businesses our universities create and support will be essential to allowing us to emerge strongly from the recession", the report looks at ways to make that a reality.

NATURE|Vol 459|28 May 2009

Not a chance

It just doesn't add up.

Peter Haff



JACEY

We have been together a long time, and I have got to know my student well. I would like to be doing other things, but the problem he came to discuss won't let us go. Roscoe had been a student in my introductory physics class when, a number of years ago, he appeared uninvited at my office door.

"It's about mathematics," he began without prologue.

I pondered the bearded young man a moment, then took off my glasses and began polishing them. "I am happy to talk about mathematics," I began, "but you realize I'm not a mathematician …"

"Well, it's also about physics," he interrupted, and sat down without invitation in the hard-backed seat that faced my desk.

I leaned back in my swivel chair and looked at him. This is going to take a while, I thought. How true that would be I had no idea. I put my glasses back on.

"It started when I heard about the proof of the four-colour theorem."

I nodded. "Only four colours are needed to distinguish between any arrangement of countries on a two-dimensional map."

"Yes, but the problem is, we don't really know if the proof is right." Roscoe looked at me carefully.

I picked up a pencil and twiddled it between my fingers. I knew that very long computer proofs, like that first used several decades ago in proving the four-colour theorem, are controversial in the mathematics community. It is difficult for any one person to be sure that every line of code is correct.

"The mathematicians seem to have come to a consensus that there are no mistakes in the proof," I remarked mildly.

"Consensus!" he blurted out. "Consensus can't reduce the probability of a theorem being wrong to zero!"

Roscoe's face had turned red and he was leaning towards me at an alarming angle. I swivelled slightly in my chair.

"I see your point," I remarked carefully, "but …"

I stopped swivelling and waited.

"There is always a non-zero probability of a making a physical mistake," he continued. "That's why it's a matter of physics. Maybe the mathematician misreads a comma. She wrongly concludes the proof is correct."

I started to swivel back and forth again. "But many smart mathematicians have combed through the proof in detail, so the chances that …" I stopped.

Roscoe beamed as if he had just caught a schoolboy in an elementary mistake. "Exactly! It's a matter of chance. The chances are small that the proof is wrong, but not zero!"

I rocked more slowly in my chair.

"But look," I finally said, somewhat defensively, "let's say that the proof is wrong, which it probably …" I looked a little guilty "… isn't. It doesn't really matter much does it? After all, before the proof was published no one knew whether it was true or not."

Roscoe looked at me as if I were an imbecile. "But surely you see the implication, professor?" he asked.

I looked puzzled.

"Let's say the probability of error in the four-colour theorem proof is *P*. But now consider another theorem whose proof is just half as long. The chances of error are less than *P* in that case, but certainly not zero. Now by induction," he continued …

I stopped swivelling. Roscoe leaned even further towards me.

"So let's consider a 100-line proof, a 10-line proof, a 2-line proof. Every line has some chance, however small, of being wrong. If a mathematician can make a physical mistake once — of perception, of memory, of whatever mental processes enable him to conclude the validity of one line of code, of one step in a proof, then he can make a mistake again … and again."

As if anticipating my objections, he interrupted himself. "And because it's a matter of physics, there are consequences, professor."

I don't know if what he did next was deliberate, or whether he just slipped.

Roscoe had tilted completely out of his chair, falling face first, in slow motion. Before I recovered my surprise he was halfway to the floor.

He continued his line of argument. "If a mathematician can make one mistake, then he can make the same mistake twice, or make another mistake on another line. A second mathematician also faces a similar chance of error, so does a third …"

I was transfixed by the impending collision of Roscoe with the floor, but somehow felt far away, as if I could do nothing about it.

"So your argument is that there is some probability that mathematics itself is wrong?"

"Yes! Not just the four-colour theorem, but the most elementary proof could be in error."

Roscoe was three-quarters of the way to the floor. I could see only the back of his head.

"What other sort of proof could be wrong?"

"Well, maybe Euclid's proof that there is an infinite number of primes is wrong."

For some reason I was reminded of Wigner's observation on the uncanny ability of mathematics to mimic the physical world. What would it mean for physics if there was a Eureka moment when someone finally realized that mathematics was wrong?

I pondered Roscoe's argument. My pencil twiddled more slowly between my fingers.

He continued: "Perhaps there is something fundamentally wrong with summing an infinite series. Maybe all those limits when infinitesimals are supposed to go to zero don't really exist. Or maybe …"

I stared at the receding Roscoe whose face was now just inches from the floor. My pencil had stopped twiddling. "Maybe what?"

My chair was now still. I didn't blink. I didn't move.

"Or maybe Zeno was right," Roscoe said.

He was very close to the floor. ■

Peter Haff, a geologist at Duke University, is currently studying the emergence of technology as a geological process in the evolution of Earth.